

PROCESSUS STOCHASTIQUES



CHAÎNES DE MARKOV, MARTINGALES ET MOUVEMENT BROWNIEN

Chargé de cours : **Laurent Tournier**
Chargés de TD : **Bastien Mallein** et **Laurent Tournier**

Basé sur un polycopié de **Yueyun Hu**

Table des matières

0	Introduction	5
0.1	Espace de probabilités	5
0.2	Processus stochastiques	10
0.3	Deux résultats fondamentaux	11
1	Chaînes de Markov	13
1.1	Définition	13
1.2	Probabilités de transition à m pas	15
1.3	Exemples de chaînes de Markov	16
1.3.1	Chaîne de markov à deux états	16
1.3.2	Marche aléatoire sur \mathbb{Z} et sur \mathbb{Z}^d	16
1.3.3	Marches aléatoires simples sur \mathbb{Z} avec barrières	17
1.3.4	Modèle d'Ehrenfest	18
1.3.5	Processus du renouvellement	18
1.3.6	Processus de branchement	18
1.4	Propriété de Markov	19
1.5	Visites à un sous-ensemble	21
1.6	Récurrence et transience	22
1.7	Décomposition en classes de communication	23
1.8	Exemples	25
1.8.1	Chaîne à deux états	25
1.8.2	Marche aléatoire simple sur \mathbb{Z}	25
1.8.3	Marche aléatoire simple symétrique sur \mathbb{Z}^d	26
1.8.4	Processus de renouvellement	26
1.8.5	Processus de branchement	26
1.9	Probabilités d'absorption	26
1.9.1	(*) Processus de branchement	27
1.10	Mesures et Probabilités invariantes	29
1.10.1	Existence de mesures invariantes	29
1.10.2	Probabilités invariantes	32
1.11	Périodicité et existence de la loi limite au sens fort	35
1.11.1	(*) Classes cycliques	38
1.12	(*) Complément sur le processus canonique	40
1.12.1	Première construction	40
1.12.2	Autre approche : extension de lois fini-dimensionnelles compatibles	41
2	Espérance conditionnelle	43
2.1	Définition	44
2.1.1	Cas des variables aléatoires dans L^2	44
2.1.2	Cas général : variables aléatoires positives ou dans L^1	45
2.2	Propriétés élémentaires	47
2.3	Espérance sachant une v.a. discrète. Loi conditionnelle	49
2.4	Cas des lois à densité. Loi conditionnelle sachant Y	50

2.5	Propriété de Markov forte et ses applications	51
2.5.1	Temps d'arrêt et propriété de Markov forte	51
2.5.2	Application au théorème ergodique	53
3	Martingales en temps discret	55
3.1	Introduction	55
3.2	Définitions et exemples	56
3.3	Premières propriétés	58
3.4	Stratégies : temps d'arrêt et transformée de martingale	58
3.5	Théorème d'arrêt	61
3.6	Inégalités maximales	62
3.7	Théorèmes de convergence	63
3.7.1	Convergence dans L^2	63
3.7.2	Convergence presque sûre	63
3.7.3	Preuve du Théorème 3.28 :	64
4	Vecteurs gaussiens	69
4.1	Rappels sur la loi normale unidimensionnelle	69
4.2	Extension de la définition à \mathbb{R}^d	70
4.3	Indépendance et conditionnement	72
4.4	Lois normales et limites	73
5	Introduction au mouvement brownien	75
5.1	Définition	75
5.1.1	Motivation : limite d'échelle de marches aléatoires	75
5.1.2	Définition	77
5.1.3	Construction	78
5.2	Propriétés	81
5.2.1	Régularité	81
5.2.2	Invariances	82
5.3	Propriétés de Markov et de martingale, et conséquences	83
5.3.1	Propriété de Markov	83
5.4	Propriété de martingale	84
5.5	Applications de la propriété de martingale	85

Chapitre 0

Introduction et Rappels

Les probabilités ont pour but l'étude des phénomènes aléatoires, c'est-à-dire des expériences qui, lorsqu'elles sont répétées, produisent une succession de résultats imprévisibles. Une variable aléatoire réelle représente ainsi une certaine quantité mesurée lors d'une expérience aléatoire. Dans le cas d'expériences qui se déroulent au fil du temps, il est pertinent de s'intéresser à l'évolution de ces quantités en fonction du temps : elles deviennent donc des *fonctions aléatoires* (c'est-à-dire des variables aléatoires dont les valeurs sont des fonctions), que l'on nomme **processus stochastiques**. Dans le cas où on n'observe l'expérience que lors d'une suite de temps donnés, et non de façon continue, on parle de processus stochastiques **en temps discret**, et il s'agit alors plus précisément de *suites aléatoires*.

Dans ce premier chapitre, on rappelle brièvement les principales notions et résultats de probabilités qui seront utilisés dans la suite, et on définit formellement ces processus stochastiques.

0.1 Espace de probabilités

Une expérience aléatoire se modélise par un **espace de probabilité** $(\Omega, \mathcal{F}, \mathbb{P})$, donné par trois éléments :

- **Un ensemble Ω** . Il représente l'ensemble des résultats possibles de l'expérience aléatoire considérée. Un élément ω de Ω s'appelle un **résultat**, ou une **réalisation** de l'expérience.
- **Une tribu (ou σ -algèbre) \mathcal{F} de sous-ensembles de Ω** . Elle est composée de tous les sous-ensembles de résultats dont on pourra considérer la probabilité. Un élément A de \mathcal{F} s'appelle un **événement** et, pour $\omega \in \Omega$, on dit que A est **réalisé** par ω si $\omega \in A$. On rappelle qu'une classe \mathcal{F} de sous-ensembles de Ω est appelée une **tribu (ou σ -algèbre)** sur l'ensemble Ω si

(i) \mathcal{F} contient \emptyset et Ω ,

(ii) \mathcal{F} est stable par passage au complémentaire : pour tout $A \in \mathcal{F}$, $A^c \in \mathcal{F}$,

(iii) \mathcal{F} est stable par union dénombrable : pour toute suite $(A_n)_{n \in \mathbb{N}}$ à valeurs dans \mathcal{F} , $\bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$.

Exemple 0.1. Sur Ω , on note $\mathcal{P}(\Omega)$ l'ensemble de tous les sous-ensembles de Ω y compris l'ensemble vide. Alors $\mathcal{P}(\Omega)$ est la plus grande tribu dite **tribu discrète**, et $\{\emptyset, \Omega\}$ est la plus petite tribu dite **tribu grossière** ou **tribu triviale**. Un espace de probabilité est dit **discret** si la tribu \mathcal{F} dont il est muni est la tribu discrète. Lorsque Ω est dénombrable, c'est en général le cas (et on peut toujours s'y ramener).

Exemple 0.2. On constate facilement que l'intersection d'une famille de tribus est encore une tribu. En particulier, étant donné un ensemble de sous-ensembles $\mathcal{C} \subset \mathcal{P}(\Omega)$, l'intersection de toutes les tribus contenant \mathcal{C} est une tribu. C'est la plus petite tribu sur Ω contenant \mathcal{C} , appelée **tribu engendrée par \mathcal{C}** et notée $\sigma(\mathcal{C})$. Sur \mathbb{R}^n , la **tribu des boréliens** notée $\mathcal{B}(\mathbb{R}^n)$ est la tribu engendrée par les ouverts (elle est aussi engendrée par les fermés, par les pavés ouverts, ou encore par les pavés fermés).

• **Une probabilité \mathbb{P} sur \mathcal{F}** . Elle associe à chaque événement $A \in \mathcal{F}$ la proportion de chance correspondant à sa réalisation, c'est-à-dire aussi la proportion asymptotique de fois où le résultat de l'expérience réalise l'événement A lorsque l'on répète cette expérience indéfiniment. On rappelle qu'une **probabilité** \mathbb{P} sur (Ω, \mathcal{F}) est une mesure (positive) de masse totale égale à 1. C'est donc une application

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1],$$

telle que $\mathbb{P}(\Omega) = 1$ et qui est σ -additive : pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements *disjoints*,

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

Une propriété sur les éléments de Ω est vraie **presque sûrement** par rapport à la probabilité \mathbb{P} (ce que l'on écrit en abrégé : " \mathbb{P} -p.s.") si elle est vérifiée pour tout $\omega \in \Omega \setminus N$ où N est un certain ensemble tel que $\mathbb{P}(N) = 0$, autrement dit si elle est vérifiée pour tout $\omega \in \tilde{\Omega}$ où $\tilde{\Omega}$ vérifie $\mathbb{P}(\tilde{\Omega}) = 1$.

Rappelons deux résultats essentiels pour le maniement des probabilités :

Proposition 0.3. *a) Pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'événements (c.-à-d. telle que $A_n \subset A_{n+1}$ pour tout n), la suite $(\mathbb{P}(A_n))_{n \in \mathbb{N}}$ est croissante, et*

$$\mathbb{P}\left(\bigcup_n A_n\right) = \lim_n \mathbb{P}(A_n).$$

b) Pour toute suite décroissante $(A_n)_{n \in \mathbb{N}}$ d'événements (c.-à-d. telle que $A_{n+1} \subset A_n$ pour tout n), la suite $(\mathbb{P}(A_n))_{n \in \mathbb{N}}$ est décroissante, et

$$\mathbb{P}\left(\bigcap_n A_n\right) = \lim_n \mathbb{P}(A_n).$$

Définition 0.4. Soit (E, \mathcal{E}) un espace mesurable. Une **variable aléatoire (v.a.) à valeurs dans E** est une application $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ mesurable, c'est-à-dire telle que

$$\forall B \in \mathcal{E}, \quad \{X \in B\} \in \mathcal{F},$$

où on utilise la notation pratique

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B).$$

On parle de **variable aléatoire (v.a.) réelle** pour une v.a. à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et de **vecteur aléatoire réel** pour une v.a. à valeurs dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, avec $n \geq 2$.

Définition 0.5. La **loi** d'une variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ est la probabilité \mathbb{P}_X définie sur (E, \mathcal{E}) par :

$$\forall B \in \mathcal{E}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Au besoin, on peut préciser qu'il s'agit de la **loi de X sous \mathbb{P}** .

Pour une variable aléatoire réelle X , la loi de X est donc une probabilité sur \mathbb{R} .

On distingue deux cas particuliers importants.

• Si X ne prend qu'un ensemble *dénombrable* de valeurs dans E , c.-à-d. s'il existe $I \subset E$ dénombrable tel que $X \in I$ p.s. (c'est le cas si X est une v.a. réelle dont les valeurs sont entières, ou rationnelles), on dit que la loi de X est **discrète** sur E . En introduisant, pour tout $x \in E$, la mesure de Dirac δ_x en x , et en posant, pour tout $x \in E$, $p_x = \mathbb{P}(X = x)$ ($= 0$ si $x \notin I$), on voit que la loi de la v.a. discrète X s'écrit $\mathbb{P}_X = \sum_{x \in I} p_x \delta_x$. Pour tout $B \subset I$ (et $B \subset E$), on a en effet :

$$\mathbb{P}(X \in B) = \sum_{x \in B} p_x.$$

- Un autre cas important est celui où \mathbb{P}_X admet une densité f par rapport à la mesure de Lebesgue dx sur \mathbb{R}^d . On note alors $\mathbb{P}_X = f(x)dx$ et on dit que la loi de X est **absolument continue** ou **admet une densité** sur \mathbb{R}^d . Dans ce cas pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$:

$$\mathbb{P}_X(B) = \int_B f(x)dx.$$

Le lemme suivant s'avèrera souvent utile.

Lemme 0.6 (Lemme d'unicité). *Dans (Ω, \mathcal{F}) considérons une classe \mathcal{C} de parties de \mathcal{F} , stable par intersections finies. Si \mathbf{P} et \mathbf{Q} sont deux probabilités telles que*

$$\mathbf{P}(C) = \mathbf{Q}(C), \forall C \in \mathcal{C},$$

alors

$$\mathbf{P}(A) = \mathbf{Q}(A), \forall A \in \sigma(\mathcal{C}).$$

Par exemple, supposons que X et Y sont des v.a. à valeurs dans \mathbb{R}^d telles que, pour tout pavé fermé C , $\mathbb{P}(X \in C) = \mathbb{P}(Y \in C)$. Alors leurs lois \mathbb{P}_X et \mathbb{P}_Y sont des probabilités sur \mathbb{R}^d qui coïncident sur la famille \mathcal{C} des pavés fermés de \mathbb{R}^d . Or \mathcal{C} est stable par intersections finies et engendre $\mathcal{B}(\mathbb{R}^d)$, donc $\mathbb{P}_X = \mathbb{P}_Y$: X et Y ont même loi.

Définition 0.7. *Pour une v.a. X réelle positive, ou à valeurs dans \mathbb{R}^d et intégrable pour la probabilité \mathbb{P} , on définit son **espérance** comme son intégrale sur Ω par rapport à \mathbb{P} :*

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

Dire que la v.a. X est intégrable correspond donc à dire que $\mathbb{E}(|X|) < \infty$.

Pour une v.a. X à valeurs dans \mathbb{R}^d , pour toute fonction borélienne $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ (c.-à-d. mesurable pour les tribus des boréliens), la variable $Y = \varphi(X)$ est intégrable pour la probabilité \mathbb{P} sur Ω si, et seulement si, φ est intégrable pour \mathbb{P}_X sur \mathbb{R}^d , et

$$\mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x)d\mathbb{P}_X(x).$$

C'est le *théorème de transfert*. Si $\varphi \geq 0$, cette formule reste valable sans supposer $\varphi(X)$ intégrable.

Si X est une v.a. discrète à valeurs dans un ensemble dénombrable I , pour toute fonction $\varphi : I \rightarrow \mathbb{R}^n$ on a donc

$$\mathbb{E}(\varphi(X)) = \sum_{x \in I} \varphi(x)\mathbb{P}(X = x),$$

dès lors que $\varphi \geq 0$ ou $\sum_{x \in I} |\varphi(x)|\mathbb{P}(X = x) < \infty$. Et si X est une v.a. sur \mathbb{R}^d ayant pour densité f_X , pour toute fonction borélienne $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^n$ on a

$$\mathbb{E}(\varphi(X)) = \int \varphi(x)f_X(x)dx.$$

dès lors que $\varphi \geq 0$ ou $\int |\varphi(x)|f_X(x)dx < \infty$.

Citons pour mémoire les lois classiques comme : Bernoulli, géométrique, binomiale, Poisson, pour les lois discrètes ; uniformes, exponentielles, normales, pour les lois absolument continues.

Définition 0.8. *Pour une v.a. X sur (Ω, \mathcal{F}) , à valeurs dans (E, \mathcal{E}) , on définit la **tribu engendrée par X** comme la plus petite tribu sur Ω rendant X mesurable. On la note $\sigma(X)$ et on a :*

$$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{E}\} \subset \mathcal{F}.$$

C'est la tribu de tous les événements relatifs à X , c'est-à-dire de la forme $\{X \in B\}$.

Connaître la valeur de X revient à savoir si les événements de la tribu $\sigma(X)$ sont réalisés ou non. De façon générale, on utilisera souvent les sous-tribus de \mathcal{F} pour décrire des ensembles d'informations dont on peut disposer concernant le résultat de l'expérience aléatoire.

Le résultat suivant montre qu'une variable aléatoire Z est $\sigma(X)$ -mesurable si, et seulement si c'est une fonction (mesurable) de X .

Proposition 0.9. *Considérons une application $X : \Omega \rightarrow E$, où (E, \mathcal{E}) est un espace mesurable. Alors une application $Z : \Omega \rightarrow \mathbb{R}$ est $\sigma(X)$ -mesurable si et seulement si il existe une application mesurable $f : E \rightarrow \mathbb{R}$ telle que $Z = f(X)$.*

Preuve : On écrit

$$Z = \lim_{n \rightarrow +\infty} \sum_{k=-\infty}^{+\infty} \frac{k}{2^n} \mathbf{1}_{[k2^{-n}, (k+1)2^{-n}[}(Z).$$

Puisque Z est $\sigma(X)$ -mesurable, pour tous $n \geq 0$ et $k \in \mathbb{Z}$ il existe $A_{n,k} \in \mathcal{E}$ tel que

$$\{Z \in [k2^{-n}, (k+1)2^{-n}[\} = \{X \in A_{n,k}\},$$

c'est à dire tel que $\mathbf{1}_{[k2^{-n}, (k+1)2^{-n}[}(Z) = \mathbf{1}_{A_{n,k}}(X)$. On a alors $Z = f(X)$ en définissant

$$\text{pour tout } x \in E, \quad f(x) = \liminf_{n \rightarrow +\infty} \sum_{k=-\infty}^{+\infty} \frac{k}{2^n} \mathbf{1}_{A_{n,k}}(x),$$

d'où le lemme. □

La relation d'équivalence entre les variables aléatoires définie par l'égalité presque sûre,

$$X = Y \quad \mathbb{P}\text{-p.s.},$$

définit des classes de v.a. sur $(\Omega, \mathcal{F}, \mathbb{P})$. On note $L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $L^2(\Omega, \mathcal{F}, \mathbb{P})$ les espaces vectoriels des classes de v.a. réelles intégrables et de carré intégrable pour la probabilité \mathbb{P} . Contrairement aux espaces $L^1(dx)$ et $L^2(dx)$ pour la mesure de Lebesgue sur \mathbb{R}^n , on a toujours

$$L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

De plus $L^1(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Banach pour la norme $\|X\|_1 = \mathbb{E}(|X|)$ et $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert pour le produit scalaire suivant : $\langle X, Y \rangle = \mathbb{E}(XY)$.

Définition 0.10. *Sur $(\Omega, \mathcal{F}, \mathbb{P})$.*

- Deux évènements A et A' sont **indépendants** si

$$\mathbb{P}(A \cap A') = \mathbb{P}(A)\mathbb{P}(A').$$

- Des évènements A_1, \dots, A_n sont **indépendants** si

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}),$$

pour tous $2 \leq k \leq n$ et $1 \leq i_1 < \dots < i_k \leq n$.

- Des tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ contenues dans \mathcal{F} sont **indépendantes** si

$$\mathbb{P}(A_1 \cap A_2 \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \cdots \mathbb{P}(A_n),$$

pour tous $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$.

- Une v.a. X est dite **indépendante de la tribu \mathcal{G}** si $\sigma(X)$ et \mathcal{G} sont indépendantes.
- Des variables aléatoires X_1, \dots, X_n sont **indépendantes** si les tribus $\sigma(X_1), \dots, \sigma(X_n)$ sont indépendantes; autrement dit, dans le cas de v.a. réelles, si

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i),$$

pour tous $B_1 \in \mathcal{B}(\mathbb{R}), \dots, B_n \in \mathcal{B}(\mathbb{R})$.

Pour X_1 à valeurs dans $(E_1, \mathcal{E}_1), \dots, X_n$ à valeurs dans (E_n, \mathcal{E}_n) , l'indépendance de X_1, \dots, X_n équivaut à dire que la loi du vecteur (X_1, \dots, X_n) est la **loi produit** des lois de X_1, \dots, X_n : $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$. En conséquence, par les théorèmes de Fubini, on a :

$$\mathbb{E}[\varphi(X_1, \dots, X_n)] = \int \dots \int \varphi(x_1, \dots, x_n) d\mathbb{P}_{X_1}(x_1) \dots d\mathbb{P}_{X_n}(x_n),$$

dès lors que $\varphi \geq 0$ ou que $\varphi(X_1, \dots, X_n)$ est intégrable, où les intégrales peuvent être calculées dans un ordre quelconque. En particulier, pour toutes fonctions $f_1 : E_1 \rightarrow \mathbb{R}, \dots, f_n : E_n \rightarrow \mathbb{R}$ mesurables,

$$\mathbb{E}[f_1(X_1) \dots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \dots \mathbb{E}[f_n(X_n)],$$

dès lors que ces fonctions sont positives, ou que $f_1(X_1), \dots, f_n(X_n)$ sont intégrables.

Le lemme d'unicité fournit la proposition suivante qui simplifie souvent la vérification de l'indépendance :

Proposition 0.11. a) Soient $\mathcal{C}_1 \subset \mathcal{F}$ et $\mathcal{C}_2 \subset \mathcal{F}$ deux classes stables par intersection finie, engendrant deux tribus \mathcal{A}_1 et \mathcal{A}_2 . Si pour tout $A_1 \in \mathcal{C}_1$ et $A_2 \in \mathcal{C}_2$,

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \mathbb{P}(A_2),$$

alors les tribus \mathcal{A}_1 et \mathcal{A}_2 sont indépendantes.

b) Soient X_1, X_2 deux variables aléatoires, à valeurs dans (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) respectivement. Soient $\mathcal{C}_1 \subset \mathcal{E}_1$ et $\mathcal{C}_2 \subset \mathcal{E}_2$ deux classes stables par intersection finie, engendrant \mathcal{E}_1 et \mathcal{E}_2 respectivement. Si pour tout $A_1 \in \mathcal{C}_1$ et $A_2 \in \mathcal{C}_2$,

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A_2),$$

alors les v.a. X_1 et X_2 sont indépendantes.

Preuve : a) Soit $B \in \mathcal{C}_2$. Si $\mathbb{P}(B) = 0$, l'égalité est évidente car $\mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0$. Supposons $\mathbb{P}(B) > 0$. Alors l'égalité s'écrit $\mathbb{P}(A | B) = \mathbb{P}(A)$. Il résulte alors du lemme d'unicité (avec $\mathbf{Q} = \mathbb{P}(\cdot | B)$) que, pour tout $A \in \mathcal{A}_1$, $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$.

Soit $A \in \mathcal{A}_1$. De même, si $\mathbb{P}(A) > 0$, il faut montrer que $\mathbb{P}(B) = \mathbb{P}(B | A)$ pour tout $B \in \mathcal{A}_2$; or c'est vrai pour tout $B \in \mathcal{C}_2$ par ce qui précède, et le lemme permet de conclure, comme précédemment.

b) On déduit b) de a) et du fait que $\{\{X_i \in A_i\} | A_i \in \mathcal{C}_i\}$ engendre $\sigma(X_i)$, pour $i = 1, 2$. Cette propriété se vérifie ainsi : l'ensemble $\{A_i \in \mathcal{E}_i | \{X_i \in A_i\} \in \sigma(\{\{X_i \in B_i\} | B_i \in \mathcal{C}_i\})\}$ est une tribu contenant \mathcal{C}_i , donc elle contient $\sigma(\mathcal{C}_i) = \mathcal{E}_i$. \square

Par exemple, supposons que X et Y sont des v.a. à valeurs dans \mathbb{R}^d telles que, pour tous pavés fermés A et B , $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$. Alors elles sont indépendantes. De même sur \mathbb{R} pour deux v.a. X et Y qui vérifieraient $\mathbb{P}(X \leq s, Y \leq t) = \mathbb{P}(X \leq s) \mathbb{P}(Y \leq t)$ pour tous $s, t \in \mathbb{R}$.

On étend ces définitions à des familles infinies :

Définition 0.12. Une famille infinie d'événements (resp. de variables aléatoires, resp. de tribus) est **indépendante** si toute sous-famille finie de celle-ci est indépendante.

En général, disposer de l'information qu'un événement A est réalisé modifie la probabilité des événements. Il s'agit alors de la probabilité conditionnelle sachant A :

Définition 0.13. Pour deux événements A et B tels que $\mathbb{P}(A) > 0$, la **probabilité conditionnelle de B sachant A** est

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

Remarquons que $\mathbb{P}(\cdot | A)$ est une probabilité sur (Ω, \mathcal{F}) . On peut donc considérer l'espérance associée, notée $\mathbb{E}(\cdot | A)$.

Proposition 0.14. *Pour toute variable aléatoire X intégrable, et tout événement A tel que $\mathbb{P}(A) > 0$, on a*

$$\mathbb{E}(X | A) = \frac{\mathbb{E}(X \mathbf{1}_A)}{\mathbb{P}(A)}.$$

Preuve : Cette égalité est vraie lorsque $X = \mathbf{1}_B$ par la définition même de la probabilité conditionnelle, et s'étend donc par linéarité aux fonctions étagées, puis par convergence monotone (resp. dominée) aux fonctions positives (resp. intégrables). \square

On peut aussi considérer la loi d'une variable aléatoire X sous $\mathbb{P}(\cdot | A)$ (on parlera de **loi conditionnelle sachant A**), ou dire que deux variables aléatoires X et Y sont indépendantes sous $\mathbb{P}(\cdot | A)$ (on parlera d'**indépendance conditionnelle sachant A**). Ces notions joueront un rôle important.

Durant ce cours, on donnera un sens à ces définitions dans certains cas où $\mathbb{P}(A) = 0$ et aussi dans des cas de conditionnement par une tribu.

0.2 Processus stochastiques

Comme il a été dit au début, les *processus stochastiques* permettent de représenter l'évolution dans le temps de phénomènes aléatoires. Le mouvement d'une particule dans l'espace, la transmission d'un signal, le passage dans le temps d'un système à différents états, la variation des cours d'actifs sur un marché, etc., en sont des exemples.

Dans la pratique, on se fixe un ensemble $T \subset \mathbb{R}_+$ d'instantanés d'observation du phénomène étudié.

1. Si T est un intervalle $[a, b]$ on dit que l'étude se fait en **temps continu**.
2. Si T est formé d'une suite d'observations $t_0 = 0 < t_1 < t_2 < \dots < t_n < \dots$, on dit que l'étude se fait en **temps discret**.

Le principe est alors le suivant : pour chaque temps t de T , le phénomène étudié est à un état aléatoire représenté par une v.a. X_t prenant ses valeurs parmi tous les états a priori possibles pour le phénomène. Le processus stochastique correspondant est alors défini par la donnée de la famille de v.a. $(X_t)_{t \in T}$. Autrement dit :

Définition 0.15. *Un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une partie T de \mathbb{R}_+ étant donnés, un **processus stochastique** (ou **processus aléatoire**, ou simplement **processus**) sur l'espace des états $E \subset \mathbb{R}^n$ est une famille $X = (X_t)_{t \in T}$ de v.a.*

$$X_t : \Omega \rightarrow E, \quad t \in T.$$

Pour tout $\omega \in \Omega$ fixé, l'application

$$t \mapsto X_t(\omega)$$

s'appelle une **trajectoire** du processus.

Nous n'étudierons ici que des processus à temps discret (cas 2. pour T) ; on se limite alors à prendre $T \subset \mathbb{N}$; on voit que dans ce cas un processus stochastique revient à la donnée d'une suite de v.a..

On rappelle que l'espace $E^{\mathbb{N}}$ des suites à valeurs dans un espace mesurable (E, \mathcal{E}) peut être muni de la **tribu produit**, $\mathcal{E}^{\otimes \mathbb{N}}$, qui est la tribu engendrée par l'ensemble des **cylindres**

$$B_1 \times \dots \times B_k \times E^{\mathbb{N}} \subset E^{\mathbb{N}}, \quad \text{pour } B_1, \dots, B_k \in \mathcal{E}, \quad k \in \mathbb{N}.$$

Proposition 0.16. *Soit $X = (X_n)_{n \in \mathbb{N}}$ un processus stochastique à temps discret à valeurs dans E .*

- *Alors $X : (\Omega, \mathcal{F}) \rightarrow (E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ est une variable aléatoire. On peut donc considérer sa loi.*
- *La loi de X est caractérisée par la donnée, pour tout $n \geq 0$, de la loi de (X_0, \dots, X_n) .*

Le second point vient du lemme d'unicité 0.6 appliqué aux cylindres.

Si E est dénombrable, se donner la loi de X revient donc à connaître, pour tout $n \geq 0$, pour tous $x_0, \dots, x_n \in E$, la probabilité élémentaire

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n),$$

et si ces probabilités sont les mêmes pour deux processus $X = (X_n)_n$ et $Y = (Y_n)_n$, alors les probabilités de tous les événements relatifs à ces processus seront donc égales.

0.3 Deux résultats fondamentaux

On rappelle une abréviation extrêmement courante :

Définition 0.17. Une famille $(X_i)_{i \in I}$ de v.a. est dite **i.i.d.** (pour *indépendante et identiquement distribuée*) si ces variables sont indépendantes et ont toutes la même loi.

Théorème 0.18 (Loi (forte) des grands nombres). Soit $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d., intégrables à valeurs dans \mathbb{R}^d , ayant même loi qu'une v.a. X . On a

$$\lim_{n \rightarrow \infty} \frac{1}{n} (X_1 + \cdots + X_n) = \mathbb{E}(X), \quad p.s.$$

Rappelons qu'une v.a. réelle X est dite **gaussienne** si sa densité est de la forme

$$x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}, \quad \forall x \in \mathbb{R}.$$

Dans ce cas, $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$ et on note $X \sim \mathcal{N}(m, \sigma^2)$. La fonction caractéristique déterminant de façon unique la loi, $X \sim \mathcal{N}(m, \sigma^2)$ si et seulement si

$$\mathbb{E}e^{itX} = e^{itm - \sigma^2 t^2/2}, \quad \forall t \in \mathbb{R}.$$

Un vecteur (X_1, \dots, X_n) est dit **gaussien** si toute combinaison linéaire $\sum_{j=1}^n \lambda_j X_j$, $\lambda_j \in \mathbb{R}$, est une v.a. réelle gaussienne. On peut montrer que la loi d'un tel vecteur est caractérisée par son espérance $m = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)$ et sa matrice de covariance $\Gamma = (\Gamma_{kl})_{1 \leq k, l \leq n}$ avec $\Gamma_{k,l} = \text{cov}(X_k, X_l)$; sa loi est alors notée $\mathcal{N}(m, \Gamma)$.

Théorème 0.19 (Théorème Central Limite). Soit $(X_n)_n$ une suite de v.a. i.i.d., de carré intégrable, à valeurs dans \mathbb{R}^d , ayant même loi qu'une v.a. X . Alors quand $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \xrightarrow{\text{(loi)}} \mathcal{N}(0, \Gamma),$$

où $\mathcal{N}(0, \Gamma)$ est une loi gaussienne sur \mathbb{R}^d , centrée et de matrice de variance-covariance Γ donnée par celle de X :

$$\Gamma_{k,l} = \text{Cov}(X^k, X^l), \quad 1 \leq k, l \leq d,$$

en ayant noté (X^1, \dots, X^d) les composantes de X .

Chapitre 1

Chaînes de Markov sur un espace d'états discret

Intuitivement, les processus de Markov sont des processus stochastiques dont, à chaque instant, le comportement futur n'est influencé que par la valeur présente et par toutes les valeurs passées. Les *chaînes* de Markov sont des processus de Markov à temps discret. Ces processus permettent de modéliser de nombreux phénomènes ; ils sont par exemple utilisés pour des problèmes de télécommunication, de théorie du signal, d'imagerie informatique, etc.

Nous nous limiterons ici à des chaînes ayant un espace d'états E dénombrable (souvent, on aura $E \subset \mathbb{Z}$). Cet espace sera muni de la tribu discrète $\mathcal{P}(E)$, et la loi d'une variable aléatoire X à valeurs dans E sera donc donnée par les probabilités élémentaires $\mathbb{P}(X = x)$ pour tout x de E .

Soit $(X_n, n \geq 0)$ un processus stochastique à valeurs dans E . Les éléments de E sont souvent notés par i, j, k ou x, y, z .

1.1 Définition

Définition 1.1. *Le processus $(X_n, n \geq 0)$ est appelé une **chaîne de Markov** s'il existe une famille $(P(x, y))_{x, y \in E}$ de réels, appelés **probabilités de transition**, telle que, pour tout $n \geq 0$, et tous $x_0, \dots, x_n, x_{n+1} \in E$,*

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

dès que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$.

Cette définition équivaut à vérifier les deux propriétés suivantes simultanément :

(i) Pour tout $n \geq 1$, et tous $x_0, \dots, x_{n+1} \in E$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n), \quad (1.1)$$

dès que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$

(ii) pour tous $x, y \in E$, la probabilité $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ ne dépend pas du temps n , tel que $\mathbb{P}(X_n = x) > 0$. (On peut donc la noter $P(x, y)$)

La propriété (i) est appelée la **propriété de Markov**. Elle exprime le fait que la loi de X_{n+1} ne dépend de X_0, \dots, X_n qu'à travers la valeur de X_n : le "présent" (X_n) donne autant d'information sur le "futur" (X_{n+1}) que si l'on connaissait tout le "passé" (X_0, \dots, X_n).

La propriété (ii) exprime l'**homogénéité** en temps du processus : la probabilité de transition de l'état x à l'état y est constante au cours du temps.

La famille $(P(x, y))_{x, y \in E}$ de la définition est appelée **noyau de transition** ou **matrice de transition** : si E est fini, on peut en effet numéroter les états $E = \{1, \dots, r\}$ de telle sorte que $(P(x, y))_{1 \leq x, y \leq r}$ est une matrice carrée. On notera indifféremment $P(x, y)$ ou $P_{x, y}$.

Dans la suite de ce chapitre, sauf mention contraire, $X = (X_n)_{n \geq 0}$ désignera une chaîne de Markov ayant une matrice de transition notée P .

Soit $x \in E$ pour lequel il existe $n \geq 0$ tel que $\mathbb{P}(X_n = x) > 0$. D'après la définition, la ligne x de la matrice P , c'est-à-dire la famille $(P(x, y))_{y \in E}$, donne la loi de X_{n+1} sachant $X_n = x$, donc elle vérifie

$$P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x) \geq 0, \quad \forall y \in E,$$

et

$$\sum_{y \in E} P(x, y) = 1.$$

Définition 1.2. Une matrice $P = (p_{x,y})_{x,y \in E}$ vérifiant

$$p_{x,y} \geq 0, \quad \forall x, y \in E \quad \text{et} \quad \sum_{y \in E} p_{x,y} = 1, \quad \forall x \in E$$

s'appelle une matrice **stochastique**.

De manière intuitive la probabilité de transition indique comment la chaîne "tourne", et il reste à dire comment elle "démarré". C'est le rôle de la loi initiale.

Définition 1.3. La loi **initiale** de la chaîne de Markov $(X_n)_{n \geq 0}$ est la loi de X_0 .

Vu que E est dénombrable, la loi initiale de $(X_n)_{n \geq 0}$ est donc donnée par la famille $\nu = (\nu(x))_{x \in E}$ où $\nu(x) = \mathbb{P}(X_0 = x)$, $x \in E$. On identifiera souvent la loi initiale avec cette famille.

Le lemme suivant montre que le noyau de transition et la loi initiale déterminent complètement la loi du processus.

Lemme 1.4. Un processus $(X_n)_{n \geq 0}$ à valeurs dans E est une chaîne de Markov de matrice de transition $P = (P(x, y))_{x,y \in E}$ et de loi initiale $\nu = (\nu(x), x \in E)$ si, et seulement si pour tous $x_0, \dots, x_n \in E$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \nu(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Preuve : Le sens "chaîne de Markov" \Rightarrow "la formule" s'obtient par récurrence sur n , laissée comme exercice. La réciproque s'en déduit car ces probabilités caractérisent la loi du processus. \square

En raison de cette propriété, on pourra donc parler sans ambiguïté de la loi de la chaîne de Markov de loi initiale ν et de matrice de transition P , sans nécessairement préciser la façon dont ce processus est construit. Il sera constamment utile dans la suite du cours de conserver la même matrice de transition mais de faire varier la loi initiale de la chaîne de Markov :

Définition 1.5. Supposons qu'une matrice stochastique $P = (P(x, y))_{x,y \in E}$ est donnée. Pour toute loi ν sur E , on notera \mathbb{P}_ν une probabilité sur Ω telle que, sous \mathbb{P}_ν (c'est-à-dire sur l'espace de probabilités (Ω, \mathbb{P}_ν)), $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale ν .

Cette définition suppose admise la propriété suivante d'existence (pour des détails, voir la section 1.12 de complément sur le processus canonique) :

Proposition 1.6. Soit P une matrice stochastique sur E . Il existe un espace mesurable (Ω, \mathcal{F}) , et un processus $(X_n)_{n \geq 0}$ à valeurs dans E tel que, pour toute probabilité ν sur E , il existe une probabilité \mathbb{P}_ν sur Ω sous laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov ayant pour loi initiale ν et pour matrice de transition P .

Pour $x \in E$, on notera \mathbb{P}_x au lieu de \mathbb{P}_{δ_x} la loi de la chaîne de Markov de matrice P et de loi initiale δ_x , c'est-à-dire telle que $X_0 = x$ p.s.. On dira que c'est la chaîne de Markov **issue de** x .

Avec ces notations, on vérifie par exemple que conditionner par $X_0 = x$ revient à considérer \mathbb{P}_x :

Lemme 1.7. Soit ν une loi sur E . Pour tout $x \in E$ tel que $\nu(x) > 0$, pour tout événement A dépendant de X_0, X_1, \dots ,

$$\mathbb{P}_\nu(A | X_0 = x) = \mathbb{P}_x(A).$$

Pour tout événement A dépendant de X_0, X_1, \dots ,

$$\mathbb{P}_\nu(A) = \sum_{x \in E} \mathbb{P}_x(A) \nu(x),$$

Preuve : Il suffit de traiter le cas où $A = \{X_0 = x_0, \dots, X_n = x_n\}$, pour tous $x_0, \dots, x_n \in E$ (par le Lemme d'unicité 0.6). On a

$$\begin{aligned} \mathbb{P}_\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n | X_0 = x) &= \frac{\delta_x(x_0) \mathbb{P}_\nu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)}{\mathbb{P}_\nu(X_0 = x)} \\ &= \frac{\delta_x(x_0) \nu(x) P(x, x_1) \cdots P(x_{n-1}, x_n)}{\nu(x)} \\ &= \delta_x(x_0) P(x, x_1) \cdots P(x_{n-1}, x_n) \\ &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n). \end{aligned}$$

Ceci implique la première égalité. La seconde égalité du lemme se déduit alors de la première et de la formule des probabilités totales. \square

Dans les cas simples, on pourra représenter graphiquement les transitions possibles en 1 pas (c.-à-d. entre X_n et X_{n+1}) sous la forme d'un graphe orienté.

Définition 1.8. Le **graphe** de la chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'états E et de matrice de transition $P = (P(x, y))_{x, y \in E}$ est le graphe orienté $\mathcal{G} = (S, A)$ dont l'ensemble des sommets est $S = E$ et l'ensemble des arêtes est

$$A = \{(x, y) \in E \times E \mid P(x, y) > 0\}.$$

1.2 Probabilités de transition à m pas

On a vu que la matrice P décrit les probabilités de transition "à 1 pas", c'est-à-dire pour passer de X_0 à X_1 , ou plus généralement de X_n à X_{n+1} . En effet la ligne x de P donne la loi de X_1 sachant que $X_0 = x$:

$$P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x) = \mathbb{P}_x(X_1 = y).$$

Pour étendre ceci au calcul des transitions de X_0 à X_m , on introduit une notion de produit de noyaux :

Définition 1.9. Soit P, Q deux noyaux de transition sur E . Leur **produit** PQ est le noyau de transition sur E défini par :

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

Dans le cas où $E = \{1, \dots, r\}$, P et Q sont des matrices carrées et cela correspond au produit matriciel. On définit de même par récurrence $P^n = P^{n-1}P$, avec $P^0 = \text{Id}_E$.

Proposition 1.10. Soit $m \in \mathbb{N}$. Pour tous $x, y \in E$,

$$\mathbb{P}_x(X_m = y) = P^m(x, y).$$

Ainsi, la loi de X_m sous \mathbb{P}_x est donnée par la ligne x de P^m .

Preuve : On procède par récurrence sur m . Pour $m = 0$, la propriété est triviale. Soit $m \in \mathbb{N}$. On suppose la propriété vraie pour m . Pour tous $x, y \in E$, pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}_x(X_{m+1} = y) &= \sum_{z \in E} \mathbb{P}_x(X_{m+1} = y \mid X_m = z) \mathbb{P}_x(X_m = z) \\ &= \sum_{z \in E} P(z, y) P^m(x, z) \end{aligned}$$

d'après la définition de P et l'hypothèse de récurrence. Ceci se réécrit $\mathbb{P}_x(X_{m+1} = y) = P^m P(x, y) = P^{m+1}(x, y)$, ce qui conclut. \square

La notion de produit s'étend aux vecteurs comme dans le cas usuel des matrices :

Définition 1.11. Soit P un noyau de transition sur E , ν une probabilité sur E et $f : E \rightarrow \mathbb{R}^+$ une fonction. La probabilité νP est la probabilité définie par

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

La fonction $Pf : E \rightarrow \mathbb{R}^+$ est définie par

$$Pf(x) = \sum_{y \in E} P(x, y) f(y).$$

Le réel νf est défini par

$$\nu f = \sum_{y \in E} \nu(y) f(y) = \int_E f d\nu.$$

Dans le cas où $E = \{1, \dots, r\}$, νP est le produit du vecteur-ligne ν par la matrice P , Pf est le produit de la matrice P par le vecteur-colonne f , et de même pour νf .

Corollaire 1.12. Soit $m \in \mathbb{N}$. Pour toute loi ν sur E , la loi de X_m sous \mathbb{P}_ν est νP^m . Ainsi, pour toute fonction $f : E \rightarrow \mathbb{R}_+$,

$$\mathbb{E}_\nu(f(X_m)) = \nu P^m f,$$

et en particulier pour tout $x \in E$,

$$\mathbb{E}_x(f(X_m)) = P^m f(x).$$

1.3 Exemples de chaînes de Markov

1.3.1 Chaîne de markov à deux états

Soit E un ensemble à deux éléments, notés par exemple $E = \{0, 1\}$: 0 peut représenter un système à l'arrêt et 1 le même système en fonctionnement. Soit $X = (X_n, n \geq 0)$ une chaîne de Markov sur E (ainsi, X_n modélise par exemple l'état de ce système après n jours de fonctionnement). En notant

$$\alpha = \mathbb{P}(X_{n+1} = 1 \mid X_n = 0), \quad \beta = \mathbb{P}(X_{n+1} = 0 \mid X_n = 1),$$

la matrice de transition de X est alors donnée par

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

À titre d'exercice, on peut vérifier (par récurrence en n) le résultat suivant :

Proposition 1.13. On a

$$P^n = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$

On peut en déduire (exercice) la limite de P^n quand $n \rightarrow \infty$; on notera alors que le cas $\alpha = \beta = 1$ est particulier.

1.3.2 Marche aléatoire sur \mathbb{Z} et sur \mathbb{Z}^d

Un exemple important est la notion de marche aléatoire sur \mathbb{Z} , qui est une chaîne de Markov sur \mathbb{Z} vérifiant de plus une certaine *homogénéité spatiale* : sachant $X_k = x$, la loi du processus $(X_{k+n} - x)_{n \geq 0}$ ne dépend ni de k (homogénéité temporelle), ni de x (homogénéité spatiale). Ceci revient à la définition suivante. Soit $q = (q(x))_{x \in \mathbb{Z}}$ une probabilité sur \mathbb{Z} . Soit $(\xi_n, n \geq 1)$ une suite de v.a. i.i.d. de loi commune q :

$$\mathbb{P}(\xi_n = x) = q(x), \quad \forall n \geq 1, \forall x \in \mathbb{Z},$$

et soit x_0 un élément de \mathbb{Z} . On définit la **marche aléatoire** partant de x_0 comme le processus $X = (X_n)_{n \geq 0}$ avec :

$$X_n = x_0 + \sum_{i=1}^n \xi_i, \quad \forall n \geq 0.$$

C'est une chaîne de Markov sur \mathbb{Z} du fait de l'indépendance de la suite des ξ_i qui représentent les pas de la marche (exercice : vérifier que c'est une chaîne de Markov). On vérifie facilement que sa probabilité de transition est

$$P(a, b) = q(b - a), \quad \forall a, b \in \mathbb{Z}.$$

Un exemple classique est la **marche aléatoire simple sur \mathbb{Z}** , ou “marche de l'ivrogne” : pour $p \in]0, 1[$ fixé, $q(+1) = p$ et $q(-1) = 1 - p$, autrement dit

$$\mathbb{P}(\xi_n = 1) = 1 - \mathbb{P}(\xi_n = -1) = p.$$

(À chaque instant, on fait un pas à droite avec la probabilité p ou un pas à gauche avec la probabilité $1 - p$). On vérifiera facilement que la probabilité de transition s'écrit :

$$P(z, \cdot) = p\delta_{z+1} + (1 - p)\delta_{z-1}, \quad \forall z \in \mathbb{Z}.$$

X_n décrit par exemple la fortune accumulée après n parties d'un jeu de hasard qui rapporte un euro en cas de gain, et fait perdre un euro sinon, en partant d'une fortune de x_0 euros.

Selon le même principe d'homogénéité spatiale, on peut considérer des marches aléatoires sur \mathbb{Z}^d , définies par $X_n = x_0 + \xi_1 + \dots + \xi_n$ où ξ_1, ξ_2, \dots sont des variables aléatoires indépendantes et suivant la même loi sur \mathbb{Z}^d . L'exemple le plus courant est la **marche aléatoire simple symétrique sur \mathbb{Z}^d** :

$$\mathbb{P}(\xi_n = e_i) = \mathbb{P}(\xi_n = -e_i) = \frac{1}{2d} \quad \text{pour } i = 1, \dots, d,$$

où (e_1, \dots, e_d) est la base canonique de \mathbb{R}^d . Autrement dit, à chaque pas, la marche choisit uniformément l'une des $2d$ directions.

1.3.3 Marches aléatoires simples sur \mathbb{Z} avec barrières

Soit $p \in [0, 1]$, et $a < x_0 < b$ des entiers. On peut définir plusieurs processus “restrictions” à $\{a, \dots, b\}$ de la marche aléatoire simple issue de x_0 , selon le choix des conditions au bord.

Considérons une matrice de transition P sur $E = \{a, \dots, b\}$ donnée par

$$P(x, x - 1) = 1 - p \quad \text{et} \quad P(x, x + 1) = p, \quad \forall x \in \{a + 1, \dots, b - 1\}$$

et par

$$P(a, a) = 1 \quad P(b, b) = 1.$$

Ainsi, si $X_n = a$, alors $X_{n+k} = a$ p.s. pour tout $k \geq 0$, par récurrence, et de même pour b . On parle de **barrières absorbantes**.

Alternativement, on peut définir $P(x, y)$ de la même façon quand $x \neq a, b$, et par

$$P(a, a + 1) = 1 \quad P(b, b - 1) = 1.$$

Ainsi, si $X_n = a$, alors $X_{n+1} = a + 1$ p.s., et si $X_n = b$, alors $X_{n+1} = b - 1$ p.s.. On parle de **barrières réfléchissantes**.

On peut bien sûr considérer des modèles où a est réfléchissante, et b absorbante, et d'autres notions de conditions au bord (par exemple périodique en b : $P(b, a) = p = 1 - P(b, b - 1)$).

1.3.4 Modèle d'Ehrenfest

Ce modèle a été introduit par le couple Ehrenfest pour l'expérience suivante : N molécules de gaz sont réparties dans deux récipients A et B séparés par une cloison percée, et on observe l'évolution de la quantité de gaz dans le récipient A . Dans ce modèle, on représente les déplacements de molécules entre récipients de la façon suivante : après un temps fixé, une molécule choisie au hasard (parmi l'ensemble des molécules) change de récipient, et ceci se répète indéfiniment.

Notons X_n le nombre de molécules dans A après n échanges. Alors $(X_n)_n$ est une chaîne de Markov à valeurs dans $E = \{0, \dots, N\}$ dont le noyau de transition est donné par $p_{0,1} = 1 = p_{N,N-1}$ et, pour $1 \leq i \leq N-1$ et $0 \leq j \leq N$,

$$p_{i,j} = \begin{cases} i/N, & \text{si } j = i - 1; \\ (N - i)/N, & \text{si } j = i + 1; \\ 0, & \text{sinon.} \end{cases}$$

1.3.5 Processus du renouvellement

On examine le fonctionnement d'une machine à temps discrets $n = 0, 1, \dots$. Si la machine tombe en panne, on la remplace immédiatement par une nouvelle. Supposons que les durées de vie des machines (Y_1, Y_2, \dots) sont i.i.d. et de même loi qu'une v.a. Y à valeurs dans \mathbb{N}^* . On s'intéresse à X_n , l'âge de la machine qui fonctionne au temps n (avec $X_n = 0$ si une machine est tombée en panne au temps n). Alors en posant

$$Z_k := \sum_{i=1}^k Y_i$$

et

$$\ell_n := \max\{k : Z_k \leq n\},$$

on constate que l'on a

$$X_n = n - Z_{\ell_n}.$$

Proposition 1.14. *Le processus $(X_n)_{n \geq 0}$ est une chaîne de Markov à valeurs dans \mathbb{N} et de probabilité de transition P de la forme*

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & 0 & 0 & 0 & 0 & \dots \\ p_{1,0} & 0 & p_{1,2} & 0 & 0 & 0 & \dots \\ p_{2,0} & 0 & 0 & p_{2,3} & 0 & 0 & \dots \\ p_{3,0} & 0 & 0 & 0 & p_{3,4} & 0 & \dots \\ \dots & 0 & 0 & 0 & \dots & \dots & \dots \\ \dots & 0 & 0 & \dots & \dots & \dots & \dots \end{pmatrix},$$

avec

$$p_{i,0} = \mathbb{P}(Y = i + 1 \mid Y > i), \quad p_{i,i+1} = \mathbb{P}(Y > i + 1 \mid Y > i).$$

1.3.6 Processus de branchement

Nous étudions l'évolution d'une population. La population originelle s'appelle la génération 0. Les enfants de la génération $n - 1$ constituent la génération n pour tout $n \geq 1$.

On note X_n le nombre d'individus de la n -ième génération.

On suppose que le nombre d'enfants de différents individus sont des variables aléatoires indépendantes et ont tous la même loi qu'une v.a. Z à valeurs dans \mathbb{N} . On aura ainsi, pour tout $n \geq 0$,

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

où $Z_{n,i}$ représente le nombre d'enfants du i -ième individu de la génération n , et les v.a. $(Z_{n,i})_{n,i \in \mathbb{N}}$ sont i.i.d. de même loi que Z .

Proposition 1.15. *Le processus $(X_n, n \geq 0)$ est une chaîne de Markov à valeurs dans $E = \mathbb{N}$ et de noyau de transition $P = (p_{i,j})_{i,j \geq 0}$ donné par*

$$p_{i,j} = \begin{cases} \mathbb{P}(Z_1 + \dots + Z_i = j), & \text{si } i > 0 \text{ et } j \geq 0; \\ 1, & \text{si } i = j = 0; \\ 0, & \text{si } i = 0 \text{ et } j > 0, \end{cases}$$

où $(Z_j)_{j \geq 1}$ sont i.i.d. et de même loi que Z .

Un problème intéressant sera de connaître la *probabilité d'extinction* : $\mathbb{P}(\exists n \geq 0 \text{ tel que } X_n = 0)$.

1.4 Propriété de Markov

On souhaite étendre la propriété qui définit les chaînes de Markov à des événements qui concernent tout le “futur” ou tout le “passé” du processus.

On commence par une extension directe de la définition :

Proposition 1.16 (Propriété de Markov simple au temps k). *Soit $k, n \geq 1$. Pour tous $x_0, \dots, x_k, x_{k+1}, \dots, x_{k+n} \in E$,*

$$\mathbb{P}_\nu(X_{k+1} = x_{k+1}, \dots, X_{k+n} = x_{k+n} \mid X_0 = x_0, \dots, X_k = x_k) = \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}).$$

Preuve : Par le Lemme 1.4, la probabilité de gauche s'écrit :

$$\begin{aligned} \frac{\mathbb{P}_\nu(X_0 = x_0, \dots, X_{k+n} = x_{k+n})}{\mathbb{P}_\nu(X_0 = x_0, \dots, X_k = x_k)} &= \frac{\nu(x_0)P(x_0, x_1) \cdots P(x_{k+n-1}, x_{k+n})}{\nu(x_0)P(x_0, x_1) \cdots P(x_{k-1}, x_k)} \\ &= P(x_k, x_{k+1}) \cdots P(x_{k+n-1}, x_{k+n}) \\ &= \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}), \end{aligned}$$

comme annoncé. □

On étend ceci à des événements plus généraux :

Proposition 1.17 (Propriété de Markov simple au temps k). *Soit $k, n \geq 1$. Pour tout état $x \in E$, et tous sous-ensembles $A \subset E^{n+1}$, $B \subset E^{k+1}$,*

$$\mathbb{P}_\nu((X_k, \dots, X_{k+n}) \in A \mid (X_0, \dots, X_k) \in B, X_k = x) = \mathbb{P}_x((X_0, \dots, X_n) \in A),$$

dès lors que l'événement par lequel on conditionne n'a pas une probabilité nulle.

Preuve : La proposition précédente s'écrit aussi, en explicitant la probabilité conditionnelle :

$$\mathbb{P}_\nu(X_0 = x_0, \dots, X_{k+n} = x_{k+n}) = \mathbb{P}_\nu(X_0 = x_0, \dots, X_k = x_k) \mathbb{P}_{x_k}(X_1 = x_{k+1}, \dots, X_n = x_{k+n}),$$

et il suffit alors de prendre $x_k = x$ et de sommer cette identité sur tous les $x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_{k+n}$ tels que $(x_0, \dots, x_{k-1}, x) \in B$ et $(x, x_{k+1}, \dots, x_{k+n}) \in A$. □

La suite de cette partie contient différents énoncés de la propriété de Markov simple au temps k , qui sont simplement différentes formulations de la proposition précédente.

Proposition 1.18. *Soit $k \geq 1$. Pour tout $x \in E$ et tout $B \subset E^{k+1}$, la loi du processus (X_k, X_{k+1}, \dots) sous $\mathbb{P}_\nu(\cdot \mid (X_0, \dots, X_k) \in B, X_k = x)$ est \mathbb{P}_x : c'est une chaîne de Markov de matrice P issue de x .*

On rappelle que les événements de la forme $\{(X_0, \dots, X_k) \in B\}$ forment la tribu $\sigma(X_0, \dots, X_k)$; on prendra l'habitude d'utiliser la notation suivante et, en général, de penser en terme de tribus :

Définition 1.19. *Pour tout $k \geq 0$, la tribu du passé avant le temps k est*

$$\mathcal{F}_k = \sigma(X_0, \dots, X_k).$$

La proposition précédente a pour conséquence directe, en terme d'espérance, les formules suivantes :

Proposition 1.20 (Propriété de Markov au temps k). *Soit $k \geq 1$. Pour tout $x \in E$, pour toute fonction $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$ mesurable positive ou bornée, pour tout événement $H \in \mathcal{F}_k$,*

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots) \mid H \cap \{X_k = x\}) = \mathbb{E}_x(f(X_0, X_1, \dots)).$$

Plus généralement,

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots)\mathbf{1}_H) = \mathbb{E}_\nu(F(X_k)\mathbf{1}_H)$$

où la fonction $F : E \rightarrow \mathbb{R}$ est définie par

$$F(x) = \mathbb{E}_x(f(X_0, X_1, \dots)), \quad x \in E.$$

Plus généralement, pour toute fonction $g : E^k \rightarrow \mathbb{R}$ positive ou bornée, avec la même notation,

$$\mathbb{E}_\nu(f(X_k, X_{k+1}, \dots)g(X_0, \dots, X_k)) = \mathbb{E}_\nu(F(X_k)g(X_0, \dots, X_k)).$$

Preuve : La première égalité vient directement de la dernière proposition : la loi de (X_k, X_{k+1}, \dots) sous $\mathbb{P}_\nu(\cdot \mid H \cap \{X_k = x\})$ est \mathbb{P}_x , donc l'espérance de f sous ces deux lois est la même. En explicitant l'espérance conditionnelle (rappelons que $\mathbb{E}(Z \mid A) = \mathbb{E}(Z\mathbf{1}_A)/\mathbb{P}(A)$), cette égalité devient :

$$\begin{aligned} \mathbb{E}_\nu(f(X_k, X_{k+1}, \dots)\mathbf{1}_H\mathbf{1}_{\{X_k=x\}}) &= \mathbb{E}_x(f(X_0, X_1, \dots))\mathbb{E}_\nu(\mathbf{1}_H\mathbf{1}_{\{X_k=x\}}) \\ &= F(x)\mathbb{E}_\nu(\mathbf{1}_H\mathbf{1}_{\{X_k=x\}}) = \mathbb{E}_\nu(F(x)\mathbf{1}_H\mathbf{1}_{\{X_k=x\}}) \\ &= \mathbb{E}_\nu(F(X_k)\mathbf{1}_H\mathbf{1}_{\{X_k=x\}}). \end{aligned}$$

En sommant cette égalité sur toutes les valeurs $x \in E$, on obtient la seconde égalité de l'énoncé.

Ceci prouve la dernière égalité lorsque $g(X_1, \dots, X_k) = \mathbf{1}_H$. On en déduit le cas général de façon classique : par linéarité on obtient le cas où g est étagée, puis par convergence monotone ou dominée le cas où g est positive ou bornée. \square

On généralise alors facilement le Lemme 1.4 au cas de transitions à plusieurs pas :

Lemme 1.21. *Pour tous états $x_0, \dots, x_r \in E$ et tous temps $0 \leq n_1 \leq \dots \leq n_r$,*

$$\mathbb{P}_{x_0}(X_{n_1} = x_1, \dots, X_{n_r} = x_r) = P^{n_1}(x_0, x_1)P^{n_2-n_1}(x_1, x_2) \cdots P^{n_r-n_{r-1}}(x_{r-1}, x_r),$$

et plus généralement, pour toute loi ν sur E ,

$$\mathbb{P}_\nu(X_{n_1} = x_1, \dots, X_{n_r} = x_r) = (\nu P^{n_1})(x_1)P^{n_2-n_1}(x_1, x_2) \cdots P^{n_r-n_{r-1}}(x_{r-1}, x_r).$$

Preuve : On procède par récurrence sur $r \geq 1$. Pour $r = 1$, c'est la Proposition 1.10. Supposons le résultat vrai pour r . Soit $x_0, \dots, x_{r+1} \in E$ et $0 \leq n_1 \leq \dots \leq n_{r+1}$. On a, d'après la propriété de Markov simple au temps n_1 ,

$$\begin{aligned} \mathbb{P}_{x_0}(X_{n_1} = x_1, \dots, X_{n_{r+1}} = x_{r+1}) &= \mathbb{P}_{x_0}(X_{n_1} = x_1)\mathbb{P}_{x_1}(X_{n_2-n_1} = x_2, \dots, X_{n_{r+1}-n_r} = x_{r+1}) \\ &= P^{n_1}(x_0, x_1)\mathbb{P}_{x_1}(X_{n_2-n_1} = x_2, \dots, X_{n_{r+1}-n_r} = x_{r+1}) \end{aligned}$$

(la seconde égalité vient de la Proposition 1.10), ce qui conclut, par l'hypothèse de récurrence. Le cas d'une loi initiale ν s'en déduit en décomposant selon la valeur de X_0 (ou via le Lemme 1.7). \square

Quitte à construire la chaîne de Markov sur $\Omega = E^{\mathbb{N}}$ (cf. Appendice), muni de la tribu produit, on peut toujours supposer qu'il existe une application $\theta : \Omega \rightarrow \Omega$ mesurable telle que, pour tout $n \in \mathbb{N}$,

$$X_n \circ \theta = X_{n+1}.$$

Cette application s'appelle le **décalage** (ou "shift" en anglais). On posera $\theta_0 = \text{Id}$, $\theta_1 = \theta$, \dots , $\theta_{n+1} = \theta_n \circ \theta$. Alors $X_n \circ \theta_k = X_{n+k}$, pour tous $n, k \geq 0$. Avec ces notations, la propriété de Markov simple s'écrit comme suit :

Proposition 1.22 (Propriété de Markov simple au temps k). *Pour tout $k \geq 1$ et pour toute v.a. réelle Z qui est $\sigma(X)$ -mesurable, positive ou bornée, et pour toute v.a. réelle W qui est \mathcal{F}_k -mesurable, pour toute loi initiale ν , on a*

$$\mathbb{E}_\nu(Z \circ \theta_k W) = \mathbb{E}_\nu(F(X_k)W),$$

où

$$F(x) := \mathbb{E}_x(Z).$$

Preuve : Toute v.a. Z qui est $\sigma(X)$ -mesurable s'écrit en effet $Z = f(X) = f(X_0, X_1, \dots)$ et toute v.a. W qui est \mathcal{F}_k -mesurable est de la forme $W = g(X_0, \dots, X_k)$. Comme $Z \circ \theta_k = f(X \circ \theta_k) = f(X_k, X_{k+1}, \dots)$, cette proposition ré-exprime la dernière formule de la Proposition 1.20. \square

1.5 Visites à un sous-ensemble

Soit $X = (X_n, n \geq 0)$ une chaîne de Markov de probabilité de transition P sur l'espace d'états E . On introduit de nouvelles variables aléatoires, fonctions de X , relatives aux visites à un sous-ensemble A de E .

Définition 1.23. *Soit $A \subset E$. Le **premier temps d'entrée** de la chaîne dans A est la v.a.*

$$T_A := \begin{cases} \inf\{n \geq 0 : X_n \in A\}, & \text{si } \{\dots\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

*Le **premier temps de retour** de la chaîne dans A est la v.a.*

$$\tau_A := \begin{cases} \inf\{n \geq 1 : X_n \in A\}, & \text{si } \{\dots\} \neq \emptyset; \\ \infty, & \text{sinon.} \end{cases}$$

*Le **nombre de visites** de la chaîne à A est la v.a.*

$$N_A = \sum_{n=0}^{\infty} \mathbf{1}_{(X_n \in A)} \in \mathbb{N} \cup \{\infty\}.$$

Lorsque $A = \{x\}$ est un singleton, on note simplement T_x , τ_x et N_x . Remarquons que si $y \neq x$, $\tau_y = T_y$, \mathbb{P}_x -p.s.

Le théorème de convergence monotone pour les séries donne immédiatement la formule suivante pour tous $x, y \in E$:

$$\mathbb{E}_x(N_y) = \sum_{n=0}^{\infty} P^n(x, y).$$

Donnons une application de la propriété de Markov :

Proposition 1.24. *Pour tout couple (x, y) d'éléments de E , on a*

$$\mathbb{E}_x(N_y) = \mathbb{P}_x(T_y < \infty) \mathbb{E}_y(N_y).$$

Preuve : Pour tout entier $k \geq 0$,

$$\begin{aligned} \mathbb{P}_x(N_y = k) &= \sum_{n=0}^{\infty} \mathbb{P}_x(T_y = n, N_y = k) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x\left(T_y = n, \sum_{j=0}^{\infty} \mathbf{1}_{(X_{n+j}=y)} = k\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x\left(X_0 \neq y, \dots, X_{n-1} \neq y, X_n = y, \sum_{j=0}^{\infty} \mathbf{1}_{(X_{n+j}=y)} = k\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(T_y = n) \mathbb{P}_y\left(\sum_{j=0}^{\infty} \mathbf{1}_{(X_j=y)} = k\right) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(T_y = n) \mathbb{P}_y(N_y = k), \end{aligned}$$

où la troisième égalité est due à la propriété de Markov au temps n . Nous avons donc

$$\mathbb{P}_x(N_y = k) = \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(N_y = k),$$

d'où la proposition en sommant $\sum_{k \geq 0} k\mathbb{P}_x(N_y = k)$, et en remarquant que si $\mathbb{P}_x(N_y = \infty) > 0$ alors $\mathbb{P}_y(N_y = \infty) > 0$. \square

Donnons une autre démonstration, qui utilise la propriété de Markov sous forme d'espérance :

Preuve : On a

$$\mathbb{E}_x(N_y) = \sum_{n=0}^{\infty} \mathbb{E}_x(N_y \mathbf{1}_{\{T_y=n\}}),$$

et on note que si $T_y = n$ alors $N_y = N_y \circ \theta_n$ (si la première visite en y a lieu au temps n , alors le nombre total de visites en y est le même que le nombre de visites en y à partir du temps n). Ainsi, pour tout $n \geq 0$, par la propriété de Markov au temps n , vu que $\mathbf{1}_{\{T_y=n\}}$ est \mathcal{F}_n -mesurable,

$$\mathbb{E}_x(N_y \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(N_y \circ \theta_n \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(F(X_n) \mathbf{1}_{\{T_y=n\}}) = \mathbb{E}_x(F(y) \mathbf{1}_{\{T_y=n\}}) = F(y)\mathbb{P}_x(T_y = n),$$

où, pour tout état z , $F(z) = \mathbb{E}_z(N_y)$. En sommant sur $n \in \mathbb{N}$, on obtient

$$\mathbb{E}_x(N_y \mathbf{1}_{\{T_y < \infty\}}) = \mathbb{E}_y(N_y)\mathbb{P}_x(T_y < \infty),$$

et le premier terme est égal à $\mathbb{E}_x(N_y)$ car sur l'événement $\{T_y = \infty\}$ on a $N_y = 0$. \square

Définition 1.25. La *fonction de Green* de la chaîne de Markov X est la fonction $G : E \times E \rightarrow \mathbb{R} \cup \{\infty\}$ donnée par : pour tous $x, y \in E$,

$$G(x, y) = \mathbb{E}_x(N_y) = \sum_{n=0}^{+\infty} P^n(x, y) = \sum_{n=0}^{+\infty} \mathbb{E}_x(\mathbf{1}_y(X_n)) = \mathbb{E}_x\left(\sum_{n=0}^{+\infty} \mathbf{1}_y(X_n)\right)$$

Avec cette notation, la proposition précédente s'écrit $G(x, y) = \mathbb{P}_x(T_y < \infty)G(y, y)$, d'où en particulier

$$G(x, y) \leq G(y, y).$$

1.6 Récurrence et transience

Nous allons étudier le comportement asymptotique des trajectoires d'une chaîne de Markov $(X_n)_n$. On va ici essentiellement distinguer deux types de comportement suivant que la chaîne "tend vers l'infini" (cas transient) ou non (cas récurrent).

Définition 1.26. Un état x est **récurrent** si $\mathbb{P}_x(\tau_x < +\infty) = 1$. Dans le cas contraire, x est dit **transient**.

Théorème 1.27. Si pour la chaîne X l'état x est transient, le nombre de visites N_x suit, sous la probabilité \mathbb{P}_x , la loi géométrique de paramètre

$$a = 1 - \mathbb{P}_x(\tau_x < \infty) = \mathbb{P}_x(\tau_x = \infty),$$

c'est-à-dire que, pour tout $k \geq 1$,

$$\mathbb{P}_x(N_x = k) = a(1-a)^{k-1}, \quad k \geq 1.$$

En particulier, $\mathbb{P}_x(N_x < \infty) = 1$ et $G(x, x) = \mathbb{E}_x(N_x) = 1/a$.

Preuve : Soit x transient. Sous \mathbb{P}_x , $X_0 = x$ et $\sum_{j=0}^{\tau_x-1} \mathbf{1}_{(X_j=x)} = 1$. Il vient, pour tout $k \geq 1$,

$$\mathbb{P}_x(N_x = k + 1) = \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x = n, N_x \circ \theta_n = k),$$

où $N_x \circ \theta_n = \sum_{j=0}^{\infty} \mathbf{1}_{(X_{j+n}=x)}$ est le nombre de visite en x par la chaîne après le temps n . D'après la propriété de Markov au temps n (Proposition 1.22), on a $\mathbb{P}_x(\tau_x = n, N_x \circ \theta_n = k) = \mathbb{E}_x(\mathbf{1}_{\{\tau_x=n\}} \mathbf{1}_{\{N_x \circ \theta_n=k\}}) = \mathbb{E}_x(\mathbf{1}_{\{\tau_x=n\}} \mathbb{E}_{X_n}(\mathbf{1}_{\{N_x=k\}})) = \mathbb{P}_x(\tau_x = n) \mathbb{P}_x(N_x = k)$ d'où

$$\mathbb{P}_x(N_x = k + 1) = \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x = n) \mathbb{P}_x(N_x = k) = \mathbb{P}_x(\tau_x < \infty) \mathbb{P}_x(N_x = k).$$

Par récurrence en k , on obtient

$$\mathbb{P}_x(N_x = k + 1) = \mathbb{P}_x(\tau_x < \infty)^k \mathbb{P}_x(N_x = 1).$$

Ce qui prouve le théorème puisque $\mathbb{P}_x(N_x = 1) = \mathbb{P}_x(\tau_x = \infty) = a$. \square

Par la Proposition 1.24, on en déduit que si x est transient alors le nombre de visites N_x est fini presque sûrement, quel que soit le point de départ : $\mathbb{P}_y(N_x < \infty) = 1$ pour tout $y \in E$.

Corollaire 1.28. *Si l'espace d'états E est fini, alors il existe au moins un état récurrent.*

Preuve : Prenons un état initial z quelconque. On a, \mathbb{P}_z -p.s., $\infty = \sum_{x \in E} N_x$ (le temps total passé dans E est infini), mais $N_x < \infty$ \mathbb{P}_z -p.s. pour tout x transient, d'où la conclusion. \square

Théorème 1.29. *Les assertions suivantes sont équivalentes :*

- (1) x est récurrent.
- (2) $N_x = +\infty$, \mathbb{P}_x -p.s.
- (3) $G(x, x) = +\infty$.

Preuve :

- $1 \implies 2$: On a montré dans la preuve du Théorème 1.27 que la propriété de Markov donne, pour tout $k \in \mathbb{N}$, $\mathbb{P}_x(N_x = k + 1) = \mathbb{P}_x(\tau_x < \infty) \mathbb{P}_x(N_x = k)$. Si x est récurrent on a donc, pour tout $k \in \mathbb{N}$, $\mathbb{P}_x(N_x = k) = \mathbb{P}_x(N_x = k + 1)$. Comme $\sum_{k \in \mathbb{N}} \mathbb{P}_x(N_x = k) = \mathbb{P}_x(N_x < \infty) < \infty$ et que tous les termes de la somme sont égaux entre eux, ils sont nécessairement tous nuls et donc $\mathbb{P}_x(N_x < \infty) = 0$.
- $2 \implies 3$: trivial.
- $3 \implies 1$: provient du théorème précédent. \square

Remarquons que, si x est récurrent, on peut parfois avoir $N_x < \infty$ si le point de départ n'est pas x ; cela dépend des classes de communication de x et du point de départ.

1.7 Décomposition en classes de communication

Définition 1.30. *Un état x est dit **absorbant** si $P(x, x) = 1$.*

Bien sûr, un état absorbant est récurrent.

Définition 1.31. *On dit que l'état y est **accessible à partir de l'état x** , et on note $x \rightarrow y$, s'il existe $n > 0$ tel que $P^n(x, y) > 0$. On dit que les deux états x et y **communiquent** si $x \rightarrow y$ et $y \rightarrow x$, et on note alors $x \leftrightarrow y$.*

Proposition 1.32 (Transitivité). *Si $x \rightarrow y$ et $y \rightarrow z$, alors $x \rightarrow z$.*

Preuve : Par hypothèses, il existe $n, m \geq 1$ tels que $P^n(x, y) > 0$ et $P^m(y, z) > 0$. Or

$$P^{n+m}(x, z) = \sum_{w \in E} P^n(x, w)P^m(w, z) \geq P^n(x, y)P^m(y, z) > 0,$$

d'où le résultat. \square

On a donc

Corollaire 1.33. *La relation \leftrightarrow est une relation symétrique et transitive.*

Étant donné un état x , on note $C(x)$ l'ensemble des états qui communiquent avec x . Il est possible que $C(x) = \emptyset$; dans ce cas, $x \not\leftrightarrow x$, et x est appelé un **état de non-retour**. Remarquons que tout état de non-retour est clairement transient. On note E_{nr} l'ensemble des états de non-retour, et $E_r = E \setminus E_{nr}$. Alors sur E_r , la relation \leftrightarrow est une relation d'équivalence. Ses classes d'équivalence sont appelées les **classes de communication** (ou plus simplement les **classes**) de la chaîne de Markov : ce sont donc les ensembles $\mathcal{C}_1, \dots, \mathcal{C}_r$ (où $r \in \mathbb{N}^* \cup \{\infty\}$) tels que E est partitionné en

$$E = E_{nr} \cup \bigcup_{k=1}^r \mathcal{C}_k,$$

et pour $i = 1, \dots, r$, tous les états de \mathcal{C}_i communiquent entre eux, et ne communiquent avec aucun autre : pour tout $i \in \{1, \dots, r\}$ et $x, y \in \mathcal{C}_i$, on a $x \rightarrow y$, et pour tous $i, j \in \{1, \dots, r\}$ distincts, et tous $x \in \mathcal{C}_i, y \in \mathcal{C}_j$, on a $x \not\leftrightarrow y$ ou $y \not\leftrightarrow x$.

Définition 1.34. *Une chaîne de Markov est **irréductible** si tous ses états communiquent, autrement dit s'il n'y a pas d'état de non-retour et qu'il n'y a qu'une classe de communication.*

Définition 1.35. *Un ensemble $\mathcal{C} \subset E$ est dit **fermé** si aucun état $y \notin \mathcal{C}$ n'est accessible à partir d'un état quelconque $x \in \mathcal{C}$, autrement dit si*

$$P^n(x, y) = 0, \quad \forall n \geq 1, x \in \mathcal{C}, y \notin \mathcal{C}.$$

Ainsi, si $\mathcal{C} \subset E$ est fermé alors pour tout $x \in \mathcal{C}$, \mathbb{P}_x -p.s., pour tout $n \geq 1$, $X_n \in \mathcal{C}$ et donc, sous \mathbb{P}_x , $(X_n)_{n \geq 0}$ est une chaîne de Markov d'espace d'états \mathcal{C} et de matrice de transition $(P(x, y))_{x, y \in \mathcal{C}}$. On l'appelle la chaîne de Markov **restreinte à \mathcal{C}** .

Naturellement, les transitions en 1 pas suffisent à déterminer si un ensemble d'états est fermé :

Proposition 1.36. *Un ensemble $\mathcal{C} \subset E$ est fermé si, et seulement si*

$$P(x, y) = 0, \quad \forall x \in \mathcal{C}, y \notin \mathcal{C}.$$

Preuve : La partie "seulement si" est immédiate. Pour la partie "si", on procède par récurrence en n . \square

Proposition 1.37. *Toute partie non vide $\mathcal{C} \subset E$ fermée et finie contient au moins un état récurrent.*

Preuve : Cela vient du corollaire 1.28 appliqué à la chaîne de Markov restreinte à \mathcal{C} . \square

Proposition 1.38. *Si x est récurrent et que $x \rightarrow y$, alors $y \rightarrow x$ et y est aussi récurrent.*

Preuve : Nous montrons par l'absurde que $y \rightarrow x$. Supposons que $y \not\leftrightarrow x$, i.e. pour tout $n \geq 1$, $P^n(y, x) = 0$. Comme $x \rightarrow y$, il existe un $n \geq 1$ tel que $P^n(x, y) = \mathbb{P}_x(X_n = y) > 0$.

Remarquons que $N_x = \infty$ si, et seulement si $N_x \circ \theta_n = \infty$. En appliquant la propriété de Markov au temps n , on a ainsi :

$$\mathbb{P}_x(X_n = y, N_x = \infty) = \mathbb{P}_x(X_n = y, N_x \circ \theta_n = \infty) = \mathbb{P}_x(X_n = y)\mathbb{P}_y(N_x = \infty),$$

or

$$\mathbb{P}_y(N_x = \infty) \leq \mathbb{P}_y(N_x \geq 1) = \mathbb{P}_y(\exists k \geq 1 \text{ tel que } X_k = x) \leq \sum_{k \geq 1} \mathbb{P}_y(X_k = x) = \sum_{k \geq 1} P^k(y, x) = 0,$$

donc $\mathbb{P}_x(X_n = y, N_x = \infty) = 0$, et donc

$$0 < P^n(x, y) = \mathbb{P}_x(X_n = y) = \mathbb{P}_x(X_n = y, N_x < \infty) \leq \mathbb{P}_x(N_x < \infty),$$

ce qui montre que x est transient, une contradiction. Donc $y \rightarrow x$.

Par conséquent, $x \leftrightarrow y$. Soient $j, l \geq 1$ tels que $P^j(x, y) > 0$ et $P^l(y, x) > 0$. On a pour tout $n \geq 0$,

$$P^{n+j+l}(y, y) \geq P^l(y, x)P^n(x, x)P^j(x, y).$$

En sommant par rapport à n , on obtient

$$\begin{aligned} G(y, y) &\geq \sum_{m=n+j+l}^{\infty} P^m(y, y) \\ &\geq \sum_{n=0}^{\infty} P^l(y, x)P^n(x, x)P^j(x, y) \\ &= P^l(y, x)P^j(x, y)G(x, x) = \infty, \end{aligned}$$

d'après la récurrence de x . D'après le théorème 1.29, y est aussi récurrent. □

Corollaire 1.39.

- Deux états qui communiquent ont la même nature (tous récurrents ou tous transients).
 \rightsquigarrow On dira qu'une classe de communication est **récurrente** (resp. **transiente**) si tous ses états le sont. Toute classe est donc ou bien récurrente, ou bien transiente.
- Une classe récurrente est fermée. Par conséquent, une classe non fermée est transiente.
- Une classe fermée et finie est récurrente.

Remarque. Ce résultat permet donc de déterminer la nature de tous les états d'une chaîne de Markov sur un espace d'états *fini* à partir de son graphe (ou de sa matrice de transition). En revanche elle ne permet pas toujours de conclure quand l'espace d'états est infini, car savoir qu'une classe de communication infinie est fermée ne suffit pas à connaître sa nature.

1.8 Exemples

1.8.1 Chaîne à deux états

Si $0 < \alpha \leq 1$ et si $0 < \beta \leq 1$, la chaîne est irréductible récurrente.

Si $\alpha = 0$ et $0 < \beta \leq 1$, les classes sont $\mathcal{T} = \{1\}$, transiente, et $\mathcal{C} = \{0\}$, récurrente.

Si $\alpha = \beta = 0$, les classes sont $\mathcal{C}_1 = \{0\}$ et $\mathcal{C}_2 = \{1\}$, toutes deux récurrentes.

1.8.2 Marche aléatoire simple sur \mathbb{Z}

Comme $0 < p < 1$, tous les états communiquent. Il suffit donc de regarder par exemple la nature de l'état 0. On a

$$P^n(0, 0) = \mathbb{P}_0(X_n = 0) = \begin{cases} 0, & \text{si } n \text{ est impair;} \\ \binom{n}{n/2} p^{n/2} (1-p)^{n/2}, & \text{si } n \text{ est pair.} \end{cases}$$

À l'aide de la formule de Stirling (c'est-à-dire $n! \sim (n/e)^n \sqrt{2\pi n}$), on obtient

$$P^{2m}(0, 0) \sim \frac{1}{\sqrt{\pi m}} (4p(1-p))^m$$

d'où :

- si $p = 1/2$, $\sum_n P^n(0, 0) = \infty$ et la marche est donc récurrente ;
- si $p \neq 1/2$, $\sum_n P^n(0, 0) < \infty$ et la marche est transiente.

Lorsque $p \neq 1/2$, la loi forte des grands nombres permet d'être plus précis : comme $X_n = x_0 + \xi_1 + \dots + \xi_n$ avec ξ_1, ξ_2, \dots indépendantes et de même loi avec $\mathbb{E}(\xi_i) = 2p - 1$, on a $\frac{X_n}{n} \rightarrow 2p - 1$ p.s., donc

- si $p < 1/2$, alors $X_n \rightarrow -\infty$ p.s.,
- si $p > 1/2$, alors $X_n \rightarrow +\infty$ p.s..

On retrouve d'ailleurs le fait que X est transiente dans ces deux cas, car chaque état n'est visité qu'un nombre fini de fois.

1.8.3 Marche aléatoire simple symétrique sur \mathbb{Z}^d

Un calcul direct de $\mathbb{P}_0(X_n = 0)$ montre que la marche est récurrente quand $d = 2$ et transiente quand $d \geq 3$: pour n pair, on vérifie en effet que l'on a

$$\mathbb{P}_0(X_n = 0) = (2d)^{-n} \sum_{n_1 + \dots + n_d = \frac{n}{2}} \frac{n!}{(n_1! \dots n_d!)^2} \sim 2 \left(\frac{d}{2n\pi} \right)^{d/2}.$$

1.8.4 Processus de renouvellement

On constate que, sous \mathbb{P}_0 , $\tau_0 = Y_1$ donc $\tau_0 < \infty$ p.s. : l'état 0 est récurrent.

Selon la loi de Y , il peut y avoir des états de non-retour (par exemple, si Y est toujours pair alors l'état 1 ne communique pas avec lui-même). Par contre, tous les autres états communiquent nécessairement avec 0 donc il n'y a qu'une classe, qui est récurrente.

1.8.5 Processus de branchement

On constate que l'état 0 est absorbant. Les autres classes (et les états de non-retour) dépendent de la loi de reproduction (loi de Z) choisie.

À titre d'exercice, on pourra déterminer les états de non-retour, les classes et leur nature dans les cas suivants : si $\mathbb{P}(Z = 0) = 1$, si $\mathbb{P}(Z = 1) = 1$, si $\mathbb{P}(Z > 1) = 1$, si $\mathbb{P}(Z > 1) = 0$, $\mathbb{P}(Z = 0) > 0$ et $\mathbb{P}(Z = 1) > 0$, si $\mathbb{P}(Z = 0) = 0$, $\mathbb{P}(Z = 1) > 0$ et $\mathbb{P}(Z > 1) > 0$, si $\mathbb{P}(Z = 0) > 0$, $\mathbb{P}(Z = 1) > 0$ et $\mathbb{P}(Z > 1) > 0$.

1.9 Probabilités d'absorption

Notons \mathcal{T} l'ensemble des états transients, et $\mathcal{C}_1, \dots, \mathcal{C}_N$ les classes de récurrence (avec $N \in \mathbb{N} \cup \{\infty\}$). On a vu que ce sont des classes fermées : une fois entrée dans l'une d'elle, la chaîne n'en sort plus. Pour $i = 1, \dots, N$, le temps d'atteinte $T_{\mathcal{C}_i}$ est ainsi appelé **temps d'absorption par la classe \mathcal{C}_i** , et pour tout état $x \in E$ on définit la **probabilité d'absorption par la classe \mathcal{C}_i en partant de x** par

$$q_i(x) = \mathbb{P}_x(T_{\mathcal{C}_i} < \infty).$$

Remarquons que $q_i(x) = 0$ si $x \in \mathcal{C}_j$ où $j \neq i$, et $q_i(x) = 1$ si $x \in \mathcal{C}_i$.

Proposition 1.40. Soit $i \in \{1, \dots, N\}$. On a :

$$\text{pour tout } x \in \mathcal{T}, \quad q_i(x) = \sum_{y \in \mathcal{C}_i} P(x, y) + \sum_{z \in \mathcal{T}} P(x, z) q_i(z). \quad (1.2)$$

De plus, si \mathcal{T} est fini, les relations (1.2) caractérisent les probabilités $(q_i(x))_{x \in \mathcal{T}}$.

Preuve : Soit $i \in \{1, \dots, N\}$. Pour simplifier on note $T_i = T_{\mathcal{C}_i} = \inf\{n \geq 1 : X_n \in \mathcal{C}_i\}$ le temps d'entrée dans la classe \mathcal{C}_i . Soit $x \in \mathcal{T}$. On remarque que, sous \mathbb{P}_x , on a $T_i > 0$, et que $T_i = T_i \circ \theta_1$. Par la propriété de Markov au temps 1, on a donc :

$$\begin{aligned} \mathbb{P}_x(T_i < \infty) &= \sum_{z \in E} \mathbb{P}_x(X_1 = z, T_i < \infty) = \sum_{z \in E} \mathbb{P}_x(X_1 = z, T_i \circ \theta_1 < \infty) \\ &= \sum_{z \in E} \mathbb{P}_x(X_1 = z) \mathbb{P}_z(T_i < \infty) = \sum_{z \in E} P(x, z) q_i(z). \end{aligned}$$

En écrivant $E = \mathcal{T} \cup \mathcal{C}_i \cup \bigcup_{j \neq i} \mathcal{C}_j$, et vu que $q_i(z) = 1$ si $z \in \mathcal{C}_i$ et que $q_i(z) = 0$ si $z \in \mathcal{C}_j$ avec $j \neq i$, on obtient (1.2). Remarquons que la première somme de (1.2) est égale à $\mathbb{P}_x(X_1 \in \mathcal{C}_i) = \mathbb{P}_x(T_i = 1)$.

Supposons que \mathcal{T} est fini et qu'une fonction $f : \mathcal{T} \rightarrow [0, 1]$ vérifie (1.2) pour tout $x \in \mathcal{T}$. Alors par itération,

$$\begin{aligned} f(x) &= \mathbb{P}_x(T_i = 1) + \sum_{z \in \mathcal{T}} P(x, z) \left(\mathbb{P}_z(T_i = 1) + \sum_{z' \in \mathcal{T}} P(z, z') f(z') \right) \\ &= \mathbb{P}_x(T_i \leq 2) + \sum_{y \in \mathcal{T}} P^2(x, y) f(y). \end{aligned}$$

Par récurrence, on obtient que pour tout n ,

$$f(x) = \mathbb{P}_x(T_i \leq n) + \sum_{y \in \mathcal{T}} P^n(x, y) f(y).$$

Or pour tout état transient y , d'après le lemme suivant, $\lim_{n \rightarrow \infty} P^n(x, y) = 0$. Ce qui entraîne, comme \mathcal{T} est fini, que $f(x) = \lim_{n \rightarrow \infty} \mathbb{P}_x(T_i \leq n) = \mathbb{P}_x(T_i < \infty) = q_i(x)$. \square

Lemme 1.41. *Si $y \in E$ est transient alors on a, pour tout $x \in E$,*

$$\lim_{n \rightarrow \infty} P^n(x, y) = 0.$$

Preuve : Par la proposition 1.24, $G(x, y) \leq G(y, y) < \infty$, or $G(x, y) = \mathbb{E}_x(N_y) = \sum_n P^n(x, y)$, d'où le résultat. \square

La proposition suivante raccourcit souvent le calcul des probabilités d'absorption :

Proposition 1.42. *Si \mathcal{T} est fini alors, pour tout $x \in \mathcal{T}$,*

$$\sum_{i=1}^N q_i(x) = 1.$$

Preuve : On a en effet, pour tout $n \geq 0$,

$$1 = \mathbb{P}_x(X_n \in E) = \mathbb{P}_x(X_n \in \mathcal{T}) + \sum_{i=1}^N \mathbb{P}_x(X_n \in \mathcal{C}_i) = \sum_{y \in \mathcal{T}} P^n(x, y) + \sum_{i=1}^N \mathbb{P}_x(T_i \leq n)$$

d'où, en passant à la limite quand $n \rightarrow \infty$, à l'aide du lemme précédent, et du fait que \mathcal{T} est fini,

$$1 = \sum_{y \in \mathcal{T}} 0 + \sum_{i=1}^N \mathbb{P}_x(T_i < \infty),$$

ce qui conclut. \square

NB. Si \mathcal{T} est infini, la formule de la proposition peut être fautive. C'est clairement le cas s'il n'y a pas de classe récurrente. C'est aussi le cas, par exemple, pour les processus de branchement sur-critiques, c'est-à-dire où la probabilité d'extinction est non nulle (voir ci-dessous).

1.9.1 (*) Processus de branchement

On suppose que $\mathbb{P}(Z = 0) > 0$, $\mathbb{P}(Z = 1) > 0$ et $\mathbb{P}(Z > 1) > 0$. Dans ce cas, les classes sont $\{0\}$ et \mathbb{N}^* (pourquoi?), et \mathbb{N}^* n'est pas fermée, donc 0 est récurrent et tous les autres états sont transients. Cependant, cela ne nous dit pas si l'extinction a lieu presque sûrement ou s'il y a une probabilité strictement positive de non-extinction. La suite de cette partie est vraie pour toute loi de Z (mais les classes sont différentes car, si $\mathbb{P}(Z = 1) = 0$, il y a des états de non-retour).

La *probabilité d'extinction* est la probabilité d'absorption par la classe $\{0\}$, autrement dit

$$\mathbb{P}(\mathcal{M}), \quad \text{où } \mathcal{M} = \bigcup_{n=0}^{\infty} \{X_n = 0\}.$$

Il est clair que $\{X_n = 0\} \subset \{X_{n+1} = 0\}$, donc ces événements forment une suite croissante et

$$\mathbb{P}(\mathcal{M}) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0).$$

Théorème 1.43. *Supposons que $X_0 = 1$. Alors la probabilité d'extinction est la plus petit nombre $a \geq 0$ tel que*

$$g(a) = a,$$

où $g(a) = \mathbb{E}(a^Z) = \sum_{j=0}^{\infty} a^j \mathbb{P}(Z = j)$ est la fonction génératrice de Z .

Preuve : On établit d'abord

$$g_n = g_{n-1} \circ g.$$

Pour cela, soit $s \in [0, 1]$. On a

$$g_n(s) = \mathbb{E}s^{X_n} = \sum_{i=0}^{\infty} \mathbb{E}(s^{X_n} \mid X_{n-1} = i) \mathbb{P}(X_{n-1} = i),$$

or d'après l'expression des probabilités de transition,

$$\mathbb{E}(s^{X_n} \mid X_{n-1} = i) = \sum_{j \geq 0} p_{ij} s^j = \sum_{j \geq 0} \mathbb{P}(Z_1 + \dots + Z_i = j) s^j = \mathbb{E}(s^{Z_1 + \dots + Z_i}) = (\mathbb{E}s^{Z_i})^i = (g(s))^i,$$

d'où $g_n(s) = g_{n-1}(g(s))$.

Puisque $g_n = g(\dots(g(s)\dots))$ avec n compositions, on a aussi $g_n = g(g_{n-1})$.

Montrons que la probabilité d'extinction $q := \mathbb{P}(\mathcal{M})$ est un point fixe de g : on a

$$q = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0) = \lim_{n \rightarrow \infty} g_n(0) = \lim_{n \rightarrow \infty} g(g_{n-1}(0)).$$

Or g est continue, donc on a bien

$$q = g(q).$$

Montrons que q est le plus petit point fixe. Soit $a \in [0, 1]$ avec $g(a) = a$. Comme g est croissante,

$$g(0) \leq g(a) = a,$$

d'où

$$g_2(0) = g(g(0)) \leq g(a) = a,$$

ainsi par récurrence, pour tout $n \geq 1$, $g_n(0) \leq a$ d'où, en passant à la limite, $q \leq a$. \square

Théorème 1.44. *Supposons que $X_0 = 1$ et $\mathbb{P}(Z = 1) < 1$. Alors la probabilité d'extinction vaut 1 si et seulement si $\mathbb{E}(Z) \leq 1$.*

Preuve : On remarque que $\mathbb{E}(Z) = g'(1)$ (dérivée à gauche, $\in [0, +\infty]$). Par un raisonnement d'analyse réelle élémentaire (faire un dessin), dans ce cas, $g'(1) \leq 1$ si et seulement si 1 est l'unique point fixe de g . \square

NB. On pourra remarquer que les équations (1.2) donnant la probabilité d'absorption sont ici équivalentes à $g(q) = q$, où $q = q_{\{0\}}(1)$ est la probabilité d'extinction, c'est-à-dire la probabilité d'absorption par $\{0\}$ partant de l'état 1 (on peut en effet remarquer que $q_{\{0\}}(n) = q_{\{0\}}(1)^n$). Lorsque g a plusieurs points fixes, ces équations ne caractérisent donc pas les probabilités d'absorption.

1.10 Mesures et Probabilités invariantes

Dans la théorie des chaînes et des processus de Markov, la notion de probabilité invariante est essentielle. Elle correspond à celle d'état d'équilibre en physique.

Définition 1.45. On dit qu'une mesure (resp. une probabilité) m sur E est une **mesure invariante** (resp. une **probabilité invariante**) de la chaîne de Markov de noyau de transition P si $mP = m$, i.e.

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

Notons qu'une mesure m sur E équivaut à la donnée de la famille $(m(x))_{x \in E}$ d'éléments de $[0, +\infty]$, où $m(x) = m(\{x\})$, et que dans la notation mP on considère cette famille comme un vecteur ligne. Dans ce cours, on supposera toujours que les mesures sur E vérifient $m(x) < +\infty$ pour tout $x \in E$, afin d'éviter des cas pathologiques.

Remarquons que, si m est invariante, alors tous ses multiples λm , où $\lambda \in \mathbb{R}_+$, sont des mesures invariantes.

Définition 1.46. Un processus $(X_n, n \geq 0)$ est dit **stationnaire** si, pour tout $k \in \mathbb{N}$, les deux processus (X_0, X_1, \dots) et (X_k, X_{k+1}, \dots) ont la même loi.

Proposition 1.47. Une chaîne de Markov $(X_n, n \geq 0)$ est stationnaire si et seulement si la loi initiale est invariante.

Preuve : On note m la loi initiale de X et P sa matrice de transition. D'après le Corollaire 1.12, mP est la loi de X_1 . Donc si X est stationnaire, X_1 a la même loi que X_0 et par conséquent, $mP = m$, m est une loi invariante.

D'autre part, si $mP = m$, on a par itération $mP^k = m$ pour tout k . D'après le Corollaire 1.12 et la Proposition 1.18, pour tout k , (X_0, X_1, \dots) et (X_k, X_{k+1}, \dots) ont donc la même loi. \square

1.10.1 Existence de mesures invariantes

Le résultat suivant est fondamental. Il établit l'existence d'une mesure invariante (non nulle) pour les chaînes récurrentes, en en construisant un exemple explicite qui sera important dans la suite.

Théorème 1.48. Soit x un état récurrent fixé de la chaîne de Markov. On définit, pour tout $y \in E$,

$$m(y) = \mathbb{E}_x \left(\sum_{n=0}^{\tau_x-1} \mathbf{1}_{\{X_n=y\}} \right).$$

Alors, pour tout $y \in E$, $m(y) < \infty$, et m est une mesure invariante.

Remarques. La mesure m vérifie clairement les propriétés suivantes :

- $m(x) = 1$.
- Si $x \not\rightarrow y$, alors $m(y) = 0$. Comme x est récurrent, m est donc portée par la classe de x .
- La masse totale $m(E)$ de la mesure m est égale à l'espérance du temps de retour en x :

$$m(E) = \sum_{y \in E} m(y) = \mathbb{E}_x(\tau_x) \quad (\in]0, \infty]).$$

Preuve : On démontre d'abord l'invariance, en autorisant a priori $m(y)$ à prendre la valeur $+\infty$. Sous \mathbb{P}_x , on a $\tau_x < \infty$ par récurrence de x , et $X_0 = X_{\tau_x} = x$, si bien que

$$\sum_{n=0}^{\tau_x-1} \mathbf{1}_{\{X_n=y\}} = \sum_{n=1}^{\tau_x} \mathbf{1}_{\{X_n=y\}} = \sum_{k=0}^{\tau_x-1} \mathbf{1}_{\{X_{k+1}=y\}}.$$

Par suite, pour tout $y \in E$,

$$\begin{aligned} m(y) &= \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} \mathbf{1}_{\{X_{k+1}=y\}} \right) = \mathbb{E}_x \left(\sum_{k=0}^{\infty} \mathbf{1}_{\{k < \tau_x\}} \mathbf{1}_{\{X_{k+1}=y\}} \right) \\ &= \sum_{k=0}^{\infty} \mathbb{P}_x(k < \tau_x, X_{k+1} = y) = \sum_{k=0}^{\infty} \sum_{z \in E} \mathbb{P}_x(k < \tau_x, X_k = z, X_{k+1} = y) \end{aligned}$$

et, pour tous $k \in \mathbb{N}$ et $z \in E$, par la propriété de Markov au temps k ,

$$\mathbb{P}_x(k < \tau_x, X_k = z, X_{k+1} = y) = \mathbb{P}_x(k < \tau_x, X_k = z) \mathbb{P}_z(X_1 = y) = \mathbb{P}_x(k < \tau_x, X_k = z) P(z, y),$$

d'où, via un échange de sommes justifié par le fait que les coefficients sont ≥ 0 ,

$$m(y) = \sum_{z \in E} \left(\sum_{k=0}^{\infty} \mathbb{P}_x(k < \tau_x, X_k = z) \right) P(z, y) = \sum_{z \in E} \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} \mathbf{1}_{\{X_k=z\}} \right) P(z, y) = \sum_{z \in E} m(z) P(z, y),$$

c'est-à-dire $m(y) = (mP)(y)$. Ceci prouve l'invariance.

Il reste à s'assurer que $m(y) < \infty$ pour tout $y \in E$. C'est évident pour $y = x$ car $m(x) = 1$, et pour y hors de la classe de x car alors $m(y) = 0$ (y n'est jamais visité sous \mathbb{P}_x). Enfin, si $x \leftrightarrow y$, il existe n tel que $P^n(y, x) > 0$ et l'équation $m(x) = mP^n(x)$ s'écrit

$$1 = \sum_{z \in E} m(z) P^n(z, x) \geq m(y) P^n(y, x),$$

ce qui implique $m(y) < \infty$. □

Si x est récurrent, le théorème suivant montre que la mesure m relie asymptotiquement le temps passé en chaque état y au temps passé en x :

Théorème 1.49 (Théorème ergodique). *Soit x un état récurrent. Pour tout état y , \mathbb{P}_x -presque sûrement,*

$$\frac{\sum_{k=0}^{n-1} \mathbf{1}_y(X_k)}{\sum_{k=0}^{n-1} \mathbf{1}_x(X_k)} \xrightarrow{n \rightarrow \infty} m(y),$$

où m est la mesure (dépendant de x) définie par le théorème précédent. Plus généralement, pour toutes fonctions $f, g : E \rightarrow \mathbb{R}^+$, \mathbb{P}_x -presque sûrement,

$$\frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} \xrightarrow{n \rightarrow \infty} \frac{\int f \, dm}{\int g \, dm}$$

si $\int g \, dm$ est fini et non nul.

Remarquons que, pour toute fonction mesurable $f : E \rightarrow \mathbb{R}_+$,

$$\int f \, dm = \sum_{y \in E} f(y) m(y) = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} f(X_k) \right).$$

La première formule est simplement l'intégrale par rapport à une mesure discrète ; la deuxième correspond, dans le cas $f = \mathbf{1}_{\{y\}}$, à la définition de $m(y)$, et s'en déduit dans le cas général par convergence monotone en écrivant $f = \sum_{y \in E} f(y) \mathbf{1}_{\{y\}}$.

Preuve : La démonstration de ce théorème repose sur l'application de la loi forte des grands nombres aux excursions successives de la chaîne de Markov en-dehors de l'état x . Soit $f : E \rightarrow \mathbb{R}_+$ une fonction.

Notons $\alpha_0 = 0$, $\alpha_1 = \tau_x$ et, plus généralement, pour tout entier $r \geq 0$, α_r le temps de r -ième retour en x : pour tout entier $r \geq 0$,

$$\alpha_{r+1} = \inf\{k > \alpha_r \mid X_k = x\}.$$

Comme x est récurrent, on a $\alpha_r < \infty$ p.s. pour tout $r \geq 0$. Le résultat suivant sera plus facilement démontré avec le formalisme des lois conditionnelles introduit au chapitre suivants, et notamment avec la propriété de Markov "forte" ; on reporte sa preuve à la section 2.5.

Lemme 1.50. *Supposons que x est un état récurrent de la chaîne de Markov $(X_n)_n$. Alors, sous la probabilité \mathbb{P}_x , pour toute $f : E \rightarrow \mathbb{R}^+$, les variables aléatoires*

$$Z_r = \sum_{k=\alpha_r}^{\alpha_{r+1}-1} f(X_k), \quad r \in \mathbb{N},$$

sont indépendantes et de même loi.

On note $N_n = \max\{r \in \mathbb{N} \mid \alpha_r \leq n\}$ le nombre de retours en x avant le temps n . Alors, pour $n \geq 0$,

$$\alpha_{N_n} \leq n < \alpha_{N_n+1}$$

d'où, puisque f est à valeurs positives,

$$\sum_{k=0}^{\alpha_{N_n}-1} f(X_k) \leq \sum_{k=0}^{n-1} f(X_k) \leq \sum_{k=0}^{\alpha_{N_n+1}-1} f(X_k). \quad (1.3)$$

Remarquons que

$$\sum_{k=0}^{\alpha_{N_n}-1} f(X_k) = \sum_{r=0}^{N_n-1} \sum_{k=\alpha_r}^{\alpha_{r+1}-1} f(X_k) = \sum_{r=0}^{N_n-1} Z_r.$$

Par la remarque précédant la preuve, on a

$$\int f \, dm = \mathbb{E}_x(Z_0).$$

Supposons $\int f \, dm < \infty$. Alors en particulier Z_0 est intégrable donc, par le lemme, la loi forte des grands nombres s'applique à Z_0, Z_1, \dots , et donne

$$\mathbb{P}_{x\text{-p.s.}}, \quad \frac{1}{R} \sum_{r=0}^{R-1} Z_r \xrightarrow{R \rightarrow \infty} \mathbb{E}_x(Z_0) = \mathbb{E}_x \left(\sum_{k=0}^{\tau_x-1} f(X_k) \right) = \int f \, dm.$$

Comme $N_n \rightarrow +\infty$ p.s. (car x est récurrent), on en déduit que la suite $(\frac{1}{N_n} \sum_{r=0}^{N_n-1} Z_r)_{n \geq 0}$ a même limite (c'est-une sous-suite de la précédente), d'où finalement :

$$\mathbb{P}_{x\text{-p.s.}}, \quad \frac{1}{N_n} \sum_{k=0}^{\alpha_{N_n}-1} f(X_k) \xrightarrow{n \rightarrow \infty} \int f \, dm.$$

De même, on a aussi

$$\mathbb{P}_{x\text{-p.s.}}, \quad \frac{1}{N_n+1} \sum_{k=0}^{\alpha_{N_n+1}-1} f(X_k) \xrightarrow{n \rightarrow \infty} \int f \, dm$$

donc, vu que $N_n \sim N_n + 1$ quand $n \rightarrow \infty$ (car $N_n \rightarrow +\infty$), on déduit de (1.3) par encadrement que

$$\mathbb{P}_{x\text{-p.s.}}, \quad \frac{1}{N_n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{n \rightarrow \infty} \int f \, dm.$$

Le théorème s'en déduit en écrivant la même limite pour g , et en divisant. \square

Le théorème permet de comparer le nombre de visites en deux états x et y ; mais quel est l'ordre de grandeur du nombre $N_n^{(x)}$ de visites en x avant le temps n ? La preuve donne plus précisément

$$\mathbb{P}_{x\text{-p.s.}}, \quad \frac{1}{N_n^{(x)}} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{n \rightarrow \infty} \int f \, dm,$$

et cela reste vrai si $\int f dm = \infty$, en utilisant, dans la preuve, la loi forte des grands nombres pour les variables positives non intégrables (si $X_n \geq 0$ sont i.i.d. et $\mathbb{E}(X_i) = \infty$, $\frac{1}{n}(X_1 + \dots + X_n) \rightarrow +\infty$ p.s.).

En particulier, en prenant $f \equiv 1$, on obtient

$$\mathbb{P}_x\text{-p.s.}, \quad \frac{N_n^{(x)}}{n} \xrightarrow{n \rightarrow \infty} \frac{1}{m(E)} = \frac{1}{\mathbb{E}_x(\tau_x)},$$

avec comme limite 0 si $m(E) = \infty$, c'est-à-dire si $\mathbb{E}_x(\tau_x) = \infty$. Ainsi, deux comportements très différents sont possibles :

- si m est une mesure finie, alors x (et donc tout état) est visité une proportion positive du temps ;
- alors que si m est une mesure infinie, la proportion de temps passé en x est asymptotiquement nulle.

Intéressons-nous plus précisément au premier cas.

1.10.2 Probabilités invariantes

On s'intéresse à l'ensemble des mesures invariantes d'une chaîne de Markov, et plus particulièrement à l'ensemble des *probabilités invariantes*. Remarquons que l'existence d'une probabilité invariante équivaut à celle d'une mesure invariante dont la masse totale est finie, car diviser une mesure invariante par sa masse totale fournit une probabilité invariante. La proposition qui suit nous permet de ramener l'étude à celle d'une chaîne irréductible, en se restreignant à chaque classe fermée.

Proposition 1.51. *Soit m une mesure invariante. On suppose ou bien que l'ensemble \mathcal{T} des états transients est fini, ou bien plus généralement que $m(\mathcal{T}) < \infty$ (ce qui est le cas par exemple si m est une probabilité invariante). Alors*

a) *m ne charge pas \mathcal{T} : pour tout état transient x , on a $m(x) = 0$.*

b) *Pour toute classe récurrente \mathcal{C} , la restriction $m_{\mathcal{C}}$ de m à \mathcal{C} , définie par $m_{\mathcal{C}}(x) = \begin{cases} m(x) & \text{si } x \in \mathcal{C}, \\ 0 & \text{sinon,} \end{cases}$ est invariante.*

Par suite, m se décompose en une somme de mesures invariantes portées par chaque classe récurrente :

$$m = \sum_{\mathcal{C} \text{ classe récurrente}} m_{\mathcal{C}}$$

Preuve : a) Soit x un état transient. Rappelons que, par le lemme 1.41, $P^n(y, x) \rightarrow 0$ pour tout $y \in E$, et que, par la proposition 1.38, pour tout y récurrent, $y \not\rightarrow x$, c'est-à-dire $P^n(y, x) = 0$ pour tout $n \in \mathbb{N}$. Comme m est invariante on a, pour tout $n \geq 0$, $m = mP^n$, ce qui donne en particulier, grâce à ce qui précède,

$$m(x) = \sum_{y \in E} m(y)P^n(y, x) = \sum_{y \in \mathcal{T}} m(y)P^n(y, x) \xrightarrow{n \rightarrow \infty} 0,$$

où, selon l'hypothèse considérée, l'échange de somme et de limite est justifié ou bien par le fait que \mathcal{T} est fini, ou par théorème de convergence dominée vu que $m(y)P^n(y, x) \leq m(y)$ et $\sum_{y \in \mathcal{T}} m(y) = m(\mathcal{T}) < \infty$.

b) Vu ce qui précède, pour tout $x \in \mathcal{C}$ et tout $y \in E$, si y est transient alors $m(y) = 0$, et si y est récurrent et $y \notin \mathcal{C}$ alors $y \not\rightarrow x$ vu la proposition 1.38, si bien que l'invariance de m s'exprime par :

$$\text{pour tout } x \in \mathcal{C}, \quad m(x) = \sum_{y \in E} m(y)P(y, x) = \sum_{y \in \mathcal{C}} m(y)P(y, x),$$

c'est-à-dire $m_{\mathcal{C}} = m_{\mathcal{C}}P$. □

Remarquons que l'opération inverse de la restriction est aussi possible, de façon plus évidente :

Lemme 1.52. *Soit \mathcal{C} une classe de communication fermée. Soit m une mesure invariante pour la restriction $(P(x, y))_{x, y \in \mathcal{C}}$ de P à \mathcal{C} . Si on étend m sur E par $\tilde{m}(x) := 0$ si $x \notin \mathcal{C}$, alors \tilde{m} est une mesure invariante pour P .*

Preuve : Si $x \in \mathcal{C}$, comme m est invariante on a

$$\tilde{m}P(x) = \sum_{y \in E} \tilde{m}(y)P(y, x) = \sum_{y \in \mathcal{C}} m(y)P(y, x) = m(x) = \tilde{m}(x).$$

Si $x \notin \mathcal{C}$, comme \mathcal{C} est fermée on a $P(y, x) = 0$ si $y \in \mathcal{C}$ et donc

$$\tilde{m}P(x) = \sum_{y \in E} \tilde{m}(y)P(y, x) = \sum_{y \in \mathcal{C}} \tilde{m}(y)P(y, x) = 0 = \tilde{m}(x)$$

□

Par conséquent, vu la proposition précédente, déterminer les probabilités invariantes revient à déterminer les probabilités invariantes de la restriction à chaque classe récurrente. De même pour les mesures invariantes, lorsqu'il y a un nombre fini d'états transients. On va donc dorénavant supposer la chaîne *irréductible*.

On constate tout d'abord que, dans ce cas, une mesure invariante est portée par tous les états :

Lemme 1.53. *Si m est une mesure invariante d'une chaîne irréductible, et $m \not\equiv 0$, alors $m(y) > 0$ pour tout $y \in E$.*

Preuve : Il existe au moins un état x tel que $m(x) \neq 0$. Soit $y \in E$. Il existe n tel que $P^n(x, y) > 0$, et on a alors

$$m(y) = mP^n(y) = \sum_{z \in E} m(z)P^n(z, y) \geq m(x)P^n(x, y) > 0.$$

□

Proposition 1.54. (Unicité) *Si une chaîne de Markov irréductible possède une probabilité invariante π , toute autre mesure invariante est proportionnelle à π . En particulier, π est la seule probabilité invariante.*

Preuve : On suppose que π est une probabilité invariante. Soit m une mesure invariante et x un état fixé de E . Si $m \equiv 0$, le lemme est vrai. Supposons $m \not\equiv 0$, d'où $m(y) > 0$ pour tout $y \in E$ par le lemme précédent. Posons $\lambda = \frac{\pi(x)}{m(x)}$ et $m' = \lambda m$. Remarquons que m' est invariante et $m'(x) = \pi(x)$. Il faut montrer que $m' = \pi$. Définissons une mesure μ sur E , en posant, pour tout $y \in E$, $\mu(y) = \min(m'(y), \pi(y))$. Alors

$$(\mu P)(y) = \sum_{z \in E} \mu(z)P(z, y) \leq \sum_{z \in E} m'(z)P(z, y) = m'(y)$$

car m' est une mesure invariante. On voit de la même façon que $(\mu P)(y) \leq \pi(y)$. On a donc

$$(\mu P)(y) \leq \min(m'(y), \pi(y)) = \mu(y).$$

Mais on a aussi

$$\sum_{y \in E} \mu P(y) = \sum_{y \in E} \sum_{z \in E} \mu(z)P(z, y) = \sum_{z \in E} \mu(z) \sum_{y \in E} P(z, y) = \sum_{z \in E} \mu(z)$$

et ces sommes sont finies car $\sum_y \mu(y) \leq \sum_y \pi(y) = 1$, ce qui permet de calculer leur différence, et

$$\sum_{y \in E} (\mu(y) - \mu P(y)) = 0.$$

Les termes de la somme étant positifs ou nuls, ils sont donc tous nuls : $\mu P(y) = \mu(y)$ pour tout $y \in E$. Ainsi, μ est invariante. Comme π et m' aussi sont invariantes, $m_1 = \pi - \mu (\geq 0)$ et $m_2 = m' - \mu (\geq 0)$ sont également des mesures invariantes. Or $m'(x) = \pi(x) = \mu(x)$ (grâce au choix de λ), d'où $m_1(x) = m_2(x) = 0$, si bien que le lemme précédent implique que $m_1 = m_2 \equiv 0$, d'où $\pi = \mu = m'$. □

L'unicité de la probabilité invariante est donc assurée, sous réserve d'irréductibilité. L'important résultat suivant relie l'existence d'une probabilité invariante à l'intégrabilité des temps de retour.

Théorème 1.55. *Considérons une chaîne de Markov (X_n) irréductible. Les conditions suivantes sont équivalentes :*

- (i) *il existe un état $x \in E$ tel que $\mathbb{E}_x(\tau_x) < +\infty$;*
- (ii) *pour tout état $x \in E$, $\mathbb{E}_x(\tau_x) < +\infty$;*
- (iii) *la chaîne de Markov possède une probabilité invariante π .*

*Sous ces conditions la chaîne est récurrente, et dite **récurrente positive**. π est alors la seule probabilité invariante. De plus, pour tout $y \in E$,*

$$\pi(y) = \frac{1}{\mathbb{E}_y(\tau_y)}$$

et, pour tous $x, y \in E$,

$$\pi(y) = \frac{1}{\mathbb{E}_x(\tau_x)} \mathbb{E}_x \left(\sum_{n=0}^{\tau_x-1} \mathbf{1}_y(X_n) \right).$$

Preuve : Montrons d'abord que (i) implique (iii). Supposons donc (i). En particulier $\tau_x < \infty$, \mathbb{P}_x -p.s., donc x est récurrent. On note m la mesure invariante associée à l'état récurrent x définie au Théorème 1.48. Se rappelant que $m(E) = \mathbb{E}_x(\tau_x)$, on définit alors une probabilité invariante par

$$\pi = \frac{m}{\mathbb{E}_x(\tau_x)}, \quad (1.4)$$

ce qui prouve (iii).

Montrons que (iii) entraîne (ii). Supposons qu'il existe une probabilité invariante π . Il résulte du principe du maximum (cf. la dernière formule avant la Section 1.6) que, pour tout état x ,

$$\mathbb{E}_\pi[N_x] = \sum_{y \in E} \pi(y) \mathbb{E}_y[N_x] \leq \sum_{y \in E} \pi(y) \mathbb{E}_x[N_x] = \mathbb{E}_x[N_x].$$

Or, comme π est invariante, $\pi(x) > 0$ par le lemme 1.53, et, pour tout n , la loi de X_n sous \mathbb{P}_π est π donc

$$\mathbb{E}_\pi[N_x] = \sum_{n=0}^{\infty} \mathbb{P}_\pi(X_n = x) = \sum_{n=0}^{\infty} \pi(x) = +\infty.$$

On en déduit que $\mathbb{E}_x[N_x] = \infty$, et donc que x est récurrent. Puisque la mesure m associée à x est une mesure invariante, il résulte de la proposition 1.54 que m et π sont proportionnelles, donc que $m(E)$ est fini. Or $m(E) = \mathbb{E}_x(\tau_x)$, donc $\mathbb{E}_x(\tau_x)$ est fini, ce qui montre (ii). Puisque (ii) entraîne (i) de façon évidente, ceci prouve les équivalences du théorème.

L'unicité de π résulte de la proposition 1.54. La deuxième formule donnant π est la formule (1.4). Cette formule appliquée à x donne

$$\pi(x) = \frac{\mathbb{E}_x(\sum_{n=0}^{\tau_x-1} \mathbf{1}_x(X_n))}{\mathbb{E}_x(\tau_x)} = \frac{1}{\mathbb{E}_x(\tau_x)}.$$

(Par l'unicité, π ne dépend pas du choix de x dans (1.4)) □

Dans le cas non irréductible, on peut appliquer le théorème précédent à toute classe récurrente. Les conditions du théorème peuvent alors être satisfaites ou non selon la classe, d'où les définitions suivantes.

Définition 1.56. *Un état récurrent x est **récurrent positif** si $\mathbb{E}_x(\tau_x) < \infty$, et **récurrent nul** sinon. On étend ces définitions à toute une classe de communication.*

Corollaire 1.57. *Lorsque E est fini, toute chaîne irréductible est récurrente positive.*

Preuve : L'existence d'un état récurrent résulte du Corollaire 1.28. On sait qu'à cet état récurrent, on peut associer une mesure invariante. Elle est de masse finie car l'espace est fini. \square

Pour les chaînes récurrentes positives, le théorème ergodique prend la forme d'une loi des grands nombres :

Théorème 1.58 (Théorème ergodique des chaînes de Markov récurrentes positives). *Si (X_n) est une chaîne de Markov irréductible récurrente positive de probabilité invariante π , alors pour toute loi initiale ν , pour tout état $y \in E$, \mathbb{P}_ν -presque sûrement,*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_y(X_k) = \pi(y),$$

et plus généralement pour toute fonction $f : E \rightarrow \mathbb{R}^+$, \mathbb{P}_ν -presque sûrement,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \int f d\pi. \quad (*)$$

et pour tous $x, y \in E$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k(x, y) = \pi(y).$$

Preuve : Pour tout état x , il résulte du Théorème 1.49, appliqué à $g = 1$, que (*) est vraie \mathbb{P}_x -presque sûrement (se souvenir que π ne dépend pas de x). C'est encore vrai \mathbb{P}_ν -p.s. car

$$\mathbb{P}_\nu(\cdot) = \sum_{x \in E} \nu(x) \mathbb{P}_x(\cdot).$$

En prenant $f = \mathbf{1}_y$, on obtient la première limite, et en prenant de plus $\nu = \delta_x$ le théorème de convergence dominée (domination par 1) donne la dernière limite. \square

Ce théorème donne deux moyens pratiques d'approcher π si, comme c'est souvent le cas, on ne sait pas la calculer explicitement. La première façon est la méthode de Monte Carlo, qui consiste à simuler sur ordinateur une longue trajectoire $X_n(\omega)$ de la chaîne, puis à faire la moyenne de f le long de cette trajectoire. D'après (*) on obtient ainsi à peu près $\int f d\pi$. L'autre façon est de calculer itérativement P^n , par exemple dans le cas où E est fini. Puis de faire la moyenne des $P^n(x, y)$ pour approcher $\pi(y)$ (on peut montrer que la vitesse de convergence est exponentielle). C'est très souvent beaucoup plus rapide que la recherche d'un vecteur propre de la transposée de P .

1.11 Périodicité et existence de la loi limite au sens fort

Nous avons vu que pour une matrice de transition irréductible et récurrente positive, la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P^k(x, y) = \pi(y)$$

existe et définit une loi invariante (et l'unique loi invariante). Parfois, le résultat plus fort suivant est valable :

$$\lim_{n \rightarrow \infty} P^n(x, y) = \pi(y),$$

cependant ce n'est pas toujours le cas. Cela dépend de la notion de périodicité :

Définition 1.59. *Si x est un état tel que $x \rightarrow x$, la **période** de x est*

$$d(x) = \text{pgcd}(\{n \in \mathbb{N}^* \mid P^n(x, x) > 0\}).$$

Lemme 1.60. *Pour tous $x, y \in E$, si $x \leftrightarrow y$ alors $d(x) = d(y)$. Autrement dit, la période est un nombre qui ne dépend que de la classe à laquelle appartient l'état considéré.*

Preuve : Soit $x, y \in E$ tels que $x \leftrightarrow y$. Il existe $m, l > 0$ tels que $P^m(x, y) > 0$ et $P^l(y, x) > 0$. On a alors

$$P^{m+l}(x, x) \geq P^m(x, y)P^l(y, x) > 0$$

donc $d(x)$ divise $m + l$. Pour tout n tel que $P^n(y, y) > 0$, on a

$$P^{m+n+l}(x, x) \geq P^m(x, y)P^n(y, y)P^l(y, x) > 0$$

donc $d(x)$ divise $m + n + l$, ce qui implique, avec ce qui précède, que $d(x)$ divise n . Par définition de $d(y)$, on a donc $d(x) \leq d(y)$. Par symétrie on a aussi $d(y) \leq d(x)$, donc $d(x) = d(y)$. \square

Définition 1.61. *Une chaîne de Markov irréductible (resp. une classe) est **apériodique** si la période commune à tous les états de E (resp. à toute la classe) est 1, **périodique** sinon.*

Proposition 1.62. *Soit C une classe récurrente. On note d sa période. Pour tout $x \in C$, il existe n_0 tel que, pour tout $n \geq n_0$, $P^{dn}(x, x) > 0$.*

Preuve : Soit $(n_l)_{l \geq 1}$ la famille des entiers positifs tels que $P^{n_l}(x, x) > 0$. Pour tout $s \geq 1$, posons $d_s := \text{pgcd}(n_1, \dots, n_s)$. La suite d'entiers positifs $(d_s)_{s \geq 1}$ est décroissante donc stationne en une valeur $d_t : d_s = d_t$ pour tout $s \geq t$. Or d_t divise tous les n_l donc divise aussi leur pgcd d : on a $d_t | d$. D'autre part, $d | n_l$ pour $l = 1, \dots, t$ donc $d | d_t$. Ainsi, $d = d_t = \text{pgcd}(n_1, \dots, n_t)$.

Par le lemme qui suit, il existe donc un entier N tel que, pour tout $n \geq N$, on a $nd = \alpha_1 n_1 + \dots + \alpha_t n_t$ pour certains entiers naturels $\alpha_1, \dots, \alpha_t$ et donc

$$P^{nd}(x, x) \geq \prod_{l=1}^t (P^{n_l}(x, x))^{\alpha_l} > 0.$$

\square

Lemme 1.63. *Soient $n_1, \dots, n_t \in \mathbb{N}$. On pose $d = \text{pgcd}(n_1, \dots, n_t)$. Alors il existe un entier N tel que pour tout $n \geq N$, on a*

$$nd = \alpha_1 n_1 + \dots + \alpha_t n_t$$

pour certains entiers $\alpha_1, \dots, \alpha_t \in \mathbb{N}$.

Preuve : En divisant tous les n_l par d , on réduit le problème au cas $d = 1$. D'après le lemme de Bézout, il existe une combinaison linéaire $\beta_1 n_1 + \dots + \beta_t n_t = 1$, avec des coefficients $\beta_l \in \mathbb{Z}$.

En rassemblant séparément les termes positifs et négatifs, on obtient alors $p - q = 1$ où p et q sont des sommes de termes positifs. Posons $N = q^2 - 1$. Alors si $n \geq N$, la division de n par q donne

$$n = \alpha q + \beta,$$

où $0 \leq \beta < q$, d'où $\alpha \geq q - 1 \geq \beta$, et l'écriture suivante prouve le lemme :

$$n = \alpha q + \beta(p - q) = (\alpha - \beta)q + \beta p.$$

\square

Théorème 1.64 (Convergence en loi des chaînes de Markov). *Considérons une chaîne de Markov X de matrice de transition P irréductible, récurrente positive et apériodique. On note π la loi invariante. Pour toute loi initiale ν ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}_\nu(X_n = j) = \pi_j, \quad \forall j \in E.$$

En particulier, on a

$$\lim_{n \rightarrow \infty} P^n(i, j) = \pi_j, \quad \forall i, j.$$

Preuve : On utilise un argument de couplage : Soit Y une autre chaîne de Markov de noyau de transition P , de loi initiale π , *indépendante* de X . Une telle chaîne Y existe : pour construire le couple (X, Y) , il suffit de se placer sur l'espace produit $E^{\mathbb{N}} \times E^{\mathbb{N}}$ muni de la probabilité produit $\mathbb{P}_{\nu} \times \pi$. Soit $W_n = (X_n, Y_n)$ la chaîne couplée. W est une chaîne de Markov de loi initiale $\nu \times \pi$ et de matrice de transition $\tilde{P}((i, k), (j, l)) = P(i, j)P(k, l)$. La chaîne W est irréductible et apériodique, car pour tout n , $\tilde{P}^n((i, k), (j, l)) = P^n(i, j)P^n(k, l)$ qui est strictement positif pour tout n assez grand (*c'est ici on a utilisé l'apériodicité du P , sinon on ne peut pas garantir que $P^n(i, j)P^n(k, l) > 0$ pour tout couple (i, k) et (j, l) , donc W ne serait même pas irréductible*).

Il est immédiat de vérifier que la loi $\tilde{\pi}(i, l) := \pi_i \pi_l$ définie sur $E \times E$ est une loi invariante pour \tilde{P} .

Alors W est une chaîne récurrente positive irréductible et apériodique. Le temps d'atteinte T de la diagonale de $E \times E$ par W :

$$T := \inf\{n \geq 1 : X_n = Y_n\},$$

est $\tilde{\mathbb{P}}_{\nu \times \pi}$ -p.s. fini car T est plus petit que le premier temps d'atteinte de n'importe quel (i, i) par W , qui est presque sûrement fini. On définit maintenant une nouvelle chaîne (Z_n) par

$$Z_n(\omega) := \begin{cases} X_n(\omega), & \text{sur } \{T(\omega) > n\}; \\ Y_n(\omega), & \text{sur } \{T(\omega) \leq n\}. \end{cases}$$

On remarque que $Z_T = X_T = Y_T$ et que $Z_0 = X_0$ a pour loi initiale ν . On vérifie maintenant que sous $\tilde{\mathbb{P}}_{\nu \times \pi}$, Z est une chaîne de Markov de noyau de transition P ; En fait,

$$\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j | Z_n = i_n, \dots, Z_0 = i_0) = \frac{\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, Z_n = i_n, \dots, Z_0 = i_0)}{\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = i_n, \dots, Z_0 = i_0)}.$$

Pour simplifier les notations, soit $A_k := \{Z_0 = i_0, \dots, Z_k = i_k\}$ pour tout $1 \leq k \leq n$. Alors

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, A_n) &= \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T > n, A_n) + \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = n, A_n) \\ &\quad + \sum_{k=0}^{n-1} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = k, A_n). \end{aligned} \quad (1.5)$$

On discute ces trois cas séparément. Sur $\{T > n\}$, $Z_{n+1} = X_{n+1}$ et $A_n = \{X_0 = i_0, \dots, X_n = i_n\}$. D'après la propriété de Markov de X en n (X et Y sont indépendants),

$$\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T > n, A_n) = P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T > n, X_n = i_n, \dots, X_0 = i_0) = P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T > n, A_n).$$

Sur $\{T = n\}$, $Z_{n+1} = Y_{n+1}$, $Z_n = Y_n$ et $A_n = \{Y_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\}$. Donc la propriété de Markov de Y en n entraîne que

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = n, A_n) &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = n, Y_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = n, A_n). \end{aligned}$$

Finalement sur $\{T = k\}$ avec $k \leq n-1$, $\{Z_{n+1} = j, T = k, A_n\} = \{Y_{n+1} = j, Y_n = i_n, \dots, Y_k = i_k, T = k, X_{k-1} = i_{k-1}, \dots, X_0 = i_0\}$. On applique la propriété de Markov en n et obtient

$$\begin{aligned} \tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j, T = k, A_n) &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = i_n, \dots, Y_k = i_k, T = k, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\ &= P(i, j) \tilde{\mathbb{P}}_{\nu \times \pi}(T = k, A_n), \end{aligned}$$

donc $\tilde{\mathbb{P}}_{\nu \times \pi}(Z_{n+1} = j | Z_n = i_n, \dots, Z_0 = i_0) = P(i, j)$ en sommant ces trois cas dans (1.5).

On voit que X et Z ont la même loi initiale et même noyau de transition donc ont la même loi. Donc

$$\mathbb{P}_{\nu}(X_n = j) - \pi_j = \tilde{\mathbb{P}}_{\nu \times \pi}(X_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j) = \tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j).$$

Remarquer que sur $\{T \leq n\}$, $Z_n = Y_n$. Donc $\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j) = \tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j, n < T) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j, n < T)$ et

$$|\mathbb{P}_{\nu}(X_n = j) - \pi_j| \leq |\tilde{\mathbb{P}}_{\nu \times \pi}(Z_n = j, n < T) - \tilde{\mathbb{P}}_{\nu \times \pi}(Y_n = j, n < T)| \leq \tilde{\mathbb{P}}_{\nu \times \pi}(n < T) \rightarrow 0,$$

car T est presque sûrement fini. \square

1.11.1 (*) Classes cycliques

Proposition 1.65. *Soit C une classe récurrente. On note d la période de cette classe. Fixons $x \in C$. Pour tout $y \in C$, il existe un entier $\nu_y \in \{0, \dots, d-1\}$ tel que*

a) $P^n(x, y) > 0 \implies n \equiv \nu_y \pmod{d}$.

b) il existe n_y tel que si $n \geq n_y$ alors $P^{nd+\nu_y}(x, y) > 0$.

Preuve : a) Soit $y \in C$. Choisissons r et m tels que $P^r(y, x) > 0$ et $P^m(x, y) > 0$. Si $P^n(x, y) > 0$ alors on a

$$P^{m+r}(x, x) > 0 \quad \text{et} \quad P^{n+r}(x, x) > 0,$$

d'où, par la définition de période, $d|m+r$ et $d|n+r$, et donc $d|m-n$ et $n \equiv m \pmod{d}$. L'entier ν_y est le reste de la division euclidienne de m par d .

b) Par la Proposition 1.62, il existe N tel que, pour tout $n \geq N$, $P^{nd}(x, x) > 0$. Soit $y \in C$. D'après a) et le fait que $x \rightarrow y$, il existe $m \geq 0$ tel que $P^{md+\nu_y}(x, y) > 0$. Posons $n_y = N + m$. Pour tout $n \geq n_y$, on a $n = n' + m$ avec $n' \geq N$, et donc $nd + \nu_y = n'd + md + \nu_y$, par conséquent

$$P^{nd+\nu_y}(x, y) \geq P^{n'd}(x, x)P^{md+\nu_y}(x, y) > 0,$$

d'où b). □

Supposons $x \in C$ fixé. Pour $\nu \in \{0, \dots, d-1\}$, on définit

$$C^{(\nu)} \equiv C^{(\nu)}(x) := \left\{ j \in C \mid \nu_j \text{ de la proposition 1.65 a) vaut } \nu \right\}.$$

$C^{(0)}, \dots, C^{(d-1)}$ sont appelées les **sous-classes cycliques** de C . Nous étendons la définition de $C^{(\nu)}$ à tout entier $\nu \geq 0$ en posant $C^{(s)} := C^{(\nu)}$ si $s \equiv \nu \pmod{d}$.

Proposition 1.66. *Soit C une classe récurrente de période d .*

a) $C^{(0)}, \dots, C^{(d-1)}$ sont deux à deux disjointes et leur réunion donne C .

b) si $j \in C^{(\nu)}$ et $P_{jk}^{(n)} > 0$ ($n > 0$), alors $k \in C^{(\nu+n)}$. Donc

$$\sum_{k \in C^{(\nu+n)}} P^n(j, k) = 1.$$

Preuve : a) est une conséquence de la proposition 1.65. Quant à b), considérons un $m > 0$ tel que $P^m(i, j) > 0$, on a

$$P^{m+n}(i, k) \geq P^m(i, j)P^n(j, k) > 0.$$

Ceci en vue de la proposition précédente (b) implique que

$$m + n \equiv \nu_k \pmod{d},$$

donc $k \in C^{(m+n)} = C^{(\nu+n)}$ puisque $m \equiv \nu \pmod{d}$. □

Cette proposition justifie l'attribut "cyclique", car elle met en évidence l'évolution cyclique de la chaîne partant d'un état de C : si la chaîne part d'un état j d'une sous-classe, elle atteint à nouveau les états de cette sous-classe aux temps $d, 2d, 3d, \dots$

D'un état j on passe en un pas à un état de la sous-classe suivante.

Il est intéressant d'observer que si la chaîne part de $i \in C^{(0)}$, alors $(X_{nd})_{n \geq 0}$ est une chaîne de Markov ayant $C^{(0)}$ pour ensemble d'états et $(P^d(i, j))_{i, j \in C^{(0)}}$ comme matrice de transition. (*exercice*). Remarquons que $C^{(0)}$ est une classe fermée pour P^d et elle est irréductible et de période 1. En effet, $P^{nd}(i, i) > 0$ pour tout $n \geq N$.

Nous allons montrer que les sous-classes $C^{(0)}(i), \dots, C^{(d-1)}(i)$ ne vont pas beaucoup dépendre de i . Pour cela, soit $j \in C^{(\nu)}(i)$, on considère les sous-classes associées $C^{(0)}(j), \dots, C^{(d-1)}(j)$. Soit $k \in C^{(s)}(j)$. Soient m et n tels que

$$P^m(i, j) > 0, \quad P^n(j, k) > 0.$$

On a

$$m = \nu \bmod d, \quad n = s \bmod d.$$

D'autre part,

$$P^{m+n}(i, k) \geq P^m(i, j)P^n(j, k) > 0,$$

donc $k \in C^{(m+n)}(i) = C^{(\nu+s)}(i)$, car $m + n = \nu + s \bmod d$.

Il s'ensuit que $C^{(s)}(j) \subset C^{(\nu+s)}(i)$. Mais les sous-classes forment une partition de C , donc

$$C^{(s)}(j) = C^{(\nu+s)}(i), \quad \forall s = 0, \dots, d-1.$$

Ainsi les sous-classes sont les mêmes à un décalage d'indices près.

Exemple : Considérons, sur $E = \{0, 1, 2, 3, 4\}$, une chaîne de matrice de transition

$$P = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

Supposons que la chaîne part de $0 \in C^{(0)}$. En un seul pas, elle peut aller à 2 ou 3 qui vont appartenir forcément à $C^{(1)}$. Ensuite on va à 1 donc $1 \in C^{(2)}$, puis de 1 on va à $0 \in C^{(3)}$ ou $4 \in C^{(3)}$, ceci implique $C^{(3)} = C^{(0)}$. La période vaut donc 3 et $C^{(0)} = \{0, 4\}$, $C^{(1)} = \{2, 3\}$ et $C^{(2)} = \{1\}$.

Théorème 1.67. *Considérons une chaîne de Markov X de matrice de transition P irréductible, récurrente positive, de période $d \geq 2$. On note π la loi invariante. Pour tout couple i, j , il existe un entier $a \in \{0, \dots, d-1\}$ tel que $P^m(i, j) = 0$ si $m \not\equiv a \bmod d$, et*

$$\lim_{n \rightarrow \infty} P^{nd+a}(i, j) = \frac{\pi_j}{\pi(C^{(a)})} = d\pi_j,$$

où $\pi(C^{(a)}) = \sum_{j \in C^{(a)}} \pi_j = \frac{1}{d}$.

Preuve : On vérifie d'abord que $\pi(C^{(a)}) = \frac{1}{d}$. Comme $\pi = \pi P$ et que pour $j \in C^{(0)}$, $\pi_j = \sum_{i \in E} \pi_i P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i P_{i,j}$, donc

$$\pi(C^{(0)}) = \sum_{j \in C^{(0)}} \sum_{i \in C^{(d-1)}} \pi_i P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i \sum_{j \in C^{(0)}} P_{i,j} = \sum_{i \in C^{(d-1)}} \pi_i$$

car $\sum_{j \in C^{(0)}} P_{i,j} = 1$ pour tout $i \in C^{(d-1)}$; donc $\pi(C^{(0)}) = \pi(C^{(d-1)})$. En considérant $\pi = \pi P^2 = \dots = \pi P^{d-1}$, on obtient que $\pi(C^{(0)}) = \pi(C^{(1)}) = \dots = \pi(C^{(d-1)}) = \frac{1}{d}$ puisque les sommes de $\pi(C^{(1)}), \dots, \pi(C^{(d)})$ vaut 1.

On traite par exemple le cas $a = 0$. La matrice P^d apériodique et pour tout $i \in C^{(0)}$, $P_{i,j}^d = 0$ si $j \notin C^{(0)}$. Alors la mesure $\hat{\pi}_j := \frac{\pi_j}{\pi(C^{(0)})}$, $j \in C^{(0)}$ est une loi sur $C^{(0)}$, de plus, elle est invariante pour P^d . Alors appliquer (1) on obtient la limite annoncée. \square

1.12 (*) Complément sur le processus canonique

Soit P une matrice stochastique, sur un espace d'états E , et μ une loi sur E . On a jusque-là admis (cf. Proposition 1.6) l'existence d'une chaîne de Markov de matrice P et de loi initiale μ . Justifions-la ici.

1.12.1 Première construction

Quitte à numéroter les états, E étant dénombrable on peut supposer $E \subset \mathbb{N}$, et même $E = \mathbb{N}$ quitte à ajouter des états où la mesure μ est nulle. Remarquons que, si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable aléatoire $X = f_\mu(U)$ suit la loi μ , en définissant la fonction

$$f_\mu : u \mapsto \min\{n \in \mathbb{N} \mid \mu(0) + \mu(1) + \cdots + \mu(n) \geq U\}.$$

En effet, pour tout $n \in \mathbb{N}$, $f_\mu(U) = n$ si, et seulement si $U \in]\mu(0) + \cdots + \mu(n-1), \mu(0) + \cdots + \mu(n)]$, et cet intervalle a pour largeur $\mu(n)$.

Supposons donnée une suite $(U_n)_{n \geq 0}$ de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. On discutera de l'existence (non évidente) d'une telle suite plus bas.

Alors on peut définir par récurrence la suite $(X_n)_{n \geq 0}$ de variables aléatoires par : $X_0 = f_\mu(U_0)$ et, pour tout $n \geq 0$,

$$X_{n+1} = f_{P(X_n, \cdot)}(U_{n+1}),$$

où $P(X_n, \cdot)$ est la mesure de probabilité donnée par la ligne X_n de la matrice P . Par cette construction, X_0 suit la loi μ et, pour tout n , la loi de X_{n+1} sachant $\{X_0 = x_0, \dots, X_n = x_n\}$ est la loi de $f_{P(x_n, \cdot)}(U_{n+1})$ (car U_{n+1} est indépendante de X_0, \dots, X_n), c'est-à-dire $P(x_n, \cdot)$. Le processus $(X_n)_{n \geq 0}$ est donc une chaîne de Markov de loi initiale μ et de matrice de transition P . Notons \mathbb{P}_μ sa loi : c'est une mesure de probabilité sur l'espace $E^\mathbb{N}$ des suites à valeurs dans E , muni de sa tribu produit (voir rappels). L'existence de \mathbb{P}_μ étant maintenant acquise, on va remplacer X par une version plus "standard".

Considérons l'espace canonique $\Omega = E^\mathbb{N}$ des suites à valeurs dans E , muni de sa tribu produit \mathcal{F} . Le processus canonique sur E est défini par $X = \text{Id}_\Omega$: c'est la fonction identité $X = (X_n)_{n \geq 0} : \Omega \rightarrow E^\mathbb{N}$. Pour toute probabilité μ sur E , sous \mathbb{P}_μ , ce processus X est une chaîne de Markov de loi initiale μ et de matrice P .

Avec l'espace canonique $\Omega = E^\mathbb{N}$, on travaille donc toujours avec le même processus $X : E^\mathbb{N} \rightarrow E^\mathbb{N}$ (l'identité), mais sous différentes lois \mathbb{P}_μ . De plus, on peut définir l'opérateur de décalage (ou **shift**) $\theta : \Omega \rightarrow \Omega$ par

$$\theta((x_n)_{n \geq 0}) = (x_{n+1})_{n \geq 0}, \quad \text{pour tout } (x_n)_{n \geq 0} \in \Omega,$$

qui vérifie $X_n \circ \theta = X_{n+1}$.

Justifions enfin l'existence de suites $(U_n)_{n \geq 0}$ de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. Prenons $\Omega = [0, 1]$, muni de la loi uniforme, et $U : [0, 1] \rightarrow [0, 1]$ définie par $U(\omega) = \omega$, de telle sorte que U suit la loi uniforme sur $[0, 1]$. On rappelle que, si $U = \overline{0, X_0 X_1 \cdots}$ est le développement de U en base 2 (avec $X_0, X_1, \dots \in \{0, 1\}$), alors les variables $(X_n)_{n \geq 0}$ sont indépendantes et de même loi $\mathcal{B}(1/2)$ (et vice-versa, si $(X_n)_n$ est ainsi, U suit la loi uniforme sur $[0, 1]$). Considérons une bijection $\varphi : \mathbb{N}^2 \rightarrow \mathbb{N}$ (on pourrait définir φ explicitement). Alors, en définissant, pour tout $n \geq 0$, la variable aléatoire U_n par

$$U_n = \overline{0, X_{\varphi(n,0)} X_{\varphi(n,1)} \cdots} = \sum_{k=0}^{\infty} \frac{X_{\varphi(n,k)}}{2^{k+1}},$$

les variables $(U_n)_{n \geq 0}$ sont indépendantes, du fait de la propriété d'indépendance par paquets, et suivent toutes la loi uniforme sur $[0, 1]$, comme rappelé plus haut.

1.12.2 Autre approche : extension de lois fini-dimensionnelles compatibles

On décrit ici, sans démonstration, comment cette existence s'inscrit dans un cadre plus général d'existence de processus.

Soit X un processus à valeurs dans un espace (E, \mathcal{E}) (sans hypothèse de dénombrabilité). On note μ_n la loi de (X_0, X_1, \dots, X_n) . C'est une probabilité sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$. Si on note $\pi_{n+1,n}$ la projection canonique de E^{n+1} sur E^n i.e. l'application $(x_0, \dots, x_{n-1}, x_n) \mapsto (x_0, \dots, x_{n-1})$, on a

$$\pi_{n+1,n}(\mu_n) = \mu_{n-1}.$$

Ceci est équivalent à, pour tout $A_k \in \mathcal{E}$,

$$\mu_{n-1}(A_0 \times A_1 \times \dots \times A_{n-1}) = \mu_n(A_0 \times A_1 \times \dots \times A_{n-1} \times E). \quad (1.6)$$

Les probabilités $(\mu_n)_{n \geq 0}$ s'appellent les répartitions finies du processus X .

Réciproquement, si on se donne des probabilités μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant la consistence (1.6), se pose la question de savoir s'il existe un processus ayant pour répartitions finies les μ_n . On introduit l'espace canonique $\Omega = E^{\mathbb{N}}$ pour $\omega = (\omega_n)_{n \geq 0}$, $X_n(\omega) = \omega_n$, $\mathcal{F}_n = \sigma(X_k, k \leq n)$, $\mathcal{F} = \sigma(X_k, k \geq 0)$.

Soit $A \in \mathcal{F}_n$, A est de la forme $A = B \times E \times \dots \times E \times \dots$ avec $B \in \mathcal{E}^{\otimes(n+1)}$. On définit alors une probabilité \mathbb{P}_n sur (Ω, \mathcal{F}_n) en posant $\mathbb{P}_n(A) = \mu_n(B)$ puis une fonction d'ensembles sur $\bigcup_n \mathcal{F}_n$ par

$$\mathbb{P}(A) = \mathbb{P}_n(A), \quad A \in \mathcal{F}_n. \quad (1.7)$$

Il s'agit de prolonger \mathbb{P} en une probabilité sur $\sigma(\bigcup_n \mathcal{F}_n)$. Remarquons que $\bigcup_n \mathcal{F}_n$ étant stable par intersection finie, ce prolongement sera unique. L'existence de ce prolongement a été montrée par Kolmogorov et on a :

Théorème 1.68. *Soit $(\mu_n)_{n \geq 0}$ une famille de probabilités sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ vérifiant (1.6). Il existe une unique probabilité \mathbb{P} sur l'espace canonique (Ω, \mathcal{F}) défini par (1.7) telle que $(\Omega, \mathcal{F}, (X_n)_{n \geq 0}, \mathbb{P})$ soit un processus de répartitions finies $(\mu_n)_{n \geq 0}$.*

Exemple 1. Soient $\nu_0, \dots, \nu_n \dots$ une suite de probabilités sur (E, \mathcal{E}) . On veut construire un modèle pour une suite de v.a. indépendantes de lois $\nu_0, \dots, \nu_n, \dots$. On définit μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$ par $\mu_n = \nu_0 \otimes \dots \otimes \nu_n$ et on applique le Théorème 1.68. On obtient une probabilité \mathbb{P} sur (Ω, \mathcal{F}) telle que $(X_n)_n$ soit une suite de v.a. indépendantes de loi $\nu_0, \dots, \nu_n, \dots$.

Exemple 2. Cet exemple fournit la construction des chaînes de Markov. On considère un ensemble E dénombrable muni d'une probabilité μ et d'une matrice de transition $Q(x, y)$, $x, y \in E$, c'est à dire une matrice à termes positifs telle que pour tous $x, y \in E$,

$$Q(x, y) \geq 0, \quad \sum_{y \in E} Q(x, y) = 1.$$

On définit μ_n sur E^{n+1} par $\mu_n(x_0, x_1, \dots, x_n) = \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n)$ et on applique le Théorème 1.68. On obtient une probabilité \mathbb{P}_μ sur (Ω, \mathcal{F}) telle que les vecteurs (X_0, X_1, \dots, X_n) aient pour loi μ_n .

Chapitre 2

Espérance conditionnelle

Nous allons introduire un des outils fondamentaux des probabilités qui sera en particulier nécessaire au chapitre prochain pour l'étude des martingales.

Introduction

Cette introduction intuitive peut être omise si on souhaite aborder directement la construction mathématique de l'espérance conditionnelle.

La notion de conditionnement apparaît en probabilité chaque fois que l'on dispose d'informations partielles supplémentaires telles que la réalisation de certains événements.

On a jusque-là défini l'espérance conditionnelle d'une variable aléatoire X (positive ou intégrable) sachant un événement A tel que $\mathbb{P}(A) > 0$:

$$\mathbb{E}(X | A) = \frac{\mathbb{E}(X \mathbf{1}_A)}{\mathbb{P}(A)}.$$

C'est la moyenne des valeurs de X parmi les résultats de l'expérience aléatoire pour lesquels A est réalisé. Cela peut aussi se comprendre comme la moyenne des valeurs de X pour l'expérience modifiée par la donnée de l'information que A est réalisé.

On souhaite étendre cette notion d'espérance conditionnelle dans le cas où l'on dispose d'une information plus riche que la réalisation d'un événement, à savoir la réalisation de toute une famille d'événements. Cependant, au lieu de définir l'espérance de X sachant que ces événements sont réalisés, il sera plus pertinent de définir l'espérance de X sachant **si** ces événements sont réalisés, autrement dit ce sera une **fonction** qui dépend de la réalisation ou non de ces événements ; ceci étant aléatoire, ce sera donc une **variable aléatoire**. On aura ainsi typiquement

$$\text{“L'espérance de } X \text{ sachant si } A \text{ est réalisé”} = \begin{cases} \mathbb{E}(X | A) & \text{si } A \text{ est réalisé} \\ \mathbb{E}(X | A^c) & \text{sinon.} \end{cases}$$

Disposer d'une information sur l'expérience, cela revient à savoir si les événements d'un ensemble $\mathcal{G} \subset \mathcal{F}$ sont réalisés ou non. Remarquons que, si on sait si A est réalisé, on sait aussi si A^c est réalisé (à savoir, si A ne l'est pas), et si on sait si A_n est réalisé pour tout $n \geq 0$, on sait aussi si $\bigcup_n A_n$ est réalisé (en vérifiant un à un si l'un des A_n est réalisé). On constate donc que l'ensemble des événements dont on peut connaître la réalisation est stable par complémentaire et par union dénombrable. De plus, on sait toujours si \emptyset et Ω sont réalisés (jamais et toujours, respectivement). On constate donc que \mathcal{G} forme une tribu.

Pour toute tribu $\mathcal{G} \subset \mathcal{F}$, l'espérance conditionnelle $\mathbb{E}(X | \mathcal{G})$ sera ainsi une **variable aléatoire** qui dépendra seulement du fait que les événements de \mathcal{G} sont réalisés ou non, ce qui formellement signifie que ce sera une variable aléatoire \mathcal{G} -mesurable. L'exemple précédent correspond ainsi à $\mathbb{E}(X | \sigma(A))$ où $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$ est la plus petite tribu contenant A .

Un cas important sera celui de l'espérance de X sachant une autre variable aléatoire Y . Cela revient à connaître la réalisation de tous les événements de la forme $\{Y \in B\}$ (avec B mesurable),

autrement dit à connaître la tribu $\sigma(Y)$ engendrée par Y . Alors $\mathbb{E}(X|Y)$ sera $\sigma(Y)$ -mesurable, c'est-à-dire (par le Lemme 0.9) une fonction mesurable de Y . Notons que l'exemple précédent correspond aussi à $\mathbb{E}(X|\mathbf{1}_A)$ car $\sigma(\mathbf{1}_A) = \{\emptyset, A, A^c, \Omega\} = \sigma(A)$.

2.1 Définition

Soit un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Rappelons d'abord une propriété simple de l'espérance :

Proposition 2.1. *Soit X une variable aléatoire réelle, de carré intégrable. L'espérance $\mathbb{E}(X)$ de X est le réel qui minimise la fonction $a \mapsto \mathbb{E}((X - a)^2)$:*

$$\min_{a \in \mathbb{R}} \mathbb{E}((X - a)^2) = \mathbb{E}((X - \mathbb{E}(X))^2) = \text{Var } X.$$

Preuve : Il suffit de développer $\mathbb{E}((X - a)^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 = (a - \mathbb{E}(X))^2 + \text{Var } X \geq \text{Var } X$, et de noter qu'il y a égalité si, et seulement si $a = \mathbb{E}(X)$. \square

Autrement dit $\mathbb{E}(X)$ est la constante qui approche le mieux X , au sens de la norme L^2 . C'est donc la projection orthogonale de X sur le sous-espace $\mathbb{R}1$ des variables aléatoires constantes. C'est de cette façon de définir l'espérance que l'on part pour introduire l'espérance conditionnelle.

On donne la définition en deux étapes. D'abord lorsque $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ (rappelons que c'est un espace de Hilbert) :

2.1.1 Cas des variables aléatoires dans L^2

Définition 2.2. *Soit $\mathcal{G} \subset \mathcal{F}$ une tribu. Pour toute variable aléatoire réelle X de carré intégrable, c'est-à-dire dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, l'espérance conditionnelle de X sachant \mathcal{G} , notée $\mathbb{E}(X|\mathcal{G})$, est la projection orthogonale de X sur le sous-espace fermé $L^2(\Omega, \mathcal{G}, \mathbb{P})$ des variables aléatoires \mathcal{G} -mesurables et de carré intégrable. C'est donc l'unique variable aléatoire de carré intégrable telle que :*

(i) $\mathbb{E}(X|\mathcal{G})$ est \mathcal{G} -mesurable

(ii) pour toute variable aléatoire $Z \in L^2$ qui est \mathcal{G} -mesurable, $X - \mathbb{E}(X|\mathcal{G}) \perp Z$, c'est-à-dire

$$\mathbb{E}(ZX) = \mathbb{E}(Z\mathbb{E}(X|\mathcal{G})).$$

Comme la projection est linéaire, $\mathbb{E}(\cdot|\mathcal{G})$ est linéaire sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$. En tant que projection orthogonale, elle est contractante :

$$\text{pour toute } X \in L^2, \quad \|\mathbb{E}(X|\mathcal{G})\|_2 \leq \|X\|_2.$$

Remarquons que, comme l'espérance usuelle, elle est aussi croissante et vérifie l'inégalité triangulaire :

Lemme 2.3. *Soit $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

a) *Si $X \geq 0$ p.s., alors $\mathbb{E}(X|\mathcal{G}) \geq 0$ p.s..*

b) *Si $X \leq Y$ p.s., alors $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$ p.s..*

c) *$|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ p.s..*

Preuve : a) Posons $Y = \mathbb{E}(X|\mathcal{G})$. On a $\{Y < 0\} \in \mathcal{G}$ donc $\mathbf{1}_{\{Y < 0\}}$ est \mathcal{G} -mesurable, et bornée donc dans L^2 , d'où par (ii)

$$0 \leq \mathbb{E}(X\mathbf{1}_{\{Y < 0\}}) = \mathbb{E}(Y\mathbf{1}_{\{Y < 0\}}) \leq 0,$$

ce qui implique $\mathbb{E}(Y\mathbf{1}_{\{Y < 0\}}) = 0$, or la v.a. $Y\mathbf{1}_{\{Y < 0\}}$ est négative, et strictement négative sur $\{Y < 0\}$, donc nécessairement $\mathbb{P}(Y < 0) = 0$, comme annoncé.

b) se déduit de a) par linéarité en l'appliquant à $Y - X$.

c) se déduit de b) : comme $X \leq |X|$, $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(|X||\mathcal{G})$, de même $-X \leq |X|$ donne $-\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(|X||\mathcal{G})$, d'où c). \square

Donnons quelques propriétés spécifiques à l'espérance conditionnelle.

Lemme 2.4. Soit $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

- a) Si X est \mathcal{G} -mesurable, alors $\mathbb{E}(X | \mathcal{G}) = X$ p.s..
- b) Pour tout réel c , $\mathbb{E}(c | \mathcal{G}) = c$ p.s..
- c) Si X est indépendante de \mathcal{G} , alors $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(X)$ p.s..
- d) $\mathbb{E}(\mathbb{E}(X | \mathcal{G})) = \mathbb{E}(X)$.
- e) $\mathbb{E}(|\mathbb{E}(X | \mathcal{G})|) \leq \mathbb{E}(|X|)$, c'est-à-dire $\|\mathbb{E}(X | \mathcal{G})\|_1 \leq \|X\|_1$.

Preuve : a) est immédiat car la projection sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est l'identité sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

b) est un cas particulier de a) (une v.a. constante est mesurable pour toute tribu).

c) Si X est indépendante de \mathcal{G} , alors $X - \mathbb{E}(X)$ est orthogonale à $L^2(\Omega, \mathcal{G}, \mathbb{P})$: pour toute $Z \in L^2$ qui est \mathcal{G} -mesurable, X est indépendante de Z donc $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z)$ et donc $\mathbb{E}((X - \mathbb{E}(X))Z) = 0$. Par suite, $\mathbb{E}(X - \mathbb{E}(X) | \mathcal{G}) = 0$, donc $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(\mathbb{E}(X) | \mathcal{G}) = \mathbb{E}(X)$ ($\mathbb{E}(X)$ est une constante).

d) est (ii) pour $Z = 1$.

e) s'obtient en prenant l'espérance de la propriété c) du précédent lemme et en appliquant d). \square

La dernière propriété ci-dessus montre que $\mathbb{E}(\cdot | \mathcal{G})$ est contractante sur $L^2 \subset L^1$, pour la norme L^1 . Or L^2 est dense dans L^1 (par exemple, le théorème de convergence dominée montre que, pour toute variable aléatoire $X \in L^1$, $\mathbb{E}(|X - X\mathbf{1}_{\{|X| \leq n\}}|) \rightarrow 0$, en dominant par $|X|$, or $X\mathbf{1}_{\{|X| \leq n\}}$ est bornée donc dans L^2). Il en résulte que $\mathbb{E}(\cdot | \mathcal{G})$ s'étend par continuité de façon unique à L^1 . On va en fait utiliser (i) et (ii) pour introduire une définition un peu plus générale car elle inclut toutes les variables positives (c'est-à-dire à valeurs dans $[0, \infty]$ p.s.).

2.1.2 Cas général : variables aléatoires positives ou dans L^1

Théorème 2.5. (Théorème et définition) : Pour toute variable aléatoire réelle X positive (resp. intégrable), et pour toute sous-tribu \mathcal{G} de \mathcal{F} , il existe une variable aléatoire $\mathbb{E}(X | \mathcal{G})$, unique à égalité presque sûre près, positive (resp. intégrable), telle que :

- (i) $\mathbb{E}(X | \mathcal{G})$ est \mathcal{G} -mesurable ;
- (ii) pour tout $A \in \mathcal{G}$, $\mathbb{E}(\mathbf{1}_A X) = \mathbb{E}(\mathbf{1}_A \mathbb{E}(X | \mathcal{G}))$.

La v.a. $\mathbb{E}(X | \mathcal{G})$ s'appelle l'**espérance conditionnelle de X sachant \mathcal{G}** . Elle est définie à égalité presque sûre près ; on considérera le plus souvent la classe de v.a. correspondante (c'est-à-dire que l'on identifie entre elles les variables égales presque sûrement).

Notons que la condition (ii) est équivalente à : pour toute v.a. Z qui est \mathcal{G} -mesurable et positive (resp. bornée, ou telle que ZX est intégrable),

$$(ii') \quad \mathbb{E}(ZX) = \mathbb{E}(Z \mathbb{E}(X | \mathcal{G})).$$

En effet, (ii) en est le cas particulier $Z = \mathbf{1}_A$, et (ii) implique (ii') par linéarité (si Z est étagée) et par approximation croissante (si Z est positive, $Z = \lim_n Z_n$ avec $0 \leq Z_1 \leq Z_2 \leq \dots$ étagées positives, et on utilise le théorème de convergence monotone ; et dans le cas où ZX est intégrable, on applique le théorème à Z_+ et Z_- qui sont positives, puis $Z = Z_+ - Z_-$).

Le cas particulier suivant est essentiel :

Définition 2.6. Pour toute variable aléatoire X positive (resp. intégrable), et pour toute variable aléatoire Y à valeurs dans un espace mesurable (E, \mathcal{E}) , on note

$$\mathbb{E}(X | Y) = \mathbb{E}(X | \sigma(Y)).$$

C'est donc (via le Lemme 0.9) l'unique variable aléatoire de la forme $\mathbb{E}(X | Y) = h(Y)$ (avec $h : E \rightarrow \mathbb{R}$ mesurable) telle que, pour toute fonction mesurable $g : E \rightarrow \mathbb{R}$ positive (resp. bornée),

$$\mathbb{E}(g(Y)X) = \mathbb{E}(g(Y)\mathbb{E}(X | Y)).$$

On pourra bien sûr considérer en particulier $\mathbb{E}(X | X_1, \dots, X_n) = \mathbb{E}(X | \sigma(X_1, \dots, X_n))$, qui est une fonction mesurable de X_1, \dots, X_n .

Preuve du Théorème 2.5. Unicité. Démontrons d'abord l'unicité dans le cas $X \in L^1$. Soit Y et \tilde{Y} deux variables aléatoires intégrables vérifiant (i) et (ii) : Y et \tilde{Y} sont \mathcal{G} -mesurables et, pour tout $A \in \mathcal{G}$, $\mathbb{E}(Y\mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(\tilde{Y}\mathbf{1}_A)$. Prenant $A = \{Y < \tilde{Y}\}$, on a $A \in \mathcal{G}$ et donc $\mathbb{E}(Y\mathbf{1}_A) = \mathbb{E}(\tilde{Y}\mathbf{1}_A)$, d'où $\mathbb{E}((\tilde{Y} - Y)\mathbf{1}_A) = 0$, or $(\tilde{Y} - Y)\mathbf{1}_A$ est nulle hors de A et > 0 sur A , donc on doit avoir $\mathbb{P}(A) = 0$, c'est-à-dire $Y \geq \tilde{Y}$ p.s.. Par symétrie, $Y = \tilde{Y}$ p.s..

Considérons maintenant l'unicité dans le cas $X \geq 0$. Pour tout $n \in \mathbb{N}$, on note $A_n = \{Y < \tilde{Y} \leq n\}$. On a $A_n \in \mathcal{G}$ par (i) et donc

$$0 \leq \mathbb{E}(Y\mathbf{1}_{A_n}) = \mathbb{E}(\tilde{Y}\mathbf{1}_{A_n}) \leq n < \infty,$$

d'où $Y\mathbf{1}_{A_n}, \tilde{Y}\mathbf{1}_{A_n} \in L^1$ et $\mathbb{E}((\tilde{Y} - Y)\mathbf{1}_{A_n}) = 0$, or $(\tilde{Y} - Y)\mathbf{1}_{A_n} > 0$ sur A_n et $= 0$ ailleurs, donc $\mathbb{P}(A_n) = 0$. La suite $(A_n)_n$ est croissante, d'où

$$0 = \lim_n \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_n A_n\right) = \mathbb{P}(Y < \tilde{Y} < \infty).$$

Par symétrie, $\mathbb{P}(\tilde{Y} < Y < \infty) = 0$, donc p.s., si $Y, \tilde{Y} < \infty$, alors $Y = \tilde{Y}$. Il reste à voir que, p.s., $Y = \infty$ si et seulement si $\tilde{Y} = \infty$. Or, avec $B_n = \{Y < n, \tilde{Y} = \infty\} \in \mathcal{G}$, on a

$$\infty \cdot \mathbb{P}(B_n) = \mathbb{E}(\tilde{Y}\mathbf{1}_{B_n}) = \mathbb{E}(Y\mathbf{1}_{B_n}) \leq n,$$

ce qui impose $\mathbb{P}(B_n) = 0$. La suite $(B_n)_n$ est croissante, d'où

$$0 = \mathbb{P}\left(\bigcup_n B_n\right) = \mathbb{P}(Y < \infty, \tilde{Y} = \infty).$$

Par symétrie, on a aussi $\mathbb{P}(\tilde{Y} < \infty, Y = \infty) = 0$. Finalement, $Y = \tilde{Y}$ p.s..

Existence. On a déjà vu l'existence dans le cas où $X \in L^2$. On va en déduire le cas général par approximation en tronquant X . Définissons, pour tout $n \in \mathbb{N}$, $X_n = X\mathbf{1}_{\{|X| \leq n\}} \in L^2$ et $Y_n = \mathbb{E}(X_n | \mathcal{G})$.

Dans le cas $X \geq 0$, on constate que $(X_n)_n$ est une suite croissante de variables aléatoires positives p.s. et donc (par un lemme précédent), qu'il en va de même de la suite $(Y_n)_n$. On peut donc définir $Y = \lim_n Y_n$ (limite croissante, dans $\overline{\mathbb{R}}$). La propriété (i) est vérifiée par Y (une limite de fonctions \mathcal{G} -mesurables est \mathcal{G} -mesurable), et la propriété (ii) pour Y s'obtient par convergence monotone :

$$\mathbb{E}(Y\mathbf{1}_A) = \mathbb{E}(\lim_n Y_n \mathbf{1}_A) = \lim_n \mathbb{E}(Y_n \mathbf{1}_A) = \lim_n \mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(\lim_n X_n \mathbf{1}_A) = \mathbb{E}(X\mathbf{1}_A).$$

Dans le cas $X \in L^1$, $(X_n)_n$ tend vers X dans L^1 , et comme on a vu que $\mathbb{E}(\cdot | \mathcal{G})$ est contractante sur L^2 pour la norme $\|\cdot\|_1$, il en résulte que $(Y_n)_n$ est de Cauchy dans L^1 donc converge dans L^1 vers une variable aléatoire Y . Comme la convergence L^1 implique la convergence p.s. d'une sous-suite, Y est \mathcal{G} -mesurable. Et la propriété (ii) pour Y s'obtient par la convergence L^1 : pour tout $A \in \mathcal{G}$, pour tout n , $\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(Y_n \mathbf{1}_A)$, et $|\mathbb{E}(X_n \mathbf{1}_A) - \mathbb{E}(X \mathbf{1}_A)| \leq \mathbb{E}(|X - X_n| \mathbf{1}_A) \leq \mathbb{E}(|X - X_n|) \rightarrow 0$, et de même pour $(Y_n)_n$ et Y , d'où à la limite $\mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(Y \mathbf{1}_A)$. \square

Exemple 2.7. Soit A un événement de \mathcal{F} de probabilité non nulle et X une variable aléatoire positive. Calculons $W = \mathbb{E}(X | \sigma(A))$.

On note que $\sigma(A) = \{A, A^c, \emptyset, \Omega\} = \sigma(\mathbf{1}_A)$, de sorte que W , qui est une fonction $\sigma(A)$ -mesurable est une fonction de $\mathbf{1}_A$, et est donc de la forme

$$W = \alpha \mathbf{1}_A + \beta \mathbf{1}_{A^c}.$$

Et en écrivant (ii) pour A et A^c , on obtient

$$\alpha = \frac{\mathbb{E}(X \mathbf{1}_A)}{\mathbb{P}(A)} = \mathbb{E}(X | A), \quad \text{et} \quad \beta = \frac{\mathbb{E}(X \mathbf{1}_{A^c})}{\mathbb{P}(A^c)} = \mathbb{E}(X | A^c),$$

donc

$$\mathbb{E}(X | \sigma(A)) = \mathbb{E}(X | A) \mathbf{1}_A + \mathbb{E}(X | A^c) \mathbf{1}_{A^c}.$$

Exemple 2.8. Soit Y une variable aléatoire discrète, à valeurs dans E dénombrable, et X une variable aléatoire positive. Calculons $W = \mathbb{E}(X | Y)$.

Comme $\mathbb{E}(X | Y)$ est $\sigma(Y)$ -mesurable, il existe $f : E \rightarrow \mathbb{R}_+$ telle que $\mathbb{E}(X | Y) = f(Y) = \sum_{y \in E} f(y) \mathbf{1}_{\{Y=y\}}$. Pour tout $y \in E$, on a alors, en écrivant (ii) pour $A = \{Y = y\}$, par TCM,

$$\mathbb{E}(X \mathbf{1}_{\{Y=y\}}) = \mathbb{E}(\mathbb{E}(X | Y) \mathbf{1}_{\{Y=y\}}) = f(y) \mathbb{P}(Y = y),$$

d'où $f(y) = \mathbb{E}(X | Y = y)$ et donc :

$$\mathbb{E}(X | Y) = \sum_{y \in E} \mathbb{E}(X | Y = y) \mathbf{1}_{\{Y=y\}}.$$

On voit ainsi que, si Y est discrète à valeurs dans E , $\mathbb{E}(X | Y) = f(Y)$ où, pour toute valeur $y \in E$, $f(y) = \mathbb{E}(X | Y = y)$ est l'espérance conditionnelle classique.

Notation abusive (mais pratique). Par extension, on utilisera quelquefois, pour n'importe quelle variable Y (même non discrète), la notation abrégée suivante : si $\mathbb{E}(X | Y) = f(Y)$, alors

$$\mathbb{E}(X | Y = y) = f(y).$$

Cette écriture, qui mathématiquement n'a pas de sens classique si $\mathbb{P}(Y = y) = 0$, facilite le formalisme de certains calculs. Par exemple, la propriété $\mathbb{E}(\mathbb{E}(X | Y)) = \mathbb{E}(X)$ (qui est (ii') avec $Z = 1$) s'écrit $\int \mathbb{E}(X | Y) d\mathbb{P} = \mathbb{E}(X)$ d'où, avec le théorème de transfert et la notation précédente,

$$\int \mathbb{E}(X | Y = y) dP_Y(y) = \mathbb{E}(X).$$

2.2 Propriétés élémentaires

On regroupe ci-dessous quelques propriétés générales de l'espérance conditionnelle.

Proposition 2.9. Soit $\mathcal{G} \subset \mathcal{F}$ une tribu, X une variable aléatoires positive, ou intégrable. On a les propriétés suivantes :

- $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}))$.
- Pour toute constante c , $\mathbb{E}(c | \mathcal{G}) = c$.
- Si X est \mathcal{G} -mesurable, $\mathbb{E}(X | \mathcal{G}) = X$.
- Si X est \mathcal{G} -mesurable, et Y est telle que XY est intégrable ou positive, $\mathbb{E}(XY | \mathcal{G}) = X \mathbb{E}(Y | \mathcal{G})$.
- Si X est indépendante de \mathcal{G} , on a $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}(X)$.
- Sur les v.a. positives, l'espérance conditionnelle est semi-linéaire (stable par addition et par multiplication par un scalaire positif).
- Sur $L^1(\Omega, \mathcal{F}, \mathbb{P})$, l'application $X \mapsto \mathbb{E}(X | \mathcal{G})$ est linéaire, positive, et donc croissante : Si X_1, X_2 sont intégrables et telles que $X_1 \leq X_2$ p.s., alors $\mathbb{E}(X_1 | \mathcal{G}) \leq \mathbb{E}(X_2 | \mathcal{G})$ p.s. En particulier,

$$|\mathbb{E}(X | \mathcal{G})| \leq \mathbb{E}(|X| | \mathcal{G}).$$

et donc $\|\mathbb{E}(X | \mathcal{G})\|_1 \leq \|X\|_1$, ce qui signifie que $\mathbb{E}(\cdot | \mathcal{G})$ est contractante donc continue.

- Sur $L^2(\Omega, \mathcal{F}, \mathbb{P})$, l'application $X \mapsto \mathbb{E}(X | \mathcal{G})$ est la projection orthogonale sur $L^2(\Omega, \mathcal{G}, \mathbb{P})$, en particulier elle est contractante : pour $X \in L^2$, $\|\mathbb{E}(X | \mathcal{G})\|_2 \leq \|X\|_2$.

Toutes ces propriétés sont des conséquences simples de la définition, ou ont été vues dans la partie précédente dans L^2 et peuvent s'étendre par approximation (à titre d'exercice, démontrer notamment la propriété d) en vérifiant directement (i) et (ii)). Attention, bien qu'on ne le précise pas systématiquement, les (in)égalités des propriétés ci-dessus sont valables *presque sûrement*, car l'espérance conditionnelle sachant \mathcal{G} est une variable aléatoire, qui de plus n'est définie qu'à égalité presque partout près.

Ces propriétés peuvent se retenir en remarquant qu'elles satisfont à l'intuition suivante : $\mathbb{E}(\cdot | \mathcal{G})$ se calcule comme une espérance dans laquelle on "considère comme constantes" les variables \mathcal{G} -mesurables. Notons que considérer ces variables aléatoires comme constantes altère en général la loi des autres variables aléatoires... à moins qu'elles ne soient indépendantes de \mathcal{G} :

Proposition 2.10. *Soient X et Y deux v.a. et \mathcal{G} une sous-tribu de \mathcal{F} . Soit ϕ une fonction borélienne, positive ou telle que $\phi(X, Y)$ soit intégrable.*

a) *On suppose que X est \mathcal{G} -mesurable et Y est indépendante de la tribu \mathcal{G} . Alors*

$$\mathbb{E}(\phi(X, Y) | \mathcal{G}) = \psi(X), \quad p.s.,$$

$$\text{où } \psi(x) := \mathbb{E}(\phi(x, Y)).$$

b) *On suppose que X et Y sont indépendantes. Alors*

$$\mathbb{E}(\phi(X, Y) | X) = \psi(X), \quad p.s.,$$

$$\text{où } \psi(x) := \mathbb{E}(\phi(x, Y)).$$

Preuve : Notons que le point b) est un cas particulier du a), avec $\mathcal{G} = \sigma(X)$, donc il suffit de prouver a). On remarque que $\psi(X)$ est bien \mathcal{G} -mesurable. Pour toute v.a. $Z \geq 0$, \mathcal{G} -mesurable bornée, on constate que Y est indépendante de (X, Z) , donc on a

$$\begin{aligned} \mathbb{E}(Z \phi(X, Y)) &= \int z \phi(x, y) d\mathbb{P}_{(X, Y, Z)}(x, y, z) = \int \int z \phi(x, y) d\mathbb{P}_{(X, Z)}(x, z) d\mathbb{P}_Y(y) \\ &= \int z \left(\int \phi(x, y) d\mathbb{P}_Y(y) \right) d\mathbb{P}_{(X, Z)}(x, z) = \int z \psi(x) d\mathbb{P}_{(X, Z)}(x, z) = \mathbb{E}(Z \psi(X)), \end{aligned}$$

ce qui achève de prouver a). □

Conditionner deux fois revient à conditionner par la tribu la plus petite :

Proposition 2.11 (Double conditionnement). *Si $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, pour toute v.a. X intégrable ou positive, on a*

$$\mathbb{E}(\mathbb{E}(X | \mathcal{G}_1) | \mathcal{G}_2) = \mathbb{E}(X | \mathcal{G}_1) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}_2) | \mathcal{G}_1) \quad p.s.$$

Preuve : Comme $\mathcal{G}_1 \subset \mathcal{G}_2$, $\mathbb{E}(X | \mathcal{G}_1)$ étant \mathcal{G}_1 -mesurable est aussi \mathcal{G}_2 -mesurable et donc

$$\mathbb{E}(\mathbb{E}(X | \mathcal{G}_1) | \mathcal{G}_2) = \mathbb{E}(X | \mathcal{G}_1).$$

Pour l'autre égalité, on note que $\mathbb{E}(X | \mathcal{G}_1)$ est \mathcal{G}_1 -mesurable et, pour tout $A \in \mathcal{G}_1$, en écrivant le point (ii) de la définition de $\mathbb{E}(X | \mathcal{G}_1)$ puis de la définition de $\mathbb{E}(X | \mathcal{G}_2)$ (car $A \in \mathcal{G}_2$ aussi), on a :

$$\mathbb{E}(\mathbb{E}(X | \mathcal{G}_1) \mathbf{1}_A) = \mathbb{E}(X \mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X | \mathcal{G}_2) \mathbf{1}_A).$$

Ceci prouve que $\mathbb{E}(\mathbb{E}(X | \mathcal{G}_2) | \mathcal{G}_1) = \mathbb{E}(X | \mathcal{G}_1)$ p.s. □

Notons que, sans l'inclusion, il n'y a pas de formule générale (les projections orthogonales ne commutent pas, en général).

Comme on l'a déjà vu, l'espérance conditionnelle se comporte, par bien des propriétés, comme une espérance. On en donne deux exemples de plus ci-dessous.

Proposition 2.12 (TCM pour l'espérance conditionnelle). *Pour toute suite croissante $(X_n, n \geq 0)$ de v.a. positives, on a :*

$$\lim_{n \rightarrow \infty} \uparrow \mathbb{E}(X_n | \mathcal{G}) = \mathbb{E}(\lim_{n \rightarrow \infty} \uparrow X_n | \mathcal{G}), \quad p.s.$$

Preuve : La mesurabilité étant préservée par limite de suite, le résultat s'obtient par passage à la limite (croissante) dans la relation (ii). □

Proposition 2.13 (Inégalité de Jensen). *Soit X une v.a. réelle intégrable. Si $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction convexe positive ou telle que $\varphi(X)$ est intégrable, on a*

$$\varphi(\mathbb{E}(X | \mathcal{G})) \leq \mathbb{E}(\varphi(X) | \mathcal{G}), \quad \text{p.s.}$$

Preuve : La fonction φ étant convexe, son graphe est l'intersection des demi-espaces qui le contiennent :

$$\text{pour tout } x \in \mathbb{R}, \quad \varphi(x) = \max\{ax + b \mid (a, b) \in E_\varphi\}$$

où

$$E_\varphi = \{(a, b) \in \mathbb{R}^2 \mid \forall x \in \mathbb{R}, \varphi(x) \geq ax + b\}.$$

On peut vérifier qu'on a aussi

$$\text{pour tout } x \in \mathbb{R}, \quad \varphi(x) = \sup\{ax + b \mid (a, b) \in E_\varphi \cap \mathbb{Q}^2\}.$$

Or, pour tous $(a, b) \in E_\varphi \cap \mathbb{Q}^2$, $\varphi(X) \geq aX + b$ donc

$$\mathbb{E}(\varphi(X) | \mathcal{G}) \geq a\mathbb{E}(X | \mathcal{G}) + b \quad \text{p.s.}$$

Comme $E_\varphi \cap \mathbb{Q}^2$ est dénombrable, on a aussi : p.s.,

$$\text{pour tous } (a, b) \in E_\varphi \cap \mathbb{Q}^2, \quad \mathbb{E}(\varphi(X) | \mathcal{G}) \geq a\mathbb{E}(X | \mathcal{G}) + b,$$

donc

$$\mathbb{E}(\varphi(X) | \mathcal{G}) \geq \sup_{(a,b) \in E_\varphi \cap \mathbb{Q}^2} (a\mathbb{E}(X | \mathcal{G}) + b) = \varphi(\mathbb{E}(X | \mathcal{G})).$$

□

2.3 Espérance sachant une v.a. discrète. Loi conditionnelle

On a vu (exemple 2.8) que, si Y est une variable aléatoire discrète, c'est-à-dire à valeurs dans un espace dénombrable E (et $\mathbb{P}(Y = y) > 0$ pour tout $y \in E$, quitte à réduire E), alors pour toute v.a. X , pour toute fonction mesurable φ positive ou telle que $\varphi(X)$ est intégrable,

$$\mathbb{E}(\varphi(X) | Y) = \sum_{y \in E} \mathbb{E}(\varphi(X) | Y = y) \mathbf{1}_{\{Y=y\}},$$

où $\mathbb{E}(\varphi(X) | Y = y) = \frac{\mathbb{E}(\varphi(X) \mathbf{1}_{\{Y=y\}})}{\mathbb{P}(Y=y)}$ est l'espérance conditionnelle classique. Le calcul de l'espérance conditionnelle est donc explicite dans ce cas.

Rappelons que, pour tout $y \in E$, par le théorème de transfert,

$$\mathbb{E}(\varphi(X) | Y = y) = \int_{\Omega} \varphi(X) d\mathbb{P}(\cdot | Y = y) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_{X|Y=y}(x),$$

où $\mathbb{P}_{X|Y=y}$ est la loi de X sous $\mathbb{P}(\cdot | Y = y)$, appelée **loi conditionnelle de X sachant $Y = y$** .

On constate alors que l'espérance de $\varphi(X)$ sachant Y s'écrit comme une espérance

$$\mathbb{E}(\varphi(X) | Y) = \int_{\mathbb{R}} \varphi(x) d\mathbb{P}_{X|Y}(x),$$

par rapport à la loi *aléatoire*

$$\mathbb{P}_{X|Y}(\cdot) = \sum_{y \in E} \mathbf{1}_{\{Y=y\}} \mathbb{P}_{X|Y=y}(\cdot),$$

qui dépend de Y et est appelée **loi conditionnelle de X sachant Y** .

On vérifie alors facilement que, pour toute fonction g positive ou telle que $g(X, Y)$ est intégrable,

$$\mathbb{E}(g(X, Y) | Y) = \int_{\mathbb{R}} g(x, Y) d\mathbb{P}_{X|Y}(x)$$

d'où, en prenant l'espérance, la formule de *désintégration*

$$\mathbb{E}(g(X, Y)) = \mathbb{E}(\mathbb{E}(g(X, Y) | Y)) = \int \left(\int g(x, y) d\mathbb{P}_{X|Y=y}(x) \right) d\mathbb{P}_Y(y).$$

Signalons enfin que le cas particulier où Y et X sont toutes deux discrètes est particulièrement simple. La loi de X sachant $Y = y$ est alors donnée par les valeurs

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

et on a, pour toute fonction g positive, ou telle que $g(X, Y)$ est intégrable,

$$\mathbb{E}(g(X, Y) | Y = y) = \sum_x g(x, y) \frac{\mathbb{P}((X, Y) = (x, y))}{\mathbb{P}(Y = y)}. \quad (2.1)$$

Exemple 2.14. *Tirages sans remise.* Soit un entier $N \geq 2$. On considère une v.a. $W = (X, Y)$ de loi uniforme dans l'ensemble des couples d'entiers distincts entre 1 et N :

$$\mathcal{A}_2^N = \{(k, l) \in \{1, \dots, N\}^2 | k \neq l\}.$$

Pour tout $(k, l) \in \mathcal{A}_2^N$, on a

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(W = (k, l)) = \frac{1}{\text{Card } \mathcal{A}_2^N} = \frac{1}{N(N-1)}$$

donc, pour $1 \leq k \leq N$,

$$\mathbb{P}(X = k) = \sum_{l=1}^N \mathbb{P}(X = k, Y = l) = \sum_{l \neq k} \frac{1}{N(N-1)} = \frac{1}{N}$$

et, pour tout $l \neq k$ dans $\{1, \dots, N\}$,

$$\mathbb{P}(Y = l | X = k) = \frac{\mathbb{P}(X = k, Y = l)}{\mathbb{P}(X = k)} = \frac{1}{N-1}.$$

Autrement dit, la loi de X est uniforme dans $\{1, \dots, N\}$ et pour $k = 1, \dots, N$, la loi de Y sachant $\{X = k\}$ est uniforme dans $\{1, \dots, N\} \setminus \{k\}$: la loi de Y sachant X est uniforme dans $\{1, \dots, N\} \setminus \{X\}$.

2.4 Cas des lois à densité. Loi conditionnelle sachant Y

Un second cas important où une formule explicite peut être donnée pour $\mathbb{E}(g(X, Y) | Y)$ est le cas où le couple (X, Y) , à valeurs dans \mathbb{R}^2 , admet une densité.

Rappelons que, si (X, Y) a pour densité $f_{(X,Y)}$ sur \mathbb{R}^2 , alors X et Y ont des densités f_X et f_Y données par

$$f_X(x) = \int f_{(X,Y)}(x, y) dy \quad \text{et} \quad f_Y(y) = \int f_{(X,Y)}(x, y) dx.$$

Proposition 2.15. Soient X et Y deux v.a. réelles telles que le couple (X, Y) ait une densité $f_{(X,Y)}$ sur \mathbb{R}^2 : $d\mathbb{P}_{(X,Y)}(x, y) = f_{(X,Y)}(x, y) dx dy$. Alors, pour toute fonction g positive ou telle que $g(X, Y)$ est intégrable,

$$\mathbb{E}(g(X, Y) | Y) = \psi(Y) \quad \text{p.s.}, \quad \text{où} \quad \psi(y) = \int_{\mathbb{R}} g(x, y) f_{X|Y=y}(x) dx,$$

avec

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

La fonction $f_{X|Y=y}$ est appelée la densité conditionnelle de X sachant $Y = y$.

Signalons que, p.s., $f_Y(Y) > 0$, car la probabilité $\mathbb{P}(f_Y(Y) = 0)$ est l'intégrale de f sur l'ensemble $\{y \in \mathbb{R} \mid f_Y(y) = 0\}$, donc est nulle. Cela justifie qu'il n'est pas nécessaire de définir $\phi(y)$, et donc $f_{X|Y=y}$, lorsque $f_Y(y) = 0$.

Ce résultat se généralise immédiatement au cas où X et Y sont à valeurs dans \mathbb{R}^m et \mathbb{R}^n .

Preuve : Vérifions la relation (ii'). Pour toute fonction mesurable $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} \mathbb{E}[g(X, Y)h(Y)] &= \int_{\mathbb{R}^2} g(x, y)h(y)f_{(X, Y)}(x, y)dx dy = \int_{\mathbb{R}} h(y)f_Y(y) \left(\int_{\mathbb{R}} g(x, y) \frac{f_{(X, Y)}(x, y)}{f_Y(y)} dx \right) dy \\ &= \int_{\mathbb{R}} \psi(y)h(y)f_Y(y)dy = \mathbb{E}[\psi(Y)h(Y)], \end{aligned}$$

(par la remarque précédente, la première intégrale peut se ramener, sans changer sa valeur, au domaine $\mathbb{R} \times \{f_Y(\cdot) > 0\}$, ce qui rend possible la division par $f_Y(y)$). C'est la relation attendue. \square

Notons que la formule peut se réécrire, avec l'écriture abusive déjà introduite, sous la forme pratique suivante, à rapprocher de (2.1) :

$$\mathbb{E}(g(X, Y) \mid Y = y) = \int g(x, y)f_{X|Y=y}(x)dx.$$

Pour $y \in \mathbb{R}$, la loi μ_y de densité $f_{X|Y=y}$ est appelée **loi conditionnelle de X sachant $Y = y$** . On a en effet en particulier, pour toute fonction mesurable φ positive ou telle que $\varphi(X)$ est intégrable,

$$\mathbb{E}(\varphi(X) \mid Y = y) = \int \varphi(x)f_{X|Y=y}(x)dx.$$

La loi (aléatoire) μ_Y est appelée **loi conditionnelle de X sachant Y** .

En intégrant la formule de la proposition, on obtient la formule de *désintégration* :

$$\mathbb{E}(g(X, Y)) = \mathbb{E}(\mathbb{E}(g(X, Y) \mid Y)) = \int \left(\int g(x, y)f_{X|Y=y}(x)dx \right) f_Y(y)dy.$$

2.5 Propriété de Markov forte et ses applications

Signalons d'abord que l'espérance conditionnelle permet de donner une nouvelle expression pour la propriété de Markov. Nous reprenons pour cela les notations du chapitre 1.

Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov sur E . On note $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$ la tribu engendrée par (X_0, X_1, \dots, X_n) . Le théorème suivant est une réexpression de la propriété de Markov simple (sous la forme de la Proposition 1.20) en terme d'espérance conditionnelle :

Théorème 2.16 (Propriété de Markov simple). *Pour toute fonction mesurable bornée ou positive $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$, et pour toute loi initiale μ , on a, pour tout $n \geq 0$,*

$$\mathbb{E}_{\mu}(f(X_n, X_{n+1}, \dots) \mid \mathcal{F}_n) = \mathbb{E}_{X_n}(f(X_0, X_1, \dots)).$$

Notons que, pour lever toute ambiguïté, le terme de droite devrait, comme dans le Chapitre 1, s'écrire $F(X_n)$ avec, pour tout $x \in E$, $F(x) = \mathbb{E}_x(f(X_0, X_1, \dots))$.

2.5.1 Temps d'arrêt et propriété de Markov forte

Définition 2.17. Une **filtration** (à temps discret) est une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$, de sous-tribus de \mathcal{F} .

Un **temps d'arrêt** de cette filtration est une variable aléatoire $\tau : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ telle que

$$\text{pour tout } n \in \mathbb{N}, \quad \{\tau \leq n\} \in \mathcal{F}_n.$$

La tribu \mathcal{F}_{τ} du passé avant τ est

$$\mathcal{F}_{\tau} = \{A \in \mathcal{F} \mid \text{pour tout } n \in \mathbb{N}, A \cap \{\tau \leq n\} \in \mathcal{F}_n\}.$$

Lemme 2.18. Une variable aléatoire τ à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un temps d'arrêt pour $(\mathcal{F}_n)_{n \geq 0}$ si, et seulement si, pour tout $n \in \mathbb{N}$, $\{\tau = n\} \in \mathcal{F}_n$.

Preuve : En effet, si τ est un temps d'arrêt, $\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\} \in \mathcal{F}_n$ (car $\{\tau \leq n\} \in \mathcal{F}_n$, $\{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$ et \mathcal{F}_n est une tribu). Et si la propriété du lemme est vérifiée, alors pour tout $n \in \mathbb{N}$ on a $\{\tau \leq n\} = \bigcup_{k \leq n} \{\tau = k\} \in \mathcal{F}_n$ car $\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $k \leq n$, donc τ est un temps d'arrêt. \square

On vérifiera à titre d'exercice les propriétés suivantes (où on note $a \vee b = \max(a, b)$ et $a \wedge b = \min(a, b)$) :

Proposition 2.19. a) Si τ et σ sont deux temps d'arrêts, $\sigma \wedge \tau$, $\sigma \vee \tau$, $\sigma + \tau$ le sont aussi.

b) Soit (X_n) un processus tel que, pour tout n , X_n est \mathcal{F}_n mesurable (on dira que le processus X est adapté à la filtration (\mathcal{F}_n)), alors τ et $X_\tau \mathbf{1}_{(\tau < \infty)}$ sont \mathcal{F}_τ -mesurables.

Le résultat suivant étend la propriété de Markov simple aux temps d'arrêt :

Théorème 2.20 (Propriété de Markov forte). Pour tout temps d'arrêt T , pour toute fonction mesurable $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$, bornée ou positive, pour toute loi initiale μ ,

$$\text{sur l'événement } \{T < +\infty\}, \quad \mathbb{E}_\mu(f(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T) = \mathbb{E}_{X_T}(f(X_0, X_1, \dots)).$$

Preuve : Soit $H \in \mathcal{F}_T$. On note, pour $x \in E$, $F(x) = \mathbb{E}_x(f(X_0, X_1, \dots))$. Comme $F(X_T)$ est \mathcal{F}_T -mesurable, il suffit de vérifier que

$$\mathbb{E}_\mu\left(f(X_T, X_{T+1}, \dots) \mathbf{1}_H \mathbf{1}_{(T < \infty)}\right) = \mathbb{E}_\mu\left(F(X_T) \mathbf{1}_H \mathbf{1}_{(T < \infty)}\right).$$

On décompose le membre de gauche selon la valeur de T :

$$\begin{aligned} \mathbb{E}_\mu\left(f(X_T, X_{T+1}, \dots) \mathbf{1}_H \mathbf{1}_{(T < \infty)}\right) &= \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(f(X_T, X_{T+1}, \dots) \mathbf{1}_H \mathbf{1}_{(T=n)}\right) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(f(X_n, X_{n+1}, \dots) \mathbf{1}_H \mathbf{1}_{(T=n)}\right) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(F(X_n) \mathbf{1}_H \mathbf{1}_{(T=n)}\right) \quad (\text{propriété de Markov simple au temps } n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_\mu\left(F(X_T) \mathbf{1}_H \mathbf{1}_{(T=n)}\right) \\ &= \mathbb{E}_\mu\left(F(X_T) \mathbf{1}_H \mathbf{1}_{(T < \infty)}\right), \end{aligned}$$

d'où le théorème. \square

On signale un cas particulier très courant :

Corollaire 2.21 (Propriété de Markov forte (version simplifiée)). On note μ la loi initiale de $(X_n)_{n \geq 0}$. Soit T un temps d'arrêt de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ sur E tel que :

- p.s., $T < \infty$;
- il existe $x \in E$ tel que, p.s., $X_T = x$.

Alors sous \mathbb{P}_μ , le processus (X_T, X_{T+1}, \dots) est indépendant de la tribu \mathcal{F}_T et a la même loi que la chaîne (X_0, X_1, \dots) sous \mathbb{P}_x . En d'autres termes, pour toute fonction mesurable $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$, par exemple bornée ou positive, et pour tout événement $H \in \mathcal{F}_T$, on a

$$\mathbb{E}_\mu\left(f(X_T, X_{T+1}, \dots) \mathbf{1}_H\right) = \mathbb{P}_\mu(H) \mathbb{E}_x\left(f(X_0, X_1, \dots)\right).$$

Rappelons l'opérateur de décalage, ou *shift*, $\theta : \Omega \rightarrow \Omega$ mesurable tel que, pour tout $n \in \mathbb{N}$,

$$X_n \circ \theta = X_{n+1}.$$

On pose $\theta_0 = \text{identité}$, $\theta_1 = \theta$, $\theta_2 = \theta \circ \theta$, \dots , $\theta_{n+1} = \theta_n \circ \theta$. Avec ces notations, la propriété de Markov forte s'écrit comme suit :

Proposition 2.22 (Autre forme de la propriété de Markov forte). *Pour toute v.a. $Z : \Omega \rightarrow \mathbb{R}$, $\sigma(X_k, k \in \mathbb{N})$ -mesurable, positive ou bornée, on a, sur $\{T < +\infty\}$,*

$$\mathbb{E}_\nu(Z \circ \theta_T \mid \mathcal{F}_T) = \mathbb{E}_{X_T}(Z).$$

2.5.2 Application au théorème ergodique

Rappelons la notation pour le temps de (premier) retour en $x \in E$:

$$\tau_x = \inf\{k > 0 \mid X_k = x\}.$$

On définit alors par récurrence $(\tau_x^{(r)})_{r \in \mathbb{N}}$ par $\tau_x^{(0)} = 0$ et pour tout entier $r \geq 0$,

$$\tau_x^{(r+1)} = \inf\{k > \tau_x^{(r)} \mid X_k = x\}$$

On voit que $\tau^{(1)} = \tau_x$ est le temps de retour en l'état x , puis $\tau_x^{(r)}$ est l'instant de r -ième retour en x (avec $\tau_x^{(r)} = \infty$ s'il y a moins de r retours en x). Ce sont des temps d'arrêt. Par ailleurs, on a

$$\text{pour tout } r \geq 0, \quad \tau_x^{(r+1)} = \tau_x \circ \theta_{\tau_x^{(r)}} + \tau_x^{(r)}.$$

Si x est récurrent alors, sous \mathbb{P}_x , $\tau_x^{(r)} < \infty$ pour tout $r \geq 0$, et la propriété de Markov forte nous donne le résultat admis lors de la preuve du théorème ergodique au chapitre précédent :

Théorème 2.23. *Supposons que x est un état récurrent de la chaîne de Markov $(X_n)_n$. Alors, sous la probabilité \mathbb{P}_x , pour toute $f : E \rightarrow \mathbb{R}^+$, les variables aléatoires*

$$Z_r = \sum_{k=\tau_x^{(r)}}^{\tau_x^{(r+1)}-1} f(X_k), \quad r \in \mathbb{N},$$

sont indépendantes et de même loi.

Preuve : Soit $r \in \mathbb{N}$. Remarquons que $Z_r = Z_0 \circ \theta_{\tau_x^{(r)}}$. Soit W une v.a. $\mathcal{F}_{\tau_x^{(r)}}$ -mesurable bornée, et $\psi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne. On a, par la définition de l'espérance conditionnelle,

$$\mathbb{E}_x(W\psi(Z_r)) = \mathbb{E}_x(W\mathbb{E}_x(\psi(Z_r) \mid \mathcal{F}_{\tau_x^{(r)}})) = \mathbb{E}_x(W\mathbb{E}_x(\psi(Z_0 \circ \theta_{\tau_x^{(r)}}) \mid \mathcal{F}_{\tau_x^{(r)}})).$$

Comme $\tau_x^{(r)}$ est un temps d'arrêt fini p.s. (du fait de la récurrence) et que $Z_{\tau_x^{(r)}} = x$ p.s., il résulte de la propriété de Markov forte (version simplifiée) au temps $\tau_x^{(r)}$ que

$$\mathbb{E}_x(\psi(Z_0 \circ \theta_{\tau_x^{(r)}}) \mid \mathcal{F}_{\tau_x^{(r)}}) = \mathbb{E}_x(\psi(Z_0)),$$

donc finalement

$$\mathbb{E}_x(W\psi(Z_r)) = \mathbb{E}_x(W)\mathbb{E}_x(\psi(Z_0)).$$

Ceci montre que Z_r est indépendante de la tribu $\mathcal{F}_{\tau_x^{(r)}}$ et de même loi que Z_0 . Puisque les variables aléatoires Z_0, Z_1, \dots, Z_{r-1} sont $\mathcal{F}_{\tau_x^{(r)}}$ -mesurables (à vérifier), on en déduit par récurrence sur r que les v.a. $Z_r, r \geq 0$, sont indépendantes. \square

Chapitre 3

Martingales en temps discret

3.1 Introduction

La notion de martingale peut être approchée à partir de l'exemple suivant : considérons un jeu où à chaque coup on gagne ou on perd 1 euro avec la probabilité 1/2. La suite des “gains algébriques” (c'est-à-dire positifs ou négatifs) est donnée par la suite de v.a. i.i.d. $(X_n, n \geq 1)$ telle que, pour tout $n \geq 1$,

$$\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

Soit $a_0 > 0$ la fortune initiale du joueur. Sa fortune au bout de n coups (on dira aussi, “au temps n ”) sera la v.a.

$$S_n = a_0 + X_1 + \dots + X_n, \quad \text{avec en particulier } S_0 = a_0.$$

Ce que l'on peut espérer comme fortune au $(n+1)$ -ième coup compte tenu de ce que l'on a gagné les n premiers coups est donné par

$$\mathbb{E}(S_{n+1} \mid X_1, \dots, X_n) = a_0 + X_1 + \dots + X_n + \mathbb{E}(X_{n+1})$$

car, X_{n+1} étant indépendante de la tribu $\sigma(X_1, \dots, X_n)$, on a $\mathbb{E}(X_{n+1} \mid X_1, \dots, X_n) = \mathbb{E}(X_{n+1})$. Ainsi, pour tout n ,

$$\mathbb{E}(S_{n+1} \mid X_1, \dots, X_n) = S_n.$$

Cette relation conditionnelle est une **propriété de martingale**; elle exprime l'équité du jeu à tout instant, quel qu'ait été son déroulement jusqu'alors.

Pour généraliser cette propriété, on étend le conditionnement ci-dessus à “toute l'information disponible au temps n ”, ce qui doit inclure les valeurs S_0, S_1, \dots, S_n , mais peut a priori être plus vaste (dans certains cas, S_{n+1} pourra dépendre d'une quantité disponible au temps n , qui n'est pas S_1, \dots, S_n , par exemple si le joueur décide de changer de jeu et qu'il connaît les valeurs passées de l'autre jeu, qui renseignent sur le futur). On suppose ainsi donnée une suite de tribus $\mathcal{F}_0, \mathcal{F}_1, \dots$ où, pour tout n , \mathcal{F}_n représente l'ensemble des événements connus au temps n . C'est la **tribu du passé avant le temps n (inclus)**. Vu cette intuition, cette suite est croissante (on l'appellera une **filtration**) :

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots,$$

et, pour tout n , S_n est \mathcal{F}_n -mesurable. Dans l'exemple précédent, le modèle ne comporte que le jeu répété, donc au temps n , $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Ainsi, la propriété de martingale prend aussi la forme suivante, qui sera prise comme définition :

$$\mathbb{E}(S_{n+1} \mid \mathcal{F}_n) = S_n, \quad \forall n \geq 0.$$

Imaginons maintenant que le joueur décide de faire varier sa mise. Lors du n -ième jeu, il joue une somme ϕ_n (positive) : ou bien il gagne $+\phi_n$ (si $X_n = 1$), ou bien il gagne $-\phi_n$ (si $X_n = -1$), autrement dit son gain est $X_n \cdot \phi_n$. Sa fortune au temps n sera donc

$$M_n = a_0 + X_1 \cdot \phi_1 + X_2 \cdot \phi_2 + \dots + X_n \cdot \phi_n.$$

Vu que $X_n = S_n - S_{n-1}$, on peut aussi voir M comme une transformation sur S :

$$M_n = S_0 + \sum_{k=1}^n (S_k - S_{k-1}) \cdot \phi_k,$$

autrement dit on a multiplié les incréments de S par ϕ : pour tout n ,

$$M_{n+1} - M_n = (S_{n+1} - S_n) \cdot \phi_{n+1}.$$

Notons que la mise ϕ_{n+1} est décidée juste avant de jouer au temps $n+1$, donc peut dépendre de tout le passé avant le temps n : ici, ϕ_{n+1} est une fonction de S_0, \dots, S_n , elle est \mathcal{F}_n -mesurable. Ce processus sera dit **prévisible**. On constate que, dans cet exemple, comme $M_{n+1} = M_n + \phi_{n+1} \cdot X_{n+1}$ et ϕ_{n+1} est \mathcal{F}_n -mesurable tandis que X_{n+1} est indépendante de \mathcal{F}_n et d'espérance nulle,

$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n + \phi_{n+1} \mathbb{E}(X_{n+1}) = M_n.$$

Ainsi, quelle que soit la stratégie de mise employée, la fortune satisfait encore la propriété de martingale : le jeu reste équitable. En écrivant ci-dessus $X_{n+1} = S_{n+1} - S_n$, on aurait obtenu cette propriété dans le cas général où S est une martingale.

Si on avait initialement considéré un jeu favorable, c'est-à-dire que $\mathbb{E}(X_n) \geq 0$ pour tout n , alors la suite $(S_n)_n$ aurait satisfait $\mathbb{E}(S_{n+1} | \mathcal{F}_n) \geq S_n$. On parlera de **sous-martingale**. Dans ce cas, on vérifie que M est aussi une sous-martingale.

Inversement, un jeu défavorable ($\mathbb{E}(X_n) \leq 0$) mène à la notion de **sur-martingale**.

Donnons maintenant des définitions et propriétés générales.

3.2 Définitions et exemples

Définition 3.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, et $(S_n)_{n \in \mathbb{N}}$ une processus à valeurs dans un espace (E, \mathcal{E}) .

- On appelle **filtration** (à temps discret) une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$, de sous-tribus de \mathcal{F} .
- On dit que $(S_n)_n$ est un processus **adapté à la filtration** $(\mathcal{F}_n)_n$ (ou $(\mathcal{F}_n)_n$ -adapté) si pour tout $n \geq 0$, S_n est \mathcal{F}_n -mesurable.
- La **filtration naturelle** du processus $(S_n)_{n \geq 0}$ est la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ définie par : pour tout $n \geq 0$, $\mathcal{F}_n = \sigma(S_0, \dots, S_n)$.

Ainsi, $(S_n)_n$ est adapté si à chaque instant l'information fournie par les valeurs passées du processus est contenue dans l'information donnée par la filtration $(\mathcal{F}_n)_n$. Et la filtration naturelle du processus est la plus petite filtration à laquelle il soit adapté.

On suppose dorénavant donnée une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ sur (Ω, \mathcal{F}) .

Définition 3.2. Un processus $(M_n)_{n \in \mathbb{N}}$ à valeurs réelles est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -**martingale** si

- (i) $(M_n)_{n \in \mathbb{N}}$ est adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$;
- (ii) pour tout $n \in \mathbb{N}$, M_n est intégrable ;
- (iii) pour tout $n \in \mathbb{N}$, $\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n$, p.s.

Il peut être utile d'introduire une variante de la définition, pour des variables positives, d'espérance finie ou infinie :

Définition 3.3. Un processus $(M_n)_{n \in \mathbb{N}}$ à valeurs dans $[0, \infty]$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -**martingale positive** si

- (i) $(M_n)_{n \in \mathbb{N}}$ est adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$;
- (iii) pour tout $n \in \mathbb{N}$, $\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n$, p.s.

Remplaçant “=” par “ \leq ” ou “ \geq ” dans la propriété de martingale, on obtient les notions suivantes :

Définition 3.4. Un processus $(M_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -**surmartingale** (resp. **sous-martingale**) si

- (i) $(M_n)_{n \in \mathbb{N}}$ est $(\mathcal{F}_n)_{n \in \mathbb{N}}$ adapté;
- (ii) pour tout $n \in \mathbb{N}$, M_n est intégrable;
- (iii) pour tout $n \in \mathbb{N}$, $\mathbb{E}(M_{n+1} | \mathcal{F}_n) \leq M_n$, (resp. $\geq M_n$) p.s.

On peut aussi définir sous-martingales et surmartingales positives, de la même façon que pour les martingales.

On notera qu'une martingale est à la fois une surmartingale et une sous-martingale, et que si $(M_n)_n$ est une surmartingale, alors $(-M_n)_n$ est une sous-martingale. On mémoriserà qu'une sous-martingale a une tendance à croître, tandis qu'une surmartingale a une tendance à décroître, *contrairement à ce que l'appellation pourrait suggérer*.

On peut également définir des (sur-,sous-)martingales à valeurs dans \mathbb{R}^d en demandant que chaque composante soit une (sur-,sous-)martingale.

Il est important de bien noter que toutes ces définitions sont liées à la filtration. Lorsque la filtration n'est pas précisée, il s'agit implicitement de la filtration naturelle du processus.

Exemple 3.5. Somme de v.a. i.i.d.. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles i.i.d. intégrables. Pour tout n , posons

$$S_n = X_0 + \dots + X_n.$$

Pour tout n , la v.a. S_n est \mathcal{F}_n -mesurable, pour $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$, et intégrable, de plus

$$\mathbb{E}(S_{n+1} | \sigma(X_0, \dots, X_n)) = X_0 + \dots + X_n + m = S_n + m,$$

parce que X_{n+1} est indépendante de $\sigma(X_0, \dots, X_n)$ et $m = \mathbb{E}(X_i)$ est l'espérance commune des variables X_i (elles ont la même loi). On en déduit :

- Si $m = 0$, $(S_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -martingale;
- Si $m > 0$, $(S_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -sous-martingale;
- Si $m < 0$, $(S_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -surmartingale.

Exemple 3.6. Produit de v.a. i.i.d.. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles i.i.d. positives. Pour tout n , posons

$$U_n = X_0 \cdots X_n.$$

Pour tout n , la v.a. U_n est \mathcal{F}_n -mesurable, pour $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$, et positive, de plus

$$\mathbb{E}(U_{n+1} | \sigma(X_0, \dots, X_n)) = X_0 \cdots X_n m = U_n m,$$

parce que X_{n+1} est indépendante de $\sigma(X_0, \dots, X_n)$ et $m = \mathbb{E}(X_i)$ est l'espérance commune des variables X_i (elles ont la même loi). On en déduit (avec la positivité de X_1, \dots, X_n) :

- Si $m = 1$, $(U_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -martingale;
- Si $m > 1$, $(U_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -sous-martingale;
- Si $m < 1$, $(U_n)_{n \in \mathbb{N}}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -surmartingale.

Notons que la positivité ne joue pas de rôle pour le cas martingale, et peut alors être remplacée par une hypothèse d'intégrabilité.

Exemple 3.7. Martingale de Doob (ou de Lévy). Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $(\mathcal{F}_n)_{n \geq 0}$ une filtration. La suite des espérances conditionnelles

$$X_n = \mathbb{E}(X | \mathcal{F}_n), \quad n \in \mathbb{N},$$

est une martingale adaptée à $(\mathcal{F}_n, n \in \mathbb{N})$. Cela vient de la propriété de double conditionnement.

Exemple 3.8. Origine du nom sur/sous-martingale. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur un espace d'états E , de matrice de transition P , et $f : E \rightarrow \mathbb{R}_+$ une fonction positive. On dit que f est **harmonique** (resp. **sous-harmonique**, resp. **sur-harmonique**), si $f = Pf$ (resp. $f \leq Pf$, resp. $f \geq Pf$); voir page 16 pour la définition de Pf . On note $(\mathcal{F}_n)_n$ la filtration naturelle de $(X_n)_n$. On vérifie simplement que

- Si f est harmonique, alors $(f(X_n))_{n \geq 0}$ est une $(\mathcal{F}_n)_n$ -martingale ;
- Si f est sous-harmonique, alors $(f(X_n))_{n \geq 0}$ est une $(\mathcal{F}_n)_n$ -sous-martingale ;
- Si f est sur-harmonique, alors $(f(X_n))_{n \geq 0}$ est une $(\mathcal{F}_n)_n$ -sur-martingale.

3.3 Premières propriétés

Dans la suite, on se donne une filtration $(\mathcal{F}_n)_{n \geq 0}$, et les martingales seront adaptées à cette filtration, même si cela n'est pas précisé.

Proposition 3.9. Soit $(M_n, n \in \mathbb{N})$ une (\mathcal{F}_n) -martingale (resp. surmartingale, resp. sous-martingale). Pour tout couple (n, p) d'entiers ≥ 0 on a :

$$\mathbb{E}(M_{n+p} | \mathcal{F}_n) = M_n, \quad (\text{resp. } \leq, \quad \text{resp. } \geq),$$

et

$$\mathbb{E}(M_{n+p}) = \mathbb{E}(M_n), \quad (\text{resp. } \leq, \quad \text{resp. } \geq).$$

Preuve : Il suffit en effet d'appliquer comme dans l'exemple ci-dessus le double conditionnement. \square

En particulier, on voit qu'une martingale est un processus à espérance constante.

On vérifie en outre facilement les résultats suivants, en se rappelant la notation pratique $a \vee b = \max(a, b)$ et $a \wedge b = \min(a, b)$, pour $a, b \in \mathbb{R}$:

Proposition 3.10. a) Soient $(X_n)_n$ et $(Y_n)_n$ deux martingales. Pour tous réels a et b , $(aX_n + bY_n)_n$ est une martingale. De plus, $(X_n \vee Y_n)_n$ est une sous-martingale et $(X_n \wedge Y_n)_n$ est une surmartingale.

b) Soient $(X_n)_n$ et $(Y_n)_n$ deux sur- (resp. sous-) martingales. Pour tous réels $a, b \geq 0$, $(aX_n + bY_n)_n$ est encore une sur- (resp. sous-) martingale.

Proposition 3.11. Soit $(M_n, n \in \mathbb{N})$ une (\mathcal{F}_n) -martingale (resp. surmartingale ou sous-martingale). On suppose que (M_n) est aussi adaptée à une autre filtration (\mathcal{G}_n) avec $\mathcal{G}_n \subset \mathcal{F}_n$ pour chaque $n \geq 0$. Alors $(M_n, n \in \mathbb{N})$ est aussi une (\mathcal{G}_n) -martingale (resp. surmartingale ou sous-martingale).

Preuve : On applique le double conditionnement (voir Proposition 2.11) :

$$\mathbb{E}(M_{n+1} | \mathcal{G}_n) = \mathbb{E}\left(\mathbb{E}(M_{n+1} | \mathcal{F}_n) \mid \mathcal{G}_n\right) = \mathbb{E}(M_n | \mathcal{F}_n) = M_n.$$

\square

Proposition 3.12. Soit (M_n) une martingale (resp. une sous-martingale) et ϕ une fonction convexe (resp. une fonction convexe croissante) telle que, pour tout n , $\phi(M_n)$ soit intégrable ou positive. Alors $(\phi(M_n))$ est une sous-martingale.

Preuve : Appliquer l'inégalité de Jensen. \square

Par cette proposition, pour toute martingale $(M_n)_n$, $(|M_n|)_n$ et $(M_n^2)_n$ sont des sous-martingales ; et pour toute sous-martingale $(M_n)_n$, $(e^{M_n})_n$ est une sous-martingale, et il en va de même de $(M_n^2)_n$ à condition que $M_n \geq 0$ pour tout n .

3.4 Stratégies : temps d'arrêt et transformée de martingale

Un exemple simple de stratégie pour un joueur consiste à cesser de jouer dès qu'une certaine condition sera satisfaite. Pour être applicable, la vérification de cette condition au temps n ne doit nécessairement dépendre que du passé avant le temps n . Ceci mène à la définition suivante, qui étend celle vue pour les chaînes de Markov :

Définition 3.13. Un *temps d'arrêt* de la filtration (\mathcal{F}_n) est une variable aléatoire $\tau : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ telle que

$$\text{pour tout } n \in \mathbb{N}, \quad \{\tau \leq n\} \in \mathcal{F}_n.$$

La *tribu \mathcal{F}_τ du passé avant τ* est définie comme l'ensemble des $A \in \mathcal{F}$ tels que, pour tout $n \in \mathbb{N}$, $A \cap \{\tau \leq n\} \in \mathcal{F}_n$.

De même que pour les chaînes de Markov (Lemme 2.18), il suffit de vérifier que, pour tout $n \in \mathbb{N}$, $\{\tau = n\} \in \mathcal{F}_n$, pour montrer que τ est un temps d'arrêt. Et on a encore :

Proposition 3.14. a) Si τ et σ deux temps d'arrêts, alors $\sigma \wedge \tau$, $\sigma \vee \tau$, $\sigma + \tau$ le sont aussi.

b) Si τ est un temps d'arrêt et (X_n) est un processus adapté à la filtration (\mathcal{F}_n) , alors τ et X_τ sont \mathcal{F}_τ -mesurables.

Soit τ un temps d'arrêt associé à la filtration (\mathcal{F}_n) , et (M_n) un processus adapté à la même filtration. On définit le **processus M arrêté au temps τ** comme le processus M^τ donné par :

$$M_n^\tau := M_{n \wedge \tau}, \quad n \geq 0.$$

Autrement dit, pour tout $\omega \in \Omega$,

$$M_n^\tau(\omega) = \begin{cases} M_n(\omega), & \text{si } n \leq \tau(\omega); \\ M_{\tau(\omega)}(\omega), & \text{sinon.} \end{cases}$$

Ce processus correspond à la fortune du joueur qui opte pour la stratégie de s'arrêter de jouer au temps τ .

En suivant l'introduction, on peut définir une notion plus générale de stratégie.

Définition 3.15. Un processus $(\phi_n)_{n \geq 0}$ est dit **prévisible** pour une filtration $(\mathcal{F}_n)_{n \geq 0}$ (ou $(\mathcal{F}_n)_n$ -prévisible) si pour tout $n \geq 1$, ϕ_n est \mathcal{F}_{n-1} -mesurable, et si ϕ_0 est \mathcal{F}_0 -mesurable.

Pour tout processus $(X_n)_{n \geq 0}$ on introduit une notation pour les accroissements :

$$\text{pour tout } n \geq 1, \quad \Delta X_n = X_n - X_{n-1}.$$

On note par exemple que la propriété de martingale de $(M_n)_n$ est équivalente à $\mathbb{E}(\Delta M_n | \mathcal{F}_{n-1}) = 0$.

Définition 3.16. Soit $M = (M_n)_{n \geq 0}$ une martingale (resp. sur- ou sous-martingale) associée à une filtration $(\mathcal{F}_n)_{n \geq 0}$ et $\phi = (\phi_n)_{n \geq 0}$ un processus à valeurs réelles, **prévisible** pour la filtration (\mathcal{F}_n) .

On appelle **transformée de la martingale** (resp. **sur- ou sous-martingale**) M **par le processus prévisible** ϕ , le processus noté $\langle \phi, M \rangle = (\langle \phi, M \rangle_n)_{n \geq 0}$ défini par $\langle \phi, M \rangle_0 = \phi_0 M_0$ et

$$\Delta \langle \phi, M \rangle_n = \phi_n \Delta M_n, \quad n \geq 1.$$

Autrement dit,

$$\langle \phi, M \rangle_n = \phi_0 M_0 + \sum_{i=1}^n \phi_i \Delta M_i.$$

Exemple 3.17.

- Si $\phi \equiv 1$, alors $\langle \phi, M \rangle_n = M_n$ pour tout n .
- Dans l'introduction, S_n est, après n parties, la fortune d'un joueur qui mise une somme 1 à chaque partie ; s'il choisit de miser plutôt ϕ_n lors du n -ième coup (avec $\phi_0 = 1$), alors sa fortune au temps n est $M_n = \langle \phi, S \rangle_n$ (le gain ΔS_n est multiplié par ϕ_n en cas de mise).
- Soit τ un temps d'arrêt associé à la filtration (\mathcal{F}_n) . Posons

$$\phi_n = \mathbf{1}_{(\tau \geq n)}, \quad n \geq 0.$$

On définit ainsi un processus prévisible $\phi = (\phi_n)_{n \geq 0}$ puisque $\{\tau \geq n\}^c = \{\tau \leq n-1\} \in \mathcal{F}_{n-1}$. Il vient donc pour tout n ,

$$\langle \phi, M \rangle_n = M_0 + \sum_{i=1}^n \mathbf{1}_{(\tau \geq i)} \Delta M_i = M_{n \wedge \tau} = M_n^\tau, \quad n \geq 0.$$

c'est à dire que $\langle \phi, M \rangle = M^\tau$. Le processus arrêté d'une martingale (resp. sur- ou sous-martingale) est donc la transformée de la martingale (resp. sur- ou sous-) initiale par le processus prévisible $\phi = (\phi_n) = (\mathbf{1}_{(\tau \geq n)})_{n \geq 0}$. En particulier,

$$M_n^\tau = M_0 + \sum_{i=1}^n \mathbf{1}_{(\tau \geq i)} \Delta M_i. \quad (3.1)$$

Théorème 3.18.

- Si M est une martingale et si le processus prévisible ϕ est borné, alors le processus $\langle \phi, M \rangle$ est aussi une martingale.
- Si M est une sur- (resp. une sous-)martingale et si le processus prévisible ϕ est borné et **positif**, alors le processus $\langle \phi, M \rangle$ est aussi une sur- (resp. une sous-)martingale.

Remarque : L'hypothèse ϕ borné (i.e. il existe une constante $C > 0$ telle que $\sup_n |\phi_n| \leq C$ p.s.) n'est faite que pour assurer l'intégrabilité de $\langle \phi, M \rangle_n$ pour tout n , cette hypothèse peut donc être affaiblie dans beaucoup de cas, par exemple, dans le cas d'une sur- ou sous-martingale positive, il suffit de supposer que ϕ est un processus positif.

Preuve : Soit M une martingale. Pour tout n dans \mathbb{N} , $\langle \phi, M \rangle_n$ est \mathcal{F}_n -mesurable et intégrable puisque ϕ_n est bornée. Comme

$$\Delta \langle \phi, M \rangle_n = \phi_n \Delta M_n,$$

et ϕ_n est \mathcal{F}_{n-1} -mesurable, on a

$$\mathbb{E}(\Delta \langle \phi, M \rangle_n | \mathcal{F}_{n-1}) = \phi_n \mathbb{E}(\Delta M_n | \mathcal{F}_{n-1}) = 0,$$

car M est une martingale. D'où le résultat dans le cas martingale. La démonstration s'adapte sans difficulté au cas des sur- et sous-martingales. \square

La conséquence directe suivante (qui vient de l'exemple précédent) est fondamentale :

Corollaire 3.19. *Toute martingale (resp. sur- ou sous-martingale) arrêtée par un temps d'arrêt est encore une martingale (resp. une sur- ou sous-martingale).*

Ce corollaire nous montre en particulier que pour toute martingale M (resp. sur-, resp. sous-martingale) et pour tout temps d'arrêt τ et tous n et k tels que $0 \leq k \leq n$, on a

$$\mathbb{E}(M_{\tau \wedge n} | \mathcal{F}_k) = M_{\tau \wedge k}, \quad (\text{resp. } \leq \text{ resp. } \geq). \quad (3.2)$$

Généralisation au cas de martingales à valeurs dans \mathbb{R}^d . Soit $M_n = (M_n^1, \dots, M_n^d)$ une (\mathcal{F}_n) -martingale à valeurs dans \mathbb{R}^d , et $\phi_n = (\phi_n^1, \dots, \phi_n^d)$ un processus (\mathcal{F}_n) -prévisible à valeurs dans \mathbb{R}^d . Considérons pour tout $n \geq 1$ le vecteur aléatoire de \mathbb{R}^d

$$\Delta M_n = (\Delta M_n^1, \dots, \Delta M_n^d)$$

ainsi que le produit scalaire

$$\phi_n \cdot \Delta M_n = \sum_{i=1}^d \phi_n^i \Delta M_n^i,$$

et définissons comme précédemment le processus $\langle \phi, M \rangle$ à valeurs réelles par $\langle \phi, M \rangle_0 = \phi_0 \cdot M_0$ et, pour tout $n \geq 1$, $\Delta \langle \phi, M \rangle_n = \phi_n \cdot \Delta M_n$, c'est-à-dire

$$\langle \phi, M \rangle_n = \phi_0 \cdot M_0 + \sum_{k=1}^n \phi_k \cdot \Delta M_k.$$

Théorème 3.20. Avec les notations précédentes $\langle \phi, M \rangle$ est une martingale à valeurs réelles.

Preuve : On a

$$\mathbb{E}(\Delta \langle \phi, M \rangle_n | \mathcal{F}_{n-1}) = \sum_{i=1}^d \phi_n^i \mathbb{E}(\Delta M_n^i | \mathcal{F}_{n-1}) = 0,$$

en se rappelant que chaque composante M^i est une martingale. \square

3.5 Théorème d'arrêt

La formule (3.2) précédente va nous permettre de démontrer que la propriété de martingale (resp. de sur- ou sous-martingale) se généralise à certains temps d'arrêt, ce qui s'appelle un "théorème d'arrêt".

Théorème 3.21 (Théorème d'arrêt, cas borné). Soient S et T deux temps d'arrêt et M une martingale (resp. surmartingale, resp. sous-martingale), relativement à la même filtration (\mathcal{F}_n) . On suppose qu'il existe un entier N tel que

$$S \leq T \leq N, \quad \text{p.s.}$$

Alors

$$\mathbb{E}(M_T | \mathcal{F}_S) = M_S, \quad (\text{resp. } \leq, \text{ resp. } \geq)$$

En particulier, $\mathbb{E}(M_T) = \mathbb{E}(M_S)$, (resp. \leq , resp. \geq).

Preuve : Montrons le cas surmartingale. Comme $M_T = \sum_{k=0}^N M_k \mathbf{1}_{(T=k)}$ et $M_S = \sum_{k=0}^N M_k \mathbf{1}_{(S=k)}$, les v.a. M_T et M_S sont intégrables (on a $|M_T| \leq |M_0| + \dots + |M_N|$ et de même pour M_S). Pour tout $A \in \mathcal{F}_S$, on a

$$\begin{aligned} \mathbb{E}(M_T \mathbf{1}_A) &= \sum_{k=0}^N \mathbb{E}(M_T \mathbf{1}_{A \cap (S=k)}) \\ &\leq \sum_{k=0}^N \mathbb{E}(M_T \wedge k \mathbf{1}_{A \cap (S=k)}), \end{aligned}$$

d'après (3.2) et le fait que $A \cap \{S = k\} \in \mathcal{F}_k$. Mais sur l'évènement $\{S = k\}$ on a $T \geq k$ et donc

$$M_T \wedge k = M_k = M_S, \quad \text{sur l'évènement } \{S = k\}.$$

Il vient, en regroupant les sommes,

$$\mathbb{E}(M_T \mathbf{1}_A) \leq \mathbb{E}(M_S \mathbf{1}_A),$$

ou de façon équivalente pour tout $A \in \mathcal{F}_S$,

$$\mathbb{E}(\mathbb{E}(M_T | \mathcal{F}_S) \mathbf{1}_A) \leq \mathbb{E}(M_S \mathbf{1}_A),$$

ce qui montre que $\mathbb{E}(M_T | \mathcal{F}_S) \leq M_S$ p.s., puisque M_S est \mathcal{F}_S -mesurable. \square

Remarque 3.22. En particulier, avec $S \equiv 0$ et T un temps d'arrêt borné, on obtient $\mathbb{E}(M_T) = \mathbb{E}(M_0)$ (resp. \leq ou \geq).

3.6 Inégalités maximales

Théorème 3.23. *Pour toute sous-martingale positive $X = (X_n)_{n \in \mathbb{N}}$ on a l'inégalité : pour tout $\lambda > 0$, pour tout $n \geq 0$,*

$$\lambda \mathbb{P}\left(\sup_{0 \leq k \leq n} X_k \geq \lambda\right) \leq \mathbb{E}\left(X_n \mathbf{1}_{\{\sup_{0 \leq k \leq n} X_k \geq \lambda\}}\right) \leq \mathbb{E}(X_n).$$

En particulier, pour toute martingale de carré intégrable $(M_n)_{n \in \mathbb{N}}$, pour tout $\lambda > 0$, pour tout $n \geq 0$,

$$\mathbb{P}\left(\sup_{0 \leq k \leq n} |M_k| \geq \lambda\right) \leq \frac{\mathbb{E}(M_n^2)}{\lambda^2}, \quad \lambda > 0.$$

Remarque : Le fait que $X_n \geq 0$ n'est utilisé que dans la majoration $\mathbb{E}(X_n \mathbf{1}_{\{\sup_{0 \leq k \leq n} X_k \geq \lambda\}}) \leq \mathbb{E}(X_n)$. L'inégalité $\lambda \mathbb{P}\left(\sup_{0 \leq k \leq n} X_k \geq \lambda\right) \leq \mathbb{E}\left(X_n \mathbf{1}_{\{\sup_{0 \leq k \leq n} X_k \geq \lambda\}}\right)$ reste vraie sans l'hypothèse de positivité de X .

Preuve : Posons $T = \inf\{k \geq 0 \mid X_k \geq \lambda\}$ avec par convention $\inf \emptyset = \infty$. T est un temps d'arrêt pour la filtration (\mathcal{F}_n) . Soit $n \in \mathbb{N}$. Remarquons que

$$\left\{\sup_{0 \leq k \leq n} X_k \geq \lambda\right\} = \{T \leq n\}.$$

On peut appliquer à ce stade le Théorème 3.21 à $T \wedge n \leq n$ et utiliser le fait que $\{T \leq n\} \in \mathcal{F}_{T \wedge n}$ pour obtenir :

$$\mathbb{E}(X_n \mathbf{1}_{\{T \leq n\}}) \geq \mathbb{E}(X_{T \wedge n} \mathbf{1}_{\{T \leq n\}}) = \mathbb{E}(X_T \mathbf{1}_{\{T \leq n\}}) \geq \lambda \mathbb{P}(T \leq n),$$

ce qui est exactement la première inégalité du théorème. Alternativement, on peut directement procéder comme suit :

$$\begin{aligned} \mathbb{E}\left(X_n \mathbf{1}_{\{T \leq n\}}\right) &= \sum_{k=0}^n \mathbb{E}\left(X_n \mathbf{1}_{\{T=k\}}\right) = \sum_{k=0}^n \mathbb{E}\left(\mathbb{E}(X_n \mid \mathcal{F}_k) \mathbf{1}_{\{T=k\}}\right) \\ &\geq \sum_{k=0}^n \mathbb{E}\left(X_k \mathbf{1}_{\{T=k\}}\right) \geq \lambda \sum_{k=0}^n \mathbb{E}\left(\mathbf{1}_{\{T=k\}}\right) = \lambda \mathbb{P}(T \leq n). \end{aligned}$$

Le cas particulier s'obtient en remarquant que (M_n^2) est une sous-martingale positive d'après l'inégalité de Jensen (Proposition 3.12) et que $\mathbb{P}(\sup_k |M_k| \geq \lambda) = \mathbb{P}(\sup_k M_k^2 \geq \lambda^2)$. \square

Proposition 3.24 (Inégalité de Doob dans L^2). *Pour toute sous-martingale positive X on a :*

$$\text{pour tout } n \geq 0, \quad \mathbb{E}\left(\sup_{0 \leq k \leq n} X_k^2\right) \leq 4\mathbb{E}(X_n^2).$$

En terme de norme L^2 , on a

$$\text{pour tout } n \geq 0, \quad \|X_n\|_{L^2} \leq \left\| \sup_{0 \leq k \leq n} X_k \right\|_{L^2} \leq 2\|X_n\|_{L^2}.$$

Preuve : En remarquant que, pour toute v.a. Z positive, $Z^2 = \int_0^Z 2\lambda d\lambda = 2 \int_0^\infty \lambda \mathbf{1}_{\{Z \geq \lambda\}} d\lambda$ et donc que, par le théorème de Fubini-Tonelli,

$$\mathbb{E}(Z^2) = 2 \int_0^\infty \lambda \mathbb{P}(Z \geq \lambda) d\lambda,$$

il vient en utilisant les résultats de la démonstration précédente et en posant $Z_n = \sup_{0 \leq k \leq n} X_k$,

$$\mathbb{E}(Z_n^2) = 2 \int_0^\infty \lambda \mathbb{P}(Z_n \geq \lambda) d\lambda \leq 2 \int_0^\infty \mathbb{E}\left(X_n \mathbf{1}_{\{Z_n \geq \lambda\}}\right) d\lambda = 2\mathbb{E}\left(X_n Z_n\right) \leq 2\sqrt{\mathbb{E}(X_n^2)\mathbb{E}(Z_n^2)},$$

d'après le théorème de Fubini-Tonelli puis l'inégalité de Cauchy-Schwarz. Enfin, la première égalité entre normes L^2 vient simplement du fait que

$$X_n^2 \leq \left(\sup_{0 \leq k \leq n} X_k\right)^2.$$

\square

Avec la même démonstration, on peut généraliser l'inégalité à L^p (avec $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p}$) :

Proposition 3.25 (Inégalité de Doob dans L^p). *Soit $p > 1$. Pour toute sous-martingale positive X on a :*

$$\|X_n\|_p \leq \left\| \sup_{0 \leq k \leq n} X_k \right\|_p \leq \frac{p}{p-1} \|X_n\|_p.$$

3.7 Théorèmes de convergence

De même que les suites réelles monotones bornées convergent, on va voir que les surmartingales et sous-martingales (et donc les martingales) s'avèrent converger sous diverses hypothèses de majoration ou minoration, et en divers sens, ce qui en fait un outil théorique presque aussi utile en probabilités que les suites réelles monotones en analyse réelle.

3.7.1 Convergence dans L^2

Théorème 3.26. *Soit $M = (M_n)_{n \geq 0}$ une martingale. M converge dans L^2 vers une v.a. $M_\infty \in L^2$ (i.e. $\|M_n - M_\infty\|_2 \rightarrow 0$) si, et seulement si, la suite $(M_n)_n$ est bornée dans L^2 , c'est-à-dire que*

$$\sup_{n \in \mathbb{N}} \mathbb{E}(M_n^2) < \infty. \quad (3.3)$$

De plus on a dans ce cas, pour tout n ,

$$\mathbb{E}(M_\infty | \mathcal{F}_n) = M_n. \quad (3.4)$$

Preuve : Partie "Seulement si" : (3.3) vient du fait qu'une suite convergente (dans L^2 , ici) est bornée.

Partie "Si" : Puisque M est de carré intégrable, (M_n^2) est une sous-martingale. La suite des espérances $(\mathbb{E}(M_n^2))_n$ est donc croissante; la condition (3.3) du théorème montre que cette suite d'espérance est convergente, donc de Cauchy dans \mathbb{R} . D'un autre côté en développant le carré on obtient, par la propriété de martingale, pour tous $n, p \geq 0$,

$$\mathbb{E}\left((M_{n+p} - M_n)^2 \mid \mathcal{F}_n\right) = \mathbb{E}(M_{n+p}^2 \mid \mathcal{F}_n) - 2M_n \mathbb{E}(M_{n+p} \mid \mathcal{F}_n) + M_n^2 = \mathbb{E}(M_{n+p}^2 \mid \mathcal{F}_n) - M_n^2 \quad (3.5)$$

d'où, en prenant l'espérance,

$$\|M_{n+p} - M_n\|_2^2 = \mathbb{E}(M_{n+p}^2) - \mathbb{E}(M_n^2).$$

La suite de v.a. $(M_n)_n$ est donc de Cauchy dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ et converge donc vers une v.a. $M_\infty \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ (par complétude de L^2).

Enfin, l'égalité (3.4) vient de la continuité de l'application de L^2 dans L^2 de projection $X \mapsto \mathbb{E}(X | \mathcal{F}_n)$ (cette application est contractante), qui justifie de passer à la limite dans $M_n = \mathbb{E}(M_l | \mathcal{F}_n)$ lorsque $l \rightarrow \infty$. \square

3.7.2 Convergence presque sûre

Théorème 3.27. *Si la martingale M converge dans L^2 vers une v.a. $M_\infty (\in L^2)$, alors elle converge presque sûrement vers M_∞ .*

La démonstration de ce théorème est plus délicate que la précédente; elle utilise en particulier les inégalités maximales pour contrôler les fluctuations.

Preuve : Pour $n \geq 0$, on pose $V_n := \sup_{i, j \geq n} |M_i - M_j|$. Presque sûrement, $(V_n)_n$ est une suite décroissante positive, donc admet une limite $\bar{V} = \lim_n V_n$. Soit $n \geq 0$. D'après l'inégalité de Doob (appliquée à la martingale $(M_{n+j} - M_n)_{j \geq 0}$) on a, pour tout k ,

$$\mathbb{E}\left(\sup_{0 \leq j \leq k} |M_{j+n} - M_n|^2\right) \leq 4\mathbb{E}((M_{k+n} - M_n)^2).$$

Pour $k \rightarrow \infty$, on a $\mathbb{E}\left(\sup_{j \geq 0} |M_{j+n} - M_n|^2\right) \leq 4\mathbb{E}((M_\infty - M_n)^2)$. Donc

$$\mathbb{E}(V_n^2) \leq 2\mathbb{E}\left(\sup_{j \geq 0} |M_{j+n} - M_n|^2\right) \leq 8\mathbb{E}((M_\infty - M_n)^2) = 8\|M_n - M_\infty\|_2^2.$$

Alors $\mathbb{E}(V_n^2) \rightarrow 0$ et donc $\mathbb{E}(V^2) = 0$ par convergence dominée, d'où $V = 0$ p.s.. Autrement dit, p.s., $(M_n)_n$ est une suite de Cauchy et nécessairement M_∞ est sa limite puisque $(M_n)_n$ converge dans L^2 vers M_∞ (donc p.s. une sous-suite de $(M_n)_n$ converge vers M_∞). \square

Ce théorème se déduit aussi du résultat essentiel suivant, qui est beaucoup plus général. Rappelons d'abord que, pour tout réel x on note $x^+ = \sup(x, 0)$ et $x^- = \sup(-x, 0)$.

Théorème 3.28 (Doob). *Toute sous-martingale $(X_n)_{n \in \mathbb{N}}$ telle que $\sup_{n \geq 1} \mathbb{E}(X_n^+) < \infty$ converge p.s. vers une v.a. X_∞ intégrable.*

De même, toute surmartingale $(X_n)_{n \in \mathbb{N}}$ telle que $\sup_{n \geq 1} \mathbb{E}(X_n^-) < \infty$ converge p.s. vers une v.a. X_∞ intégrable.

Immédiatement on obtient le corollaire suivant, qui est le cas d'application le plus fréquent :

Corollaire 3.29. *Toute surmartingale positive converge presque sûrement.*

Un peu plus généralement, toute surmartingale minorée par une v.a. intégrable (par exemple une constante), et toute sous-martingale majorée par une v.a. intégrable, convergent presque sûrement (si $X_n \leq Z$ alors $X_n^+ \leq Z^+ \leq |Z|$ donc $\sup_n \mathbb{E}(X_n^+) \leq \mathbb{E}(|Z|)$). Et donc toute martingale majorée ou minorée par une v.a. intégrable converge presque sûrement. Rappelons tout de même que la condition du théorème est plus générale.

Vu que $X_n^+ \leq |X_n|$, on obtient aussi que toute sur- ou sous-martingale bornée dans L^1 converge presque sûrement. À la différence du cas L^2 vu précédemment, il se peut en revanche qu'elle ne converge pas dans L^1 ; on reviendra sur ce point après la preuve du théorème 3.28.

3.7.3 Preuve du Théorème 3.28 :

Soient $a < b$. On définit une suite de temps d'arrêts $\tau_0 \leq \sigma_1 \leq \tau_1 \leq \sigma_2 \leq \tau_2 \leq \dots$ par récurrence par : $\tau_0 = 0$ puis, pour tout $k \geq 1$,

$$\sigma_k = \inf\{j > \tau_{k-1} \mid X_j \leq a\}, \quad \tau_k = \inf\{j > \sigma_k \mid X_j \geq b\}.$$

avec par convention $\inf \emptyset = \infty$. On pose alors $\nu_{a,b}^{(n)} = \max\{k \geq 0 \mid \tau_k \leq n\}$, appelé le **nombre de franchissements croissants de $[a, b]$ par (X_1, \dots, X_n)** .

Lemme 3.30. *Soit (X_n) une sous-martingale. On a pour tous $a < b$, pour tout $n \in \mathbb{N}$,*

$$\mathbb{E}\left(\nu_{a,b}^{(n)}\right) \leq 1 + \frac{\mathbb{E}((X_n - a)^+)}{b - a} \leq 1 + \frac{\mathbb{E}(X_n^+) + |a|}{b - a}.$$

Preuve : Soit $n \in \mathbb{N}$. On a, pour tout $k \geq 1$, $\tau_{k-1} \leq \sigma_k$, d'où $n \wedge \tau_{k-1} \leq n \wedge \sigma_k \leq n$, donc le théorème d'arrêt appliqué à la sous-martingale X et aux temps d'arrêt bornés $n \wedge \tau_{k-1}$ et $n \wedge \sigma_k$ donne $\mathbb{E}(X_{n \wedge \tau_{k-1}}) \leq \mathbb{E}(X_{n \wedge \sigma_k})$, c'est-à-dire

$$\mathbb{E}(X_{n \wedge \sigma_k} - X_{n \wedge \tau_{k-1}}) \geq 0.$$

Or, pour tout $k \geq 2$, par la définition de σ_k et τ_{k-1} ,

$$X_{n \wedge \sigma_k} - X_{n \wedge \tau_{k-1}} = \begin{cases} X_{\sigma_k} - X_{\tau_{k-1}} \leq -(b - a) & \text{si } \sigma_k \leq n, \\ X_n - X_{\tau_{k-1}} \leq X_n - b \leq X_n - a = (X_n - a)^+ & \text{si } \tau_{k-1} < n < \sigma_k, \\ X_n - X_n = 0 & \text{si } n \leq \tau_{k-1}, \end{cases}$$

ce qui donne en intégrant, avec l'inégalité précédente,

$$0 \leq \mathbb{E}(X_{n \wedge \sigma_k} - X_{n \wedge \tau_{k-1}}) \leq -(b-a)\mathbb{P}(\sigma_k \leq n) + \mathbb{E}((X_n - a)^+ \mathbf{1}_{(\tau_{k-1} < n < \sigma_k)}).$$

En remarquant que $\{\nu_{a,b}^{(n)} \geq k\} = \{\tau_k \leq n\} \subset \{\sigma_k \leq n\}$, on en déduit, pour tout $k \geq 2$,

$$\mathbb{P}(\nu_{a,b}^{(n)} \geq k) \leq \mathbb{P}(\sigma_k \leq n) \leq \frac{\mathbb{E}((X_n - a)^+ \mathbf{1}_{(\tau_{k-1} < n < \sigma_k)})}{b-a}.$$

Par ailleurs, les événements $\{\tau_{k-1} < n < \sigma_k\}$, pour $k \geq 2$, sont disjoints, donc en sommant sur $k \geq 2$ on obtient

$$\sum_{k \geq 2} \mathbb{P}(\nu_{a,b}^{(n)} \geq k) \leq \frac{\mathbb{E}((X_n - a)^+ \mathbf{1}_F)}{b-a},$$

où $F = \bigcup_{k \geq 2} \{\tau_{k-1} < n < \sigma_k\}$. Comme $\nu_{a,b}^{(n)}$ est à valeurs dans \mathbb{N} , on a alors

$$\mathbb{E}(\nu_{a,b}^{(n)}) = \sum_{k \geq 1} \mathbb{P}(\nu_{a,b}^{(n)} \geq k) \leq 1 + \sum_{k \geq 2} \mathbb{P}(\nu_{a,b}^{(n)} \geq k) \leq 1 + \frac{\mathbb{E}((X_n - a)^+ \mathbf{1}_F)}{b-a} \leq 1 + \frac{\mathbb{E}((X_n - a)^+)}{b-a}.$$

C'est la première inégalité annoncée. La seconde vient du fait que $(X_n - a)^+ \leq X_n^+ + |a|$. \square

Preuve du Théorème 3.28 : Soit X une sous-martingale telle que $\sup_n \mathbb{E}(X_n^+) < \infty$. Montrer que (X_n) converge presque sûrement dans $\overline{\mathbb{R}}$ est équivalent à montrer que

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} X_n > \liminf_{n \rightarrow \infty} X_n\right) = 0.$$

Remarquons que

$$\left\{\limsup_{n \rightarrow \infty} X_n > \liminf_{n \rightarrow \infty} X_n\right\} \subset \bigcup_{\substack{a < b, \\ a, b \text{ rationnels}}} \left\{\limsup_{n \rightarrow \infty} X_n > b > a > \liminf_{n \rightarrow \infty} X_n\right\},$$

et $\left\{\limsup_{n \rightarrow \infty} X_n > b > a > \liminf_{n \rightarrow \infty} X_n\right\} \subset \left\{\lim_{n \rightarrow \infty} \nu_{a,b}^{(n)} = \infty\right\}$. Mais d'après le lemme précédent, et par convergence monotone, vu l'hypothèse du théorème,

$$\mathbb{E}\left(\lim_{n \rightarrow \infty} \nu_{a,b}^{(n)}\right) = \lim_{n \rightarrow \infty} \mathbb{E}(\nu_{a,b}^{(n)}) \leq 1 + \sup_n \frac{\mathbb{E}(X_n^+) + |a|}{b-a} < \infty,$$

donc $\mathbb{P}\left(\lim_{n \rightarrow \infty} \nu_{a,b}^{(n)} = \infty\right) = 0$ et $\mathbb{P}\left(\limsup_{n \rightarrow \infty} X_n > \liminf_{n \rightarrow \infty} X_n\right) = 0$. Ainsi, il existe une v.a. X_∞ à valeurs dans $\overline{\mathbb{R}}$ telle que $(X_n)_n$ converge vers X_∞ p.s..

On a, pour tout n , $|X_n| = X_n^+ + X_n^- = X_n^+ + (-X_n + X_n^+) = 2X_n^+ - X_n$, et $\mathbb{E}(X_n) \geq \mathbb{E}(X_0)$ car $(X_n)_n$ est une sous-martingale, donc

$$\mathbb{E}(|X_n|) = 2\mathbb{E}(X_n^+) - \mathbb{E}(X_n) \leq 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0).$$

On déduit alors du lemme de Fatou que

$$\mathbb{E}(|X_\infty|) = \mathbb{E}\left(\lim_{n \rightarrow \infty} |X_n|\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|X_n|) \leq 2 \sup_n \mathbb{E}(X_n^+) - \mathbb{E}(X_0) < \infty$$

d'après l'hypothèse, donc X_∞ est intégrable. Notamment, X_∞ est donc presque sûrement finie. \square

(*) Uniforme intégrabilité et convergence dans L^1 .

On a vu qu'une martingale $(M_n)_n$ bornée dans L^2 converge dans L^2 et presque sûrement vers une v.a. $M_\infty \in L^2$ et qu'en particulier (du fait de la convergence L^2 , qui implique la convergence L^1) la propriété de martingale passe à la limite sous la forme

$$\text{pour tout } n, \quad \mathbb{E}(M_\infty | \mathcal{F}_n) = M_n.$$

En revanche, une martingale bornée dans L^1 converge p.s. vers une v.a. $M_\infty \in L^1$ mais il se peut que la convergence n'ait pas lieu dans L^1 . Cela sera précisé dans le théorème 3.33, qui utilise la notion suivante, et la Proposition 3.32 qui en donne des propriétés.

Définition 3.31. Soit I un ensemble d'indices quelconque (dénombrable ou non). Une famille de v.a. réelles $(X_i)_{i \in I}$ est dite **uniformément intégrable** si

$$\sup_{i \in I} \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq \lambda\}}) \xrightarrow{\lambda \rightarrow \infty} 0.$$

Proposition 3.32.

- Si $(X_i)_{i \in I}$ est uniformément intégrable, alors elle est bornée dans L^1 : $\sup_{i \in I} \|X_i\|_{L^1} < \infty$.
- Soit $p > 1$. Si $(X_i)_{i \in I}$ est bornée dans L^p , alors $(X_i)_{i \in I}$ est uniformément intégrable.
- Pour toute v.a. $Y \in L^1$, et toute filtration $(\mathcal{F}_n)_n$, la suite $(X_n)_n$ donnée par $X_n = \mathbb{E}(Y | \mathcal{F}_n)$ est uniformément intégrable.
- Si $X_n \rightarrow X_\infty$ p.s. et que $(X_n)_{n \geq 1}$ est uniformément intégrable, alors X_n converge vers X_∞ dans L^1 (et donc en particulier $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X_\infty)$).

Preuve : a) En choisissant λ suffisamment grand pour que $\sup_{i \in I} \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq \lambda\}}) \leq 1$, on a pour tout $i \in I$, $\mathbb{E}(|X_i|) \leq \lambda + \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq \lambda\}}) \leq \lambda + 1$, d'où la conclusion.

b) Par l'inégalité de Hölder,

$$\mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq \lambda\}}) \leq \|X_i\|_{L^p} \left(\mathbb{P}(|X_i| \geq \lambda) \right)^{1/q} \leq C \left(\mathbb{P}(|X_i| \geq \lambda) \right)^{1/q},$$

où $C := \sup_{i \in I} \|X_i\|_{L^p} < \infty$, et d'après l'inégalité de Markov, $\mathbb{P}(|X_i| \geq \lambda) \leq \frac{1}{\lambda^p} \mathbb{E}(|X_i|^p) \leq C^p \lambda^{-p} \rightarrow 0$ quand $\lambda \rightarrow \infty$, uniformément par rapport à $i \in I$, ce qui montre l'uniforme intégrabilité.

c) On a, pour tout $M > 0$, pour tout $n \in \mathbb{N}$, par l'inégalité triangulaire (ou de Jensen) puis par la définition de l'espérance conditionnelle, la croissance de l'espérance et enfin l'inégalité de Markov,

$$\begin{aligned} \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| \geq \lambda\}}) &\leq \mathbb{E}(\mathbb{E}(|Y| | \mathcal{F}_n) \mathbf{1}_{\{|X_n| \geq \lambda\}}) = \mathbb{E}(|Y| \mathbf{1}_{\{|X_n| \geq \lambda\}}) \\ &= \mathbb{E}(|Y| \mathbf{1}_{\{|Y| \leq M\}} \mathbf{1}_{\{|X_n| \geq \lambda\}}) + \mathbb{E}(|Y| \mathbf{1}_{\{|Y| > M\}} \mathbf{1}_{\{|X_n| \geq \lambda\}}) \\ &\leq M \mathbb{P}(|X_n| \geq \lambda) + \mathbb{E}(|Y| \mathbf{1}_{\{|Y| > M\}}) \\ &\leq \frac{M}{\lambda} \mathbb{E}(|Y|) + \mathbb{E}(|Y| \mathbf{1}_{\{|Y| > M\}}), \end{aligned}$$

en utilisant aussi le fait que $\mathbb{E}(|X_n|) \leq \mathbb{E}(|Y|)$. Ainsi, pour tout $M > 0$,

$$\limsup_{\lambda \rightarrow \infty} \sup_n \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| \geq \lambda\}}) \leq \mathbb{E}(|Y| \mathbf{1}_{\{|Y| > M\}}).$$

Or, par théorème de convergence dominée, $\mathbb{E}(|Y| \mathbf{1}_{\{|Y| > M\}}) \rightarrow 0$ quand $M \rightarrow \infty$, donc la limsup précédente est majorée par 0, ce qui donne l'uniforme intégrabilité de $(X_n)_n$.

d) Le lemme de Fatou montre que $\mathbb{E}(|X_\infty|) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|X_n|) \leq \sup_{n \geq 1} \|X_n\|_{L^1} < \infty$, donc $X_\infty \in L^1$ et on déduit facilement l'uniforme intégrabilité de $(X_n - X_\infty)_{n \geq 1}$ de celle de $(X_n)_{n \geq 1}$. D'autre part, $\mathbb{E}(|X_n - X_\infty|) = \mathbb{E}(|X_n - X_\infty| \mathbf{1}_{\{|X_n - X_\infty| \leq \lambda\}}) + \mathbb{E}(|X_n - X_\infty| \mathbf{1}_{\{|X_n - X_\infty| > \lambda\}})$. Par convergence dominée, on a pour tout $\lambda > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{E}(|X_n - X_\infty|) \leq \limsup_{n \rightarrow \infty} \mathbb{E}(|X_n - X_\infty| \mathbf{1}_{\{|X_n - X_\infty| > \lambda\}}),$$

qui tend vers 0 quand $\lambda \rightarrow \infty$ grâce à l'uniforme l'intégrabilité de $(X_n - X_\infty)_{n \geq 1}$. \square

Théorème 3.33 (Doob). *Soit $(M_n)_n$ une martingale. Les propriétés suivantes sont équivalentes :*

- (1) *il existe une v.a. $Y \in L^1$ telle que, pour tout n , M_n s'écrit $M_n = \mathbb{E}(Y | \mathcal{F}_n)$ (on dira que (M_n) est une martingale régulière) ;*
- (2) *$(M_n)_n$ est uniformément intégrable ;*
- (3) *$(M_n)_n$ converge vers M_∞ dans L^1 .*

Et dans ce cas on a, pour tout n , $M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$.

Preuve : (3) \implies (1) : si $M_k \rightarrow M_\infty$ dans L^1 , alors par continuité de $\mathbb{E}(\cdot | \mathcal{F}_n)$ sur L^1 ,

$$\mathbb{E}(M_k | \mathcal{F}_n) \xrightarrow[k \rightarrow \infty]{} \mathbb{E}(M_\infty | \mathcal{F}_n) \quad \text{dans } L^1,$$

or la suite ci-dessus est constante égale à M_n , d'où $M_n = \mathbb{E}(M_\infty | \mathcal{F}_n)$ par unicité de la limite.

(1) \implies (2) est le point b) de la proposition 3.32.

(2) \implies (3) : On rappelle que, par le point "a)" de la Proposition 3.32, toute famille uniformément intégrable est bornée dans L^1 . D'après le théorème de convergence p.s. de martingale (Théorème 3.28), M_n converge donc p.s. vers une certaine limite notée M_∞ . Comme (M_n) est uniformément intégrable, la convergence p.s. entraîne la convergence dans L^1 par le point d) de la Proposition 3.32, donc M_n converge dans L^1 vers M_∞ . \square

On peut alors étendre le théorème d'arrêt à des temps non bornés :

Théorème 3.34 (Théorème d'arrêt, cas uniformément intégrable). *Soient S et T deux temps d'arrêt finis et X une sous-martingale (resp. surmartingale, resp. martingale), relativement à la même filtration $(\mathcal{F}_n)_n$. On suppose que*

(i) $S \leq T$ p.s.,

(ii) *la famille $(X_{n \wedge T})_{n \geq 0}$ est uniformément intégrable.*

Alors X_T et X_S sont intégrables et

$$\mathbb{E}(X_T | \mathcal{F}_S) \geq X_S \quad (\text{resp. } \leq, \text{ resp. } =).$$

En particulier, la conclusion reste vraie en remplaçant l'hypothèse (ii) par

(ii') *la famille $(X_n)_{n \geq 0}$ est uniformément intégrable.*

Preuve : Justifions d'abord qu'il suffit de montrer le cas martingale. Supposons que X est une sous-martingale. D'après la décomposition de Doob, on a $X_n = M_n + A_n$, avec M une (\mathcal{F}_n) -martingale et A un processus prévisible et croissant et nul en 0. En appliquant le Théorème 3.21 à la martingale M et au temps d'arrêt borné $n \wedge T$, on a

$$\mathbb{E}(A_{n \wedge T}) = \mathbb{E}(X_{n \wedge T}) \leq C,$$

où la borne C est donnée par l'uniforme intégrabilité (ii). Donc $\mathbb{E}(A_T) \leq C$ par convergence monotone, ce qui, au vu de l'inégalité $|M_{n \wedge T}| \leq |X_{n \wedge T}| + A_T$, implique que la famille $(M_{n \wedge T})_{n \geq 0}$ est uniformément intégrable. Si on avait montré le théorème pour le cas martingale, alors on saurait que M_T et M_S sont intégrables et que $\mathbb{E}(M_T | \mathcal{F}_S) = M_S$ p.s., d'où X_T et X_S intégrables (puisque $A_S \leq A_T \in L^1$) et $\mathbb{E}(X_T | \mathcal{F}_S) \geq \mathbb{E}(M_T | \mathcal{F}_S) + A_S = X_S$ p.s., ce qui conclut la preuve dans le cas sous-martingale. Le cas surmartingale s'en déduit car $-X$ est une sous-martingale.

Montrons donc le théorème dans le cas où X est une martingale. Comme $X_T = \lim_{n \rightarrow \infty} X_{n \wedge T}$ p.s. et que $(X_{n \wedge T})$ est uniformément intégrable, $X_{n \wedge T}$ converge dans L^1 vers X_T et a fortiori, $X_T \in L^1$.

Pour tous $k \leq n$, $S \wedge k \leq T \wedge n \leq n$ p.s. En appliquant le Théorème 3.21, on a

$$\mathbb{E}(X_{T \wedge n} | \mathcal{F}_{S \wedge k}) = X_{S \wedge k}.$$

Comme $X_{n \wedge T}$ converge dans L^1 vers X_T , on a pour tout k ,

$$\mathbb{E}(X_T | \mathcal{F}_{S \wedge k}) = X_{S \wedge k}.$$

En particulier, $(X_{S \wedge k})_{k \geq 0}$ est uniformément intégrable et a fortiori, $X_S \in L^1$. Maintenant, pour tout $A \in \mathcal{F}_S$,

$$\mathbb{E}(X_S \mathbf{1}_A) = \lim_{k \rightarrow \infty} \mathbb{E}(X_S \mathbf{1}_{A \cap \{S \leq k\}}) = \lim_{k \rightarrow \infty} \mathbb{E}(X_T \mathbf{1}_{A \cap \{S \leq k\}}) = \mathbb{E}(X_T \mathbf{1}_A),$$

où dans la 2^e égalité, on utilise le fait que $A \cap \{S \leq k\} \in \mathcal{F}_{S \wedge k}$ (vérification : pour tout $j \leq k-1$, $A \cap \{S \leq k\} \cap \{S \wedge k = j\} = A \cap \{S = j\} \in \mathcal{F}_j$ et $A \cap \{S \leq k\} \cap \{S \wedge k = k\} = A \cap \{S = k\} \in \mathcal{F}_k$).

Montrons enfin que (ii') implique (ii). Justifions que X_T est intégrable. On se place (quitte à considérer $-X$) dans le cas d'une sous-martingale. Notons $Y_n = X_{n \wedge T}$. On a, pour tout n , $|Y_n| = Y_n^+ + Y_n^- = Y_n^+ + (-Y_n + Y_n^+) = 2Y_n^+ - Y_n$, et $\mathbb{E}(Y_n) \geq \mathbb{E}(Y_0)$ car $(Y_n)_n$ est une sous-martingale, donc

$$\mathbb{E}(|Y_n|) = 2\mathbb{E}(Y_n^+) - \mathbb{E}(Y_n) \leq 2\mathbb{E}(Y_n^+) - \mathbb{E}(Y_0).$$

Comme $Y_n = X_{n \wedge T} \rightarrow X_T$ p.s., on déduit alors du lemme de Fatou que

$$\mathbb{E}(|X_T|) = \mathbb{E}\left(\lim_{n \rightarrow \infty} |Y_n|\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(|Y_n|) \leq 2 \sup_n \mathbb{E}(Y_n^+) - \mathbb{E}(Y_0) < \infty$$

d'après l'hypothèse, donc X_T est intégrable. Montrons alors (ii). Pour tout $\lambda > 0$,

$$\begin{aligned} \mathbb{E}(|X_{n \wedge T}| \mathbf{1}_{\{|X_{n \wedge T}| > \lambda\}}) &= \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > \lambda\}} \mathbf{1}_{\{n < T\}}) + \mathbb{E}(|X_T| \mathbf{1}_{\{|X_T| > \lambda\}} \mathbf{1}_{\{n \geq T\}}) \\ &\leq \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > \lambda\}}) + \mathbb{E}(|X_T| \mathbf{1}_{\{|X_T| > \lambda\}}) \end{aligned}$$

d'où

$$\sup_{n \geq 0} \mathbb{E}(|X_{n \wedge T}| \mathbf{1}_{\{|X_{n \wedge T}| > \lambda\}}) \leq \sup_{n \geq 0} \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > \lambda\}}) + \mathbb{E}(|X_T| \mathbf{1}_{\{|X_T| > \lambda\}})$$

et le membre de droite tend vers 0 quand $\lambda \rightarrow \infty$, par (ii') et par le fait que X_T est intégrable (convergence dominée). Donc le membre de gauche aussi : c'est ce qu'il nous fallait démontrer. \square

Chapitre 4

Vecteurs gaussiens

4.1 Rappels sur la loi normale unidimensionnelle

Loi normale standard

La loi normale standard, notée $\mathcal{N}(0, 1)$, est la loi de densité $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ sur \mathbb{R} . On rappelle un calcul pour vérifier que cela définit bien une densité (c.-à-d. que $\int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}$) :

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right) e^{-\frac{x^2}{2}} dx = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_{\mathbb{R} \times]0, 2\pi[} e^{-\frac{r^2}{2}} r dr d\theta = 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi \left[-e^{-\frac{r^2}{2}} \right]_{r=0}^{\infty} = 2\pi \end{aligned}$$

à l'aide du théorème de Fubini-Tonelli (2è égalité) et d'un changement de variable en coordonnées polaires (3è égalité).

Notons que tous les polynômes sont intégrables sous cette loi : $\int_{\mathbb{R}} |x|^n e^{-x^2/2} dx < \infty$ pour tout n .

De plus cette loi a pour espérance 0 (par parité de sa densité) et variance 1 (faire une intégration par parties en intégrant $x e^{-x^2/2}$ et en dérivant x).

Loi normale

Considérons l'image de la loi $\mathcal{N}(0, 1)$ par une application affine.

Si $Z \sim \mathcal{N}(0, 1)$, et $m \in \mathbb{R}$, $\sigma \in \mathbb{R}^*$, la variable aléatoire $X = m + \sigma Z$ suit la loi de densité $x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ (preuve par changement de variable) et a pour espérance m et variance σ^2 . Cette loi ne dépend que de m et σ^2 , on la note $\mathcal{N}(m, \sigma^2)$, appelée **loi normale de moyenne m et de variance σ^2** (ou d'écart-type σ , avec $\sigma > 0$). On dit qu'une variable aléatoire X suit une "loi normale" ou que X est "gaussienne" si elle suit l'une de ces lois. Il est naturel de donner un sens au cas particulier $\sigma = 0$: alors $X = m$ est constante. Ainsi, on définira $\mathcal{N}(m, 0) = \delta_m$.

Proposition 4.1. *Quelques propriétés des gaussiennes réelles :*

- (Stabilité par application affine) L'image d'une variable gaussienne par une application affine est gaussienne : si $X \sim \mathcal{N}(m, \sigma^2)$ alors $aX + b \sim \mathcal{N}(am + b, a^2\sigma^2)$ pour tous $a, b \in \mathbb{R}$.
- Si $X \sim \mathcal{N}(m, \sigma^2)$, sa fonction caractéristique est $\Phi_X : t \mapsto \mathbb{E}[e^{itX}] = e^{itm - \frac{1}{2}t^2\sigma^2}$.
- (Stabilité par convolution, ou par somme indépendante) La somme de variables gaussiennes indépendantes est gaussienne : si $X \sim \mathcal{N}(m_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ sont indépendantes, alors $X + Y \sim \mathcal{N}(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$.

Preuve : a) c'est évident par la définition de $\mathcal{N}(m, \sigma^2)$, car la composée de deux fonctions affines est affine.

b) On raisonne pour $Z \sim \mathcal{N}(0, 1)$, le cas $X \sim \mathcal{N}(m, \sigma^2)$ s'en déduit par $\Phi_X(t) = \mathbb{E}[e^{it(m+\sigma Z)}] = e^{itm} \Phi_Z(\sigma t)$. On a $\Phi'_Z(t) = \mathbb{E}[iZ e^{itZ}]$ par dérivation sous l'espérance (justifiée par $\mathbb{E}[|Z|] < \infty$), puis

$\Phi'_Z(t) = -t\Phi_Z(t)$ par intégration par parties, d'où $\Phi_Z(t) = Ce^{-t^2/2}$ pour un certain C par résolution d'équation différentielle, et $C = 1$ par $\Phi_Z(0) = 1$.

c) Cela se déduit de la fonction caractéristique : $\Phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \Phi_X(t)\Phi_Y(t) = e^{it(m_X+m_Y) - \frac{1}{2}t^2(\sigma_X^2+\sigma_Y^2)} = \Phi_W(t)$ où $W \sim \mathcal{N}(m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$. \square

Comme conséquence du calcul de la fonction caractéristique de $Z \sim \mathcal{N}(0, 1)$ on a aussi le calcul des moments de cette loi : on a

$$\Phi_Z(t) = e^{-t^2/2} = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} t^{2n}$$

et aussi, par échange entre série et intégrale (à justifier par Fubini-Lebesgue par exemple)

$$\Phi_Z(t) = \mathbb{E}[e^{itZ}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(iZ)^k}{k!} t^k\right] = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}[Z^k]}{k!} t^k = \sum_{n=0}^{\infty} \frac{(-1)^n \mathbb{E}[Z^{2n}]}{(2n)!} t^{2n}$$

(vu que $\mathbb{E}[Z^{2k+1}] = 0$ par symétrie) d'où, d'ailleurs, par unicité du développement en série entière de rayon de convergence > 0 (ici, de rayon infini),

$$\mathbb{E}[Z^{2n}] = \frac{(2n)!}{2^n n!} = \frac{(2n)(2n-1)\cdots 2 \cdot 1}{(2n)(2n-1)\cdots (2 \cdot 2)(2 \cdot 1)} = (2n-1)(2n-3)\cdots 3 \cdot 1.$$

4.2 Extension de la définition à \mathbb{R}^d

Loi normale standard multidimensionnelle

Soit $d \geq 1$. La loi normale standard sur \mathbb{R}^d , notée $\mathcal{N}(0, I_d)$ (où $0 \in \mathbb{R}^d$ et $I_d \in \mathcal{M}_d(\mathbb{R})$ est la matrice identité de taille $d \times d$) est la loi du vecteur aléatoire $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix}$ où Z_1, \dots, Z_d sont indépendantes et

de loi $\mathcal{N}(0, 1)$. Elle a donc pour densité $z \mapsto \frac{1}{(2\pi)^{d/2}} e^{-(z_1^2 + \dots + z_d^2)/2} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|z\|^2}{2}}$.

On remarque que cette densité est radiale, donc la loi $\mathcal{N}(0, I_d)$ est invariante par les rotations vectorielles, et plus généralement par les applications orthogonales : si $A \in O_n(\mathbb{R})$, alors $AZ \sim \mathcal{N}(0, I_d)$.

Proposition 4.2 (Cas de \mathbb{R}^2). *Si X, Y sont indépendantes et de loi $\mathcal{N}(0, 1)$, notons $R > 0$ et $\Theta \in [0, 2\pi[$ les coordonnées polaires de (X, Y) . Alors*

- R et Θ sont indépendantes,
- Θ suit la loi uniforme sur $[0, 2\pi]$, et
- R a pour fonction de répartition $F_R(t) = \mathbb{P}(R \leq t) = 1 - e^{-\frac{t^2}{2}}$ pour $t \geq 0$.

En particulier, on en déduit la méthode polaire de simulation de la loi $\mathcal{N}(0, 1)$: si U, V sont indépendantes et de loi uniforme sur $[0, 1]$, alors $X = \sqrt{-2 \ln U} \cos(2\pi V)$ et $Y = \sqrt{-2 \ln U} \sin(2\pi V)$ sont indépendantes, de loi $\mathcal{N}(0, 1)$.

On calcule immédiatement la fonction caractéristique :

$$\text{pour } t = (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \Phi_Z(t) = \mathbb{E}[e^{i\langle t, Z \rangle}] = \Phi_{Z_1}(t_1) \cdots \Phi_{Z_d}(t_d) = e^{-\frac{\|t\|^2}{2}}.$$

Loi normale multidimensionnelle

Considérons l'image de la loi $\mathcal{N}(0, I_d)$ par une application affine de \mathbb{R}^d dans \mathbb{R}^d . Si $Z \sim \mathcal{N}(0, I_d)$, et $m \in \mathbb{R}^d$, $A \in \mathcal{M}_d(\mathbb{R})$, la variable aléatoire $X = m + AZ$ a pour fonction caractéristique

$$\Phi_X(t) = e^{itm} \mathbb{E}[e^{i\langle t, AZ \rangle}] = e^{itm} \mathbb{E}[e^{i\langle A^T t, Z \rangle}] = e^{itm - \frac{1}{2} \|A^T t\|^2} = e^{itm - \frac{1}{2} t^T A A^T t} = e^{itm - \frac{1}{2} t^T \Gamma t}$$

où $\Gamma = AA^T$. Cette loi ne dépend que de m et de Γ , notons-là $\mathcal{N}(m, \Gamma)$. On note que m est la moyenne de X : $\mathbb{E}[X] = m + \mathbb{E}[AZ] = m + A\mathbb{E}[Z] = m$, et que Γ est sa matrice de covariance :

$$\Gamma_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d} = \mathbb{E}[(X-m)(X-m)^T] = \mathbb{E}[AZZ^T A^T] = A\mathbb{E}[ZZ^T]A^T = AI_d A^T = AA^T = \Gamma.$$

Une matrice de covariance est toujours symétrique ($\Gamma^T = \Gamma$) et positive (pour tout $t \in \mathbb{R}^d$, $t^T \Gamma t \geq 0$). Inversement, si Γ est une matrice symétrique positive, il existe $A \in \mathcal{M}_d(\mathbb{R})$ (non unique) tel que $AA^T = \Gamma$: algorithmiquement, on peut calculer A (triangulaire) par la méthode de Cholesky ; d'un point de vue théorique, on peut diagonaliser $\Gamma = PDP^T$ en base orthonormée, avec $P \in \mathcal{O}_n(\mathbb{R})$,

$$D = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_d^2 \end{pmatrix}, \text{ puis considérer } A = P \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{pmatrix} P^T.$$

On a donc défini la loi $\mathcal{N}(m, \Gamma)$ pour tout $m \in \mathbb{R}^d$ et Γ symétrique positive. C'est la loi normale de moyenne m et de matrice de covariance Γ . Un vecteur aléatoire X est dit "gaussien" s'il suit l'une de ces lois.

Pour simuler la loi $\mathcal{N}(m, \Gamma)$, on calcule donc A tel que $\Gamma = AA^T$, on simule $Z \sim \mathcal{N}(0, I_d)$ (les composantes de Z sont indépendantes, de loi $\mathcal{N}(0, 1)$), et on pose $X = m + AZ$.

Proposition 4.3. *Quelques propriétés des vecteurs gaussiens :*

- (Stabilité par application affine) Soit $d, d' \geq 1$ L'image d'un vecteur gaussien par une application affine $\mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ est un vecteur gaussien : si $X \sim \mathcal{N}(m, \Gamma)$ dans \mathbb{R}^d , et $A \in \mathcal{M}_{d', d}(\mathbb{R})$, $b \in \mathbb{R}^{d'}$, alors $AX + b \sim \mathcal{N}(Am + b, A\Gamma A^T)$.
- (Stabilité par convolution, ou par somme indépendante) La somme de vecteurs gaussiens de \mathbb{R}^d indépendants est gaussienne : si $X \sim \mathcal{N}(m_X, \Gamma_X)$ et $Y \sim \mathcal{N}(m_Y, \Gamma_Y)$ sont indépendantes, alors $X + Y \sim \mathcal{N}(m_X + m_Y, \Gamma_X + \Gamma_Y)$.
- (Support) Si $X \sim \mathcal{N}(0, \Gamma)$, alors p.s. $X \in \text{im } \Gamma$. Plus précisément, $\text{im } \Gamma$ est le support de X .
- (Densité) Si Γ est inversible, X a une densité donnée par

$$x \mapsto \frac{1}{(2\pi)^{d/2} \sqrt{|\det \Gamma|}} e^{-\frac{1}{2}(x-m)^T \Gamma^{-1} (x-m)}$$

Preuve : a) C'est clair par définition si A est carrée ($d' = d$). Mais le calcul de Φ_X précédent fonctionne aussi si A est rectangulaire et montre que l'on obtient un vecteur gaussien de $\mathbb{R}^{d'}$ dans ce cas. Comme la composée de deux applications affines est affine, on en déduit a). Vérifions la covariance : de façon générale, la matrice de covariance de AX est

$$\Gamma_{AX} = \mathbb{E}[(AX - \mathbb{E}[AX])(AX - \mathbb{E}[AX])^T] = \mathbb{E}[A(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T A^T] = A\Gamma_X A^T.$$

b) Cela s'obtient par les fonctions caractéristiques, comme en dimension 1.

c) Comme $X = AZ$, et Z a pour support \mathbb{R}^d , le support de X est l'image de A . Il reste à voir que $\text{im } A = \text{im } \Gamma$. D'une part, $\Gamma = AA^T$ donc $\text{im } \Gamma \subset \text{im } A$. D'autre part, si $x \in \ker \Gamma$ alors $AA^T x = 0$, d'où $0 = x^T AA^T x = \|A^T x\|^2$ et donc $A^T x = 0$, c'est-à-dire $x \in \ker A^T$, ce qui montre que $\ker \Gamma \subset \ker A^T$ d'où par la formule du rang $\dim \text{im } \Gamma = d - \dim \ker \Gamma \geq d - \dim \ker A^T = \dim \text{im } A^T = \dim \text{im } A$. Par inclusion et égalité des dimensions on a donc $\text{im } \Gamma = \text{im } A$.

d) s'obtient par changement de variable $x = Az$ dans le calcul de $\mathbb{E}[f(X)] = \mathbb{E}[f(m + AZ)] = \int \dots$ pour toute fonction f mesurable positive. \square

Proposition 4.4. *Un vecteur aléatoire X est gaussien si, et seulement si toutes les combinaisons linéaires de ses composantes sont gaussiennes.*

Preuve : Une combinaison linéaire des composantes est une application linéaire $\mathbb{R}^d \rightarrow \mathbb{R}$, donc si X est gaussien alors les combinaisons linéaires de ses composantes sont gaussiennes.

Inversement, si la propriété est vraie pour X , alors pour tout $t \in \mathbb{R}^d$ la variable $\langle t, X \rangle = t_1 X_1 + \dots + t_d X_d$ est gaussienne, de moyenne $\langle t, m \rangle$ (où $m = \mathbb{E}[X]$) et de variance

$$\mathbb{E}[\langle t, X - m \rangle^2] = \mathbb{E}[t^T (X - m)(X - m)^T t] = t^T \Gamma_X t$$

(où Γ_X est la matrice de covariance de X) donc la fonction caractéristique de X en t vaut $\mathbb{E}[e^{i\langle t, X \rangle}] = \Phi_{(t, X)}(1) = e^{i\langle t, m \rangle - \frac{1}{2}t^T \Gamma_X t}$ d'après le cas unidimensionnel, ce qui est la fonction caractéristique de $\mathcal{N}(m_X, \Gamma_X)$; ceci est donc la loi de X . \square

On dira plus généralement qu'une famille infinie $(X_i)_{i \in I}$ de variables aléatoires réelles est un **processus gaussien** si toute sous-famille finie $(X_i)_{i \in J}$ ($J \subset I$) est un vecteur gaussien. Par la proposition précédente, on conclut que $(X_i)_{i \in I}$ un processus gaussien si, et seulement si toutes les combinaisons linéaires de ses composantes sont gaussiennes.

4.3 Indépendance et conditionnement

Proposition 4.5. Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ et $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_l \end{pmatrix}$ deux vecteurs gaussiens (dans \mathbb{R}^k et \mathbb{R}^l).

a) On suppose X et Y indépendants. Alors $W = \begin{pmatrix} X \\ Y \end{pmatrix}$ est un vecteur gaussien de \mathbb{R}^{k+l} , de matrice de covariance donnée, par blocs, par $\Gamma_W = \begin{pmatrix} \Gamma_X & 0 \\ 0 & \Gamma_Y \end{pmatrix}$.

b) Inversement, on suppose que $\text{Cov}(X_i, Y_j) = 0$ pour $1 \leq i \leq k$, $1 \leq j \leq l$, et que $\begin{pmatrix} X \\ Y \end{pmatrix}$ est un vecteur gaussien de \mathbb{R}^{k+l} . Alors X et Y sont indépendants.

Preuve : a) On note que $W = \begin{pmatrix} X \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ Y \end{pmatrix}$ est la somme de deux vecteurs gaussiens indépendants (images de X et Y par des applications linéaires), donc est gaussien. La matrice de covariance vient uniquement du fait que, comme X_i et Y_j sont indépendantes, $\text{Cov}(X_i, Y_j) = 0$, pour tous i, j .

b) Cela vient directement de a), car la loi de $\begin{pmatrix} X \\ Y \end{pmatrix}$ est donnée par sa matrice de covariance et sa moyenne, et l'hypothèse garantit que $\begin{pmatrix} X \\ Y \end{pmatrix}$ a même moyenne et matrice de covariance que dans le cas a), donc même loi que sous l'hypothèse a), en particulier X et Y sont indépendantes. \square

On en déduit un calcul des espérances et loi conditionnelles au sein de vecteurs gaussiens.

Proposition 4.6. Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur gaussien centré (c'est-à-dire d'espérance nulle), avec $X \in \mathbb{R}$ et $Y \in \mathbb{R}^l$. Alors $\mathbb{E}[X | Y]$ est la projection orthogonale dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de X sur $\text{Vect}(Y_1, \dots, Y_l)$.

Si le vecteur n'est pas centré, alors $\mathbb{E}[X | Y]$ est la projection orthogonale de X sur $\text{Vect}(1, Y_1, \dots, Y_l)$. Dans les deux cas, $X - \mathbb{E}[X | Y]$ est indépendant de Y_1, \dots, Y_l .

\rightsquigarrow On rappelle qu'en général, $\mathbb{E}[X | Y]$ est la fonction mesurable de Y la plus proche de X au sens de la norme L^2 . Ce résultat énonce qu'il suffit de considérer les combinaisons affines de Y_1, \dots, Y_l .

Le cas où X, Y ne sont pas centrés vient du cas centré : on l'applique à $X - \mathbb{E}[X]$ et $Y - \mathbb{E}[Y]$ (on note que $\sigma(Y) = \sigma(Y - \mathbb{E}[Y])$) : il existe $\alpha_1, \dots, \alpha_l \in \mathbb{R}$ tels que

$$\begin{aligned} \mathbb{E}[X | Y] &= \mathbb{E}[X] + \mathbb{E}\left[X - \mathbb{E}[X] \mid Y - \mathbb{E}[Y]\right] = \mathbb{E}[X] + \alpha_1(Y_1 - \mathbb{E}[Y_1]) + \dots + \alpha_l(Y_l - \mathbb{E}[Y_l]) \\ &= a + \alpha_1 Y_1 + \dots + \alpha_l Y_l \end{aligned}$$

où a se détermine grâce à l'espérance (elle vaut $\mathbb{E}[X]$), et on peut déterminer $\alpha_1, \dots, \alpha_l$ grâce aux covariances : pour tout i ,

$$0 = \text{Cov}(\mathbb{E}[X | Y] - X, Y_i) = \text{Cov}(\alpha_1 Y_1 + \dots + \alpha_l Y_l - X, Y_i) = \dots$$

Preuve : On considère le cas centré. Notons \tilde{X} la projection orthogonale de X sur $\text{Vect}(Y_1, \dots, Y_l)$ et vérifions que $\mathbb{E}[X | Y] = \tilde{X}$. Le vecteur $\begin{pmatrix} X - \tilde{X} \\ Y \end{pmatrix}$ est un vecteur gaussien (c'est l'image de $\begin{pmatrix} X \\ Y \end{pmatrix}$ par une application linéaire) et, pour $i = 1, \dots, n$,

$$\text{Cov}(X - \tilde{X}, Y_i) = \mathbb{E}[(X - \tilde{X})Y_i] = \langle X - \tilde{X}, Y_i \rangle_{L^2} = 0$$

par définition de la projection orthogonale. Par la propriété précédente (point b)), on en déduit que $X - \tilde{X}$ et Y sont indépendants. En particulier, on calcule alors

$$\mathbb{E}[X | Y] = \mathbb{E}[\tilde{X} + (X - \tilde{X}) | Y] = \tilde{X} + \mathbb{E}[X - \tilde{X}] = \tilde{X},$$

en utilisant le fait que \tilde{X} est une fonction (linéaire) de Y_1, \dots, Y_l et que $X - \tilde{X}$ est indépendant de Y . \square

Plus précisément, on a :

Proposition 4.7. Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur gaussien, avec $X \in \mathbb{R}$ et $Y \in \mathbb{R}^l$. Alors la loi conditionnelle de X sachant Y est la loi $\mathcal{N}(\mathbb{E}[X | Y], \text{Var}(X - \mathbb{E}[X | Y]))$.

On remarquera que la variance est déterministe. En pratique, on calcule $\mathbb{E}[X | Y] = a + \alpha_1 Y_1 + \dots + \alpha_l Y_l$ par la proposition précédente, puis on calcule la variance de $X - \alpha_1 Y_1 - \dots - \alpha_l Y_l$.

Preuve : Quitte à remplacer X par $X - \mathbb{E}[X]$, on peut supposer X centré. En poursuivant la preuve précédente, on constate que

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

avec $\mathbb{E}[X | Y]$ qui est $\sigma(Y)$ -mesurable, et $X - \mathbb{E}[X | Y]$ qui est indépendante de Y , et de loi $\mathcal{N}(0, \text{Var}(X - \mathbb{E}[X | Y]))$ (car c'est une variable gaussienne). La proposition en résulte directement : formellement, si on pose $\tilde{X} = \mathbb{E}[X | Y] = h(Y)$ et $\hat{X} = X - \tilde{X}$, alors \hat{X} est indépendante de Y et de loi $\mathcal{N}(0, \text{Var}(X - \mathbb{E}[X | Y]))$ donc pour toute fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}_+$, par la proposition 2.10,

$$\mathbb{E}[f(X) | Y] = \mathbb{E}[f(h(Y) + \hat{X}) | Y] = g(Y), \quad \text{où } g(y) = \mathbb{E}[f(h(y) + \hat{X})] = \mathbb{E}[f(X_y)],$$

avec X_y une variable aléatoire de loi $\mathcal{N}(h(y), \text{Var}(X - \mathbb{E}[X | Y]))$. C'est précisément le sens de l'énoncé. \square

NB. Ces propriétés s'étendent au cas où X est dans \mathbb{R}^k : $\mathbb{E}\left[\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \middle| Y\right] = \begin{pmatrix} \mathbb{E}[X_1 | Y] \\ \vdots \\ \mathbb{E}[X_k | Y] \end{pmatrix}$ donc

chaque composante se calcule par projection sur $\text{Vect}(Y_1, \dots, Y_l)$, puis $X - \mathbb{E}[X | Y]$ est indépendant de Y à nouveau, de loi $\mathcal{N}(0, \Gamma_{X - \mathbb{E}[X | Y]})$, donc la loi de X sachant Y est la loi $\mathcal{N}(\mathbb{E}[X | Y], \Gamma_{X - \mathbb{E}[X | Y]})$.

4.4 Lois normales et limites

Proposition 4.8. Soit $(X_n)_{n \geq 0}$ une suite de vecteurs gaussiens de \mathbb{R}^d , qui converge en loi. Alors la limite est un vecteur gaussien.

Preuve : Dans le cas $d = 1$: notons, pour tout n , $m_n = \mathbb{E}[X_n]$ et $\sigma_n^2 = \text{Var}(X_n)$. Notons aussi X une limite. Alors pour tout $t \in \mathbb{R}$, $\Phi_{X_n}(t) = e^{it m_n - \frac{1}{2} t^2 \sigma_n^2} \rightarrow \Phi_X(t)$. En prenant le module, on a $e^{-\frac{1}{2} t^2 \sigma_n^2} \rightarrow |\Phi_X(t)|$, et comme $\Phi_X(0) = 1$ on peut choisir t tel que $\Phi_X(t) \neq 0$, d'où la convergence de σ_n^2 vers une limite positive ou nulle σ^2 en prenant le logarithme. Pour les espérances, on constate que si une sous-suite $m_{\varphi(n)}$ converge vers une limite $\mu \in \mathbb{R}$, alors par passage à la limite on a, pour tout $t \in \mathbb{R}$, $\Phi_X(t) = e^{it\mu - \frac{1}{2} t^2 \sigma^2}$ donc X est gaussienne. Il suffit par conséquent de montrer que l'on n'a pas

$|m_n| \rightarrow +\infty$. Supposons par exemple que, quitte à extraire une sous-suite, $m_n \rightarrow +\infty$. Alors, pour tout réel x , pour tout n assez grand, $m_n > x$, et donc

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X_n \leq m_n) = \frac{1}{2}$$

par symétrie de la gaussienne, si bien que la limite de la suite des fonctions de répartition de X_n (en les points de continuité de F_X) ne peut pas être une fonction de répartition (elle doit avoir pour limite 1 en $+\infty$), en contradiction avec la convergence en loi.

Dans le cas $d \geq 2$: si $X_n \rightarrow X$ en loi, alors pour tout $t \in \mathbb{R}^d$, $\langle t, X_n \rangle \rightarrow \langle t, X \rangle$ en loi (par continuité), et ces variables sont gaussiennes, donc le cas $d = 1$ montre que $\langle t, X \rangle$ est gaussienne : toute combinaison linéaire des composantes de X est gaussienne, donc X est un vecteur gaussien. \square

Théorème 4.9 (Théorème Central Limite multidimensionnel). *Soit X_1, X_2, \dots une suite de vecteurs aléatoires indépendants et de même loi, à valeurs dans \mathbb{R}^d , de carré intégrable. On note m leur espérance et Γ leur matrice de covariance communes. Alors*

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \xrightarrow[n \rightarrow \infty]{(\text{loi})} \mathcal{N}(0, \Gamma).$$

Preuve : En notant $(W_n)_n$ la suite ci-dessus, il faut montrer que, pour tout $t \in \mathbb{R}^d$,

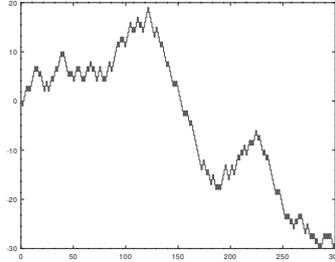
$$\Phi_{W_n}(t) = \mathbb{E}[e^{i\langle t, W_n \rangle}] \rightarrow e^{-\frac{1}{2}t^T \Gamma t},$$

mais

$$\langle t, W_n \rangle = \sqrt{n} \left(\frac{\langle t, X_1 \rangle + \dots + \langle t, X_n \rangle}{n} - \langle t, m \rangle \right),$$

et les variables $\langle t, X_i \rangle$ sont indépendantes, de même loi d'espérance $\langle t, m \rangle$ et de variance $t^T \Gamma t$ donc la convergence ci-dessus vient du TCL unidimensionnel. \square

On s'intéresse maintenant à une "limite d'échelle" de ce processus lorsqu'on accélère le temps : l'intervalle de temps entre deux sauts est $1/N$, et on souhaite faire tendre N vers l'infini. Autrement dit, on s'intéresse au processus $S^{(N)} = (S_{\lfloor Nt \rfloor})_{t \geq 0}$. Ce processus prend typiquement, à un temps donné, des valeurs de plus en plus grandes lorsque N augmente : ainsi, $S_1^{(N)} = S_N = X_1 + \dots + X_N$. L'accélération du temps nécessite donc une normalisation des valeurs prises par le processus, pour espérer obtenir une convergence et non une explosion.



Par le théorème central limite, $\frac{S_N}{\sqrt{N}}$ converge en loi vers la loi normale $\mathcal{N}(0, 1)$. Ceci suggère que la bonne normalisation est $1/\sqrt{N}$: on va étudier la suite de processus $B^{(N)} = (X_t^{(N)})_{t \geq 0}$ où

$$B_t^{(N)} = \frac{1}{\sqrt{N}} S_{\lfloor Nt \rfloor}, \quad t \geq 0.$$

On a alors

$$B_1^{(N)} = \frac{S_N}{\sqrt{N}} \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}(0, 1)$$

et, plus généralement, pour tout $t \geq 0$,

$$B_t^{(N)} = \frac{S_{\lfloor Nt \rfloor}}{\sqrt{N}} = \frac{S_{\lfloor Nt \rfloor}}{\sqrt{\lfloor Nt \rfloor}} \frac{\sqrt{\lfloor Nt \rfloor}}{\sqrt{N}} \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}(0, t),$$

car si $Z \sim \mathcal{N}(0, 1)$ alors $Z\sqrt{t} \sim \mathcal{N}(0, t)$. On a donc une convergence ponctuelle en loi. Pour décrire la loi de l'éventuel processus limite, il faut également décrire les lois jointes en différents temps. On observe tout d'abord que, si $0 < s < t$, alors on peut décomposer

$$B_t^{(N)} = \frac{X_1 + \dots + X_{\lfloor Ns \rfloor}}{\sqrt{N}} + \frac{X_{\lfloor Ns \rfloor + 1} + \dots + X_{\lfloor Nt \rfloor}}{\sqrt{N}} = B_s^{(N)} + B_{s,t}^{(N)}$$

où $B_{s,t}^{(N)} = B_t^{(N)} - B_s^{(N)}$, et vu l'écriture précédente on observe que

$$B_s^{(N)} \text{ et } B_{s,t}^{(N)} \text{ sont indépendants pour tout } N, B_s^{(N)} \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}(0, s), \text{ et } B_{s,t}^{(N)} \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}(0, t - s),$$

si bien que

$$(B_s^{(N)}, B_{s,t}^{(N)}) \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}\left(0, \begin{pmatrix} s & 0 \\ 0 & t - s \end{pmatrix}\right)$$

De même, pour tous $0 < t_1 < \dots < t_k$, les accroissements entre ces temps sont indépendants et ont une limite gaussienne :

$$(B_{t_1}^{(N)}, B_{t_1, t_2}^{(N)}, \dots, B_{t_k - t_{k-1}}^{(N)}) \xrightarrow[N \rightarrow \infty]{\text{(loi)}} \mathcal{N}\left(0, \begin{pmatrix} t_1 & 0 & \dots & 0 \\ 0 & t_2 - t_1 & 0 & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & t_k - t_{k-1} \end{pmatrix}\right).$$

Le mouvement brownien $(B_t)_{t \geq 0}$ va être défini comme un processus “limite” de $B^{(N)}$: tel que, pour tous $0 < t_1 < \dots < t_k$,

$$(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}) \sim \mathcal{N}\left(0, \begin{pmatrix} t_1 & 0 & \dots & 0 \\ 0 & t_2 - t_1 & 0 & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & t_k - t_{k-1} \end{pmatrix}\right).$$

Il s'avère que préciser ces lois (autrement dit les lois de $(B_{t_1}, \dots, B_{t_k})$) pour tous $k \in \mathbb{N}^*$ et $0 < t_1 < \dots < t_k$ ne donne pas toute l'information que l'on souhaite connaître sur un processus indexé par $[0, \infty[$, en raison du caractère *indénombrable* de cet ensemble : par exemple, l'existence d'une limite $\lim_{t \rightarrow 0^+} B_t$ dépend d'une infinité non dénombrable de composantes donc ne se réduit pas aux lois précédentes. Sans définir de notion de loi de processus (on en dira juste un mot plus loin), on notera simplement qu'il suffira, pour étudier toutes les propriétés qui suivront, d'ajouter une condition de continuité (point (iv) ci-dessous).

5.1.2 Définition

Définition 5.1. *Un processus $(B_t)_{t \in [0, \infty[}$ est un **mouvement brownien réel (standard)** si*

- (i) $B_0 = 0$ p.s. ;
- (ii) pour tout $k \in \mathbb{N}^*$ et $0 < t_1 < \dots < t_k$, les accroissements $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}$ sont indépendants ;
- (iii) pour tous $0 < s < t$, $B_t - B_s$ suit la loi $\mathcal{N}(0, t - s)$;
- (iv) p.s., la trajectoire $t \mapsto B_t$ est continue sur $[0, \infty[$.

Plus généralement, pour $x \in \mathbb{R}$ et $\sigma > 0$, un mouvement brownien **issu de x** et de **diffusivité (ou volatilité) σ^2** est un processus vérifiant (ii), (iv) et

$$(i') \quad B_0 = x \text{ p.s.,}$$

$$(iii') \quad \text{pour tous } 0 < s < t, B_t - B_s \text{ suit la loi } \mathcal{N}(0, \sigma^2(t - s)).$$

On observe immédiatement que, si B est un mouvement brownien standard, alors $(x + \sigma B_t)_{t \geq 0}$ est un mouvement brownien issu de x et de diffusivité σ^2 .

Définition 5.2. *Un processus $(B(t))_{t \in [0, \infty[} = (B_1(t), \dots, B_d(t))_{t \in [0, \infty[}$ à valeurs dans \mathbb{R}^d est un **mouvement brownien d -dimensionnel (standard)** si B_1, \dots, B_d sont des mouvements browniens réels standards indépendants.*

On pourrait plus généralement définir un mouvement brownien d -dimensionnel de diffusivité D (matrice symétrique positive) par les conditions (i), (ii), (iv) et

$$(iii'') \quad \text{pour tous } 0 < s < t, B_t - B_s \text{ suit la loi } \mathcal{N}(0, (t - s)D).$$

Le cas standard est donc le cas $D = I_d$.

Revenons au cas réel (on ne parlera plus du cas d -dimensionnel dans la suite). Les conditions (ii) et (iii) montrent que $(B_{t_1}, \dots, B_{t_k})$ est un vecteur gaussien (c'est une fonction linéaire de $(B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}})$) et qu'on a, pour tous $s < t$,

$$\text{Cov}(B_s, B_t) = \text{Cov}(B_s, B_s + (B_t - B_s)) = \text{Var}(B_s) + \text{Cov}(B_s, B_t - B_s) = \text{Var}(B_s) = s$$

donc, pour tous $s, t > 0$, $\text{Cov}(B_s, B_t) = s \wedge t$. Comme la matrice de covariance caractérise la loi d'un vecteur gaussien centré, les points (ii) et (iii) sont équivalents à

$$(ii+iii) \quad (B_t)_{t \geq 0} \text{ est un processus gaussien de covariance } \Gamma : (s, t) \mapsto s \wedge t.$$

5.1.3 Construction

Nous avons *motivé* la définition de B comme “limite” de marches aléatoires, mais nous n’avons pas *prouvé* la convergence de $B^{(N)}$ comme variables aléatoires et en particulier nous n’avons pas construit de limite. Nous avons prouvé la convergence les *lois fini-dimensionnelles*, c’est-à-dire des lois de tous les vecteurs $(B_{t_1}^{(N)}, \dots, B_{t_k}^{(N)})$, mais il faudrait une convergence en un sens plus fort pour assurer l’existence d’un processus limite qui vérifie (i),(ii),(iii), et de toute façon on a déjà remarqué que le travail sur ces lois fini-dimensionnelles ne peut suffire à obtenir (iv). Il reste donc à justifier l’existence d’un mouvement brownien. On en propose une construction qui sera instructive.

On se contentera de construire $(B_t)_{t \in [0,1]}$. Il resterait ensuite à en concaténer des copies indépendantes pour obtenir le processus sur $[0, \infty[$.

On va définir $(B_t)_{t \in [0,1]}$ comme limite uniforme d’une suite de fonctions aléatoires $X^{(0)}, X^{(1)}, \dots$ continues, affines par morceaux, et telles que

- (1) pour tout n , $X^{(n)}$ est continue, affine sur chacun des intervalles de la subdivision définie par les points dyadiques

$$D_n = \left\{ \frac{k}{2^n} \mid 0 \leq k \leq 2^n \right\};$$

- (2) pour tout n , le vecteur $(X_t^{(n)})_{t \in D_n}$ a la loi attendue, ce qui équivaut à :

$$X_0^{(n)} = 0, \quad \text{et} \quad \left(X_{\frac{k+1}{2^n}}^{(n)} - X_{\frac{k}{2^n}}^{(n)} \right)_{k=0, \dots, N-1} \text{ sont indépendants et de loi } \mathcal{N}(0, 2^{-n});$$

- (3) pour tout n , $X^{(n+1)}$ coïncide avec $X^{(n)}$ sur D_n : pour $k = 0, \dots, 2^n$, $X^{(n+1)}(\frac{k}{2^n}) = X^{(n)}(\frac{k}{2^n})$.

Pour cela on peut construire $X^{(0)}, X^{(1)}, \dots$ par récurrence : il suffit, pour tout n , de se donner des valeurs de $X^{(n+1)}$ sur $D_{n+1} \setminus D_n$ qui assurent le point (2), les valeurs de $X^{(n+1)}$ sur D_n étant celles de $X^{(n)}$ vu (3), et les valeurs intermédiaires se déduisant par interpolation linéaire vu (1).

On commence par définir $X^{(0)}$ par $X_t^{(0)} = tZ$ où $Z \sim \mathcal{N}(0, 1)$. En effet $D_0 = \{0, 1\}$.

Puis, pour $n \in \mathbb{N}$ donné, supposons $X^{(n)}$ construit, vérifiant (1) et (2). On définit donc $X^{(n+1)}$ sur D_n par les valeurs de $X^{(n)}$ (cf. (3)), et il reste à le définir sur $D_{n+1} \setminus D_n$. Pour que $X_{|D_{n+1}}^{(n+1)}$ ait la loi voulue, il suffit que $X_{|D_n}^{(n+1)}$ ait la loi voulue (c’est le cas par (2) vu la récurrence) et que la loi conditionnelle de $X_{|D_{n+1} \setminus D_n}^{(n+1)}$ sachant $X_{|D_n}^{(n+1)}$ soit celle voulue. Déterminons donc cette loi conditionnelle.

Pour simplifier les notations, on écrit ici $(B_t)_{t \in D_{n+1}}$ pour désigner un vecteur ayant la loi voulue (autrement dit un vecteur gaussien de covariance $\text{Cov}(B_s, B_t) = s \wedge t$ pour $s, t \in D_{n+1}$, ou de façon équivalente $B_0 = 0$ et les accroissements $B_{(k+1)/2^{n+1}} - B_{k/2^{n+1}}$ sont indépendants et de loi $\mathcal{N}(0, 2^{-(n+1)})$). Comme c’est un vecteur gaussien, on sait (chapitre précédent) que la loi de $(B_t)_{t \in D_{n+1} \setminus D_n}$ sachant $(B_t)_{t \in D_n}$ est la loi

$$\mathcal{N}(\mathbb{E}(X | Y), \text{Var}(X - \mathbb{E}(X | Y))), \quad \text{où } X = (B_t)_{t \in D_{n+1} \setminus D_n} \text{ et } Y = (B_t)_{t \in D_n},$$

où ici “Var” désigne la matrice de covariance. On calcule l’espérance conditionnelle par composante : pour $0 \leq k < 2^{n+1}$, on a

$$B_{\frac{2k+1}{2^{n+1}}} = B_{\frac{2k}{2^{n+1}}} + \left(B_{\frac{2k+1}{2^{n+1}}} - B_{\frac{2k}{2^{n+1}}} \right) = B_{\frac{k}{2^n}} + \left(B_{\frac{2k+1}{2^{n+1}}} - B_{\frac{2k}{2^{n+1}}} \right)$$

donc

$$\mathbb{E}\left(B_{\frac{2k+1}{2^{n+1}}} \mid B_{|D_n} \right) = B_{\frac{k}{2^n}} + \mathbb{E}\left(B_{\frac{2k+1}{2^{n+1}}} - B_{\frac{2k}{2^{n+1}}} \mid B_{\frac{k+1}{2^n}} - B_{\frac{k}{2^n}} \right).$$

En effet, $\frac{k}{2^n} \in D_n$ donc le premier terme est $\sigma(B_{|D_n})$ -mesurable. Et, pour le second terme, conditionner par $B_{|D_n}$ revient à conditionner par les accroissements de B entre les points de D_n , qui sont tous indépendants de l’accroissement dont on prend l’espérance, *sauf* celui de l’intervalle $[\frac{k}{2^n}, \frac{k+1}{2^n}]$. On utilise ici implicitement le lemme suivant :

Lemme 5.3. *Si X est \mathcal{F} -mesurable et intégrable, et \mathcal{G}, \mathcal{H} sont des tribus telles que $\sigma(\mathcal{F} \cup \mathcal{G})$ et \mathcal{H} sont indépendantes, alors $\mathbb{E}(X | \mathcal{G}, \mathcal{H}) = \mathbb{E}(X | \mathcal{G})$.*

NB. On note $\mathbb{E}(X | \mathcal{G}, \mathcal{H}) = \mathbb{E}(X | \sigma(\mathcal{G} \cup \mathcal{H}))$.

Preuve : En effet, pour toute variable aléatoire bornée Z de la forme $Z = GH$, où G et H sont des variables aléatoires bornées respectivement \mathcal{G} - et \mathcal{H} -mesurables, on a, d'après les hypothèses,

$$\mathbb{E}(XZ) = \mathbb{E}(XGH) = \mathbb{E}(XG)\mathbb{E}(H) = \mathbb{E}(\mathbb{E}(X | \mathcal{G})G)\mathbb{E}(H) = \mathbb{E}(\mathbb{E}(X | \mathcal{G})GH) = \mathbb{E}(\mathbb{E}(X | \mathcal{G})Z).$$

On admettra que ce cas particulier suffit à conclure. Cela vient du fait que les événements $A \cap B$, où $A \in \mathcal{G}$ et $B \in \mathcal{H}$, engendrent $\sigma(\mathcal{G} \cup \mathcal{H})$. \square

Si on note Z_1 et Z_2 les accroissements de B sur les intervalles $[\frac{k}{2^n}, \frac{2k+1}{2^{n+1}}]$ et $[\frac{2k+1}{2^{n+1}}, \frac{k+1}{2^n}]$, alors l'espérance conditionnelle qui reste à calculer s'écrit

$$\mathbb{E}(Z_1 | Z_1 + Z_2).$$

Or Z_1 et Z_2 sont indépendantes et de même loi. Il en résulte (vu en TD) que

$$\mathbb{E}(Z_1 | Z_1 + Z_2) = \frac{Z_1 + Z_2}{2} = \frac{1}{2}(B_{\frac{k+1}{2^n}} - B_{\frac{k}{2^n}}),$$

donc finalement

$$\mathbb{E}(B_{\frac{2k+1}{2^{n+1}}} | B|_{D_n}) = B_{\frac{k}{2^n}} + \frac{1}{2}(B_{\frac{k+1}{2^n}} - B_{\frac{k}{2^n}}) = \frac{B_{\frac{k}{2^n}} + B_{\frac{k+1}{2^n}}}{2}.$$

L'espérance conditionnelle en un point de $D_{n+1} \setminus D_n$ sachant les valeurs sur D_n est donc simplement la moyenne des valeurs aux deux points de D_n qui l'encadrent.

Calculons la matrice Σ de covariance de la loi conditionnelle, c'est-à-dire la matrice du covariance du vecteur formé par les différences entre les B_t , $t = \frac{2k+1}{2^{n+1}}$ et les moyennes des deux points de D_n qui l'encadrent (c'est-à-dire $\frac{k}{2^n}$ et $\frac{k+1}{2^n}$).

Remarquons que, quels que soient $0 \leq s < t$,

$$B_{\frac{s+t}{2}} - \frac{B_s + B_t}{2} = \frac{B_{\frac{s+t}{2}} - B_s}{2} + \frac{B_t - B_{\frac{s+t}{2}}}{2},$$

donc cette variable ne dépend que des accroissements sur $[s, \frac{s+t}{2}]$ et $[\frac{s+t}{2}, t]$, qui sont dans $[s, t]$. Par suite, vu l'indépendance des incréments sur des intervalles disjoints, la matrice de covariance Σ est diagonale, et les coefficients diagonaux se déduisent de

$$\text{Var}(B_{\frac{s+t}{2}} - \frac{B_s + B_t}{2}) = \text{Var}(\frac{B_{\frac{s+t}{2}} - B_s}{2} + \frac{B_t - B_{\frac{s+t}{2}}}{2}) = \frac{1}{4}(\frac{s+t}{2} - s) + \frac{1}{4}(t - \frac{s+t}{2}) = \frac{t-s}{4},$$

c'est-à-dire qu'ils sont donc tous égaux à $\frac{2^{-n}}{4}$.

En résumé, on veut que la loi du vecteur $(X_{\frac{2k+1}{2^{n+1}}}^{(n+1)})_{0 \leq k < 2^n}$ sachant $X^{(n)}$ soit celle du vecteur

$$\left(\frac{X_{\frac{k}{2^n}}^{(n)} + X_{\frac{k+1}{2^n}}^{(n)}}{2} + \frac{\sqrt{2^{-n}}}{2} Z_{n,k} \right)_{0 \leq k \leq 2^n - 1},$$

où les variables aléatoires $(Z_{n,k})_{k,n}$ sont indépendantes et de loi $\mathcal{N}(0, 1)$. Vu que $X^{(n)}$ est définie par interpolation affine, on note que le vecteur précédent est aussi exactement

$$\left(X_{\frac{2k+1}{2^{n+1}}}^{(n)} + \frac{\sqrt{2^{-n}}}{2} Z_{n,k} \right)_{0 \leq k \leq 2^n - 1}.$$

Vu ce calcul, on obtient donc les propriétés (1),(2),(3) en appliquant la formule précédente : on se donne une famille $(Z_{n,k})_{n,k}$ de variables indépendantes et de loi $\mathcal{N}(0, 1)$ et, pour tout n , on construit $X^{(n+1)}$ par

- $X^{(n+1)} = X^{(n)}$ sur D_n ;

- pour $k = 0, \dots, 2^n - 1$, $X_{\frac{2k+1}{2^{n+1}}}^{(n+1)} = X_{\frac{2k+1}{2^{n+1}}}^{(n)} + \frac{\sqrt{2^{-n}}}{2} Z_{n,k}$
- $X^{(n+1)}$ est prolongé continûment et de façon affine entre les valeurs précédentes.

On prouve maintenant que, presque sûrement, la suite $(X^{(n)})_n$ converge uniformément sur $[0, 1]$. Pour tout n , on constate par construction que c'est aux points de $D_{n+1} \setminus D_n$ que la différence $|X^{(n+1)} - X^{(n)}|$ est maximale, et donc

$$\|X^{(n+1)} - X^{(n)}\|_\infty = \max_{0 \leq k < 2^n} \frac{\sqrt{2^{-n}}}{2} |Z_{n,k}|.$$

On utilise le lemme suivant pour majorer l'espérance de ce terme :

Lemme 5.4. *Il existe $K > 0$ tel que, pour tout n , si Z_1, \dots, Z_n sont des variables aléatoires indépendantes et de loi $\mathcal{N}(0, 1)$, alors*

$$\mathbb{E}(\max(|Z_1|, \dots, |Z_n|)) \leq K \sqrt{\log n}.$$

Preuve : En notant $\phi : x \mapsto e^{\frac{x^2}{4}}$, la fonction ϕ est convexe, croissante sur \mathbb{R}_+ , donc par l'inégalité de Jensen on a (avec l'inégalité évidente $\max(a, b) \leq a + b$ si $a, b \geq 0$) :

$$\phi(\mathbb{E}(\max_i |Z_i|)) \leq \mathbb{E}(\phi(\max_i |Z_i|)) \leq \mathbb{E}(\phi(|Z_1| + \dots + |Z_n|)) \leq \mathbb{E}(\phi(|Z_1|) + \dots + \phi(|Z_n|)) = n\mathbb{E}(\phi(|Z_1|))$$

donc, en notant $C = \mathbb{E}(\phi(|Z_1|))$ (qui est fini, vu la densité gaussienne), on en déduit, pour tout $n \geq 2$,

$$\mathbb{E}(\max_i |Z_i|) = \sqrt{4 \log \phi(\mathbb{E}(\max_i |Z_i|))} \leq \sqrt{4 \log(nC)} \leq K \sqrt{\log n},$$

pour une certaine constante K . □

Ainsi,

$$\mathbb{E}(\|X^{(n+1)} - X^{(n)}\|_\infty) \leq 2^{-\frac{n}{2}-1} K \sqrt{\log(2^n)} = 2^{-\frac{n}{2}-1} K \sqrt{n} \sqrt{\log 2}.$$

En particulier, on en déduit que (avec le théorème de convergence monotone pour la première étape)

$$\mathbb{E}\left(\sum_{n=0}^{\infty} \|X^{(n+1)} - X^{(n)}\|_\infty\right) = \sum_{n=0}^{\infty} \mathbb{E}(\|X^{(n+1)} - X^{(n)}\|_\infty) < \infty.$$

Par suite,

$$\text{presque sûrement, } \sum_{n=0}^{\infty} \|X^{(n+1)} - X^{(n)}\|_\infty < \infty.$$

Cette condition implique que la suite $(X^{(n)})_n$ converge uniformément (si une série converge absolument dans l'espace complet $\mathcal{C}([0, 1])$ alors elle converge dans cet espace, ce qui revient ici à la convergence uniforme de $(X^{(n)})_n$) d'où la conclusion : presque sûrement, $X^{(n)}$ converge uniformément, quand $n \rightarrow \infty$, vers une fonction B qui est donc continue.

Par construction, pour tout n , pour tous $0 \leq t_1 < \dots < t_k$ dans D_n , $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}}$ sont égaux à $X_{t_1}^{(n)}, X_{t_2}^{(n)} - X_{t_1}^{(n)}, \dots, X_{t_k}^{(n)} - X_{t_{k-1}}^{(n)}$ et donc indépendants et de lois $\mathcal{N}(0, t_1), \mathcal{N}(0, t_2 - t_1), \dots, \mathcal{N}(0, t_k - t_{k-1})$.

Cette propriété s'étend à tous $0 \leq t_1 < \dots < t_k \leq 1$ par densité de $\bigcup_n D_n$ dans $[0, 1]$. En effet, cela peut se voir via la fonction caractéristique : la fonction caractéristique de $(B_{t_1}, B_{t_2} - B_{t_1}, \dots)$ est la limite de celles prises en des points dyadiques qui approchent t_1, t_2, \dots par continuité de B (et théorème de convergence dominée), or ces fonctions caractéristiques sont explicites vu le cas dyadique, et convergent vers celle de la loi attendue.

Cela achève de prouver que B est un mouvement brownien (sur l'intervalle de temps $[0, 1]$).

Remarque 5.5. *On a montré ici que $(X^{(n)})_n$ converge presque sûrement vers un mouvement brownien, dans l'espace vectoriel normé $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$. En particulier, on pourrait dire que $(X^{(n)})_n$ converge en loi vers le mouvement brownien, en tant que variable aléatoire à valeurs dans cet espace, muni de sa tribu des boréliens.*

Mais a-t-on montré que la marche aléatoire $B^{(N)}$ de l'introduction "converge en loi" vers le mouvement brownien ? Pour cela, il faudrait voir $B^{(N)}$ comme une variable aléatoire à valeurs dans un espace mesuré. On ne peut pas utiliser $\mathcal{C}([0, 1])$ car $B^{(N)}$ n'est pas continue. On pourrait a priori vouloir considérer tout l'espace $\mathbb{R}^{[0,1]}$ avec sa tribu produit $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$, et dans ce cas on aurait bien montré la convergence en loi en montrant la convergence des lois jointes d'un nombre fini de marginales ; cependant les parties mesurables ne dépendent que d'une infinité dénombrable de composantes, ce qui empêche de parler de limites ou de continuité et rend cet espace peu approprié. Une bonne solution est de considérer un espace normé de fonctions ayant des discontinuités (fonctions "càd-làg" : continues à droites, avec limites à gauche). Cela dit, notre argument ne suffit alors pas à assurer la convergence en loi, qui requiert de s'assurer que certaines quantités mesurant la continuité de $B^{(N)}$ n'explorent pas quand N est grand.

La convergence en loi de $B^{(N)}$ vers un mouvement brownien peut néanmoins être démontrée (dans l'espace des fonctions càd-làg avec une bonne norme), et connue comme le **théorème de Donsker**, ou **théorème central limite fonctionnel**.

5.2 Propriétés

Soit $B = (B_t)_{t \geq 0}$ un mouvement brownien réel.

5.2.1 Régularité

On commence par des résultats qui illustrent plutôt l'**irrégularité** de B .

Proposition 5.6. *Presque sûrement, B n'est monotone sur aucun intervalle non trivial.*

Preuve : Soit un intervalle $I = [a, b]$ avec $a < b$ réels. Soit $n \in \mathbb{N}^*$. On note $a_{n,k} = a + \frac{k}{n}(b-a)$ pour $k = 0, \dots, n$. Si B est monotone sur I , alors en particulier les n accroissements $B_{a_{k+1}} - B_{a_k}$ (où $0 \leq k < n$) sont de même signe. Or ces variables sont indépendantes (par indépendance des accroissements sur des intervalles disjoints) et ont des signes uniformes dans $\{-1, +1\}$, donc cet événement a pour probabilité $2^{-(n-1)}$ (probabilité que n pièces tombent du même côté). Par suite,

$$\mathbb{P}(B \text{ est monotone sur } I) \leq 2^{-(n-1)}.$$

Ceci vaut quel que soit n , donc la probabilité est nulle. On a obtenu : pour tout intervalle I non trivial, p.s. B n'est pas monotone sur I . Il reste à justifier que p.s., pour tout intervalle I non trivial, B n'est pas monotone sur I . Si B était monotone sur un intervalle non trivial, alors B serait en particulier monotone sur n'importe quel intervalle à extrémités rationnelles inclus dans celui-ci. Ainsi,

$$\begin{aligned} & \mathbb{P}(B \text{ est monotone sur un intervalle non trivial}) \\ & \leq \mathbb{P}(\exists a, b \in \mathbf{Q}, 0 < a < b, \text{ tels que } B \text{ est monotone sur } [a, b]) \\ & \leq \sum_{a, b \in \mathbf{Q}, 0 < a < b} \mathbb{P}(B \text{ est monotone sur } [a, b]) = 0, \end{aligned}$$

en utilisant la dénombrabilité de \mathbf{Q} et la sous-additivité de la mesure \mathbb{P} . □

Le résultat suivant est plus fort :

Proposition 5.7. *Presque sûrement, B n'est nulle part dérivable.*

Remarque 5.8. *Remarquons que la fonction $x \mapsto x^2 \sin(\frac{1}{x})$ est continue, dérivable en 0, mais n'est monotone sur aucun voisinage de 0. Il existe plus généralement des fonctions monotones sur aucun intervalle qui sont dérivables en tout point (mais ce n'est pas facile à construire).*

La non-dérivabilité de B en un temps t est simple à obtenir. La dérivabilité en t équivaut à l'existence d'une limite de $\frac{B_{t+h} - B_t}{h}$ quand $h \rightarrow 0^+$. Or $\frac{B_{t+h} - B_t}{h}$ a pour loi $\mathcal{N}(0, \frac{1}{h})$, qui est la loi de $\frac{1}{\sqrt{h}}Z$ où $Z \sim \mathcal{N}(0, 1)$, donc la probabilité que cette variable soit dans un intervalle $[-A, A]$ est $\mathbb{P}(|Z| < A\sqrt{h})$ et tend donc vers 0 quand $h \rightarrow 0^+$. On en déduit qu'il n'y a presque sûrement pas convergence en t . Par intersection dénombrables d'événements presque sûrs, B n'est presque sûrement pas dérivable en tout point rationnel, par exemple.

Preuve : Si B est dérivable en un point de $[0, 1]$, alors il existe $C \in \mathbb{N}$ et $n \in \mathbb{N}^*$ tel que B est C -lipschitzienne sur un ensemble $[x - \frac{3}{n}, x + \frac{3}{n}]$ où $x \in [0, 1]$. Il suffit donc de montrer que, pour tout $C > 0$ et $n \in \mathbb{N}$, ceci ne se produit presque sûrement pas. On note A_n cet événement.

Sur cet événement, il existe un entier k tel que $\frac{k-1}{n}, \dots, \frac{k+2}{n}$ sont tous dans $[0, 1]$ et à distance $\leq \frac{3}{n}$ du point x de la notation précédente, et on a alors $|\frac{k}{n} - s| + |\frac{k-1}{n} - s| \leq \frac{5}{n}$ (penser au cas où s est proche de 0 pour comprendre le 5) d'où $|B_{\frac{k}{n}} - B_{\frac{k-1}{n}}| \leq \frac{5C}{n}$ par inégalité triangulaire avec B_s et propriété lipschitzienne au voisinage de x ; de même pour les accroissements $|B_{\frac{k+1}{n}} - B_{\frac{k}{n}}|$ et $|B_{\frac{k+2}{n}} - B_{\frac{k+1}{n}}|$. Ainsi, $A_n \subset B_n$, où

$$B_n = \{ \exists 1 \leq k \leq n-2 \text{ tel que } \max(|B_{\frac{k}{n}} - B_{\frac{k-1}{n}}|, |B_{\frac{k+1}{n}} - B_{\frac{k}{n}}|, |B_{\frac{k+2}{n}} - B_{\frac{k+1}{n}}|) \leq \frac{5C}{n} \}.$$

Or

$$\begin{aligned} \mathbb{P}(B_n) &\leq \sum_{k=1}^{n-2} \mathbb{P}\left(\max(|B_{\frac{k}{n}} - B_{\frac{k-1}{n}}|, |B_{\frac{k+1}{n}} - B_{\frac{k}{n}}|, |B_{\frac{k+2}{n}} - B_{\frac{k+1}{n}}|) \leq \frac{5C}{n}\right) \\ &\leq n \mathbb{P}\left(|B_{1/n} - B_0| \leq \frac{5C}{n}\right)^3 \\ &= n \mathbb{P}\left(|B_1| \leq \frac{5C}{\sqrt{n}}\right)^3 \leq n \left(\frac{10C}{\sqrt{2\pi}\sqrt{n}}\right)^3 = \frac{K}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

en utilisant le fait que $B_{1/n} \sim \mathcal{N}(0, 1/n)$ a même loi que $\frac{1}{\sqrt{n}}B_1$, et l'inégalité

$$\mathbb{P}(|B_1| \leq a) = \int_{-a}^a \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \leq \int_{-a}^a \frac{1}{\sqrt{2\pi}} dx = \frac{2a}{\sqrt{2\pi}}.$$

Ainsi, $\mathbb{P}(A_n) \rightarrow 0$, or on constate que la suite A_n est croissante, donc ceci montre que $\mathbb{P}(A_n) = 0$ pour tout n . C'est ce qu'il nous fallait démontrer. \square

On peut en revanche prouver que B est localement Hölderien d'indice α , pour tout $\alpha < \frac{1}{2}$: presque sûrement, il existe $C > 0$ tel que, pour tous $0 < s < t < 1$,

$$|B_t - B_s| \leq C|t - s|^\alpha.$$

Cela pourrait se montrer à partir de notre construction précédente.

5.2.2 Invariances

Proposition 5.9. Soit B un mouvement brownien. Soit $\sigma > 0$, $s > 0$. La loi de B satisfait les invariances suivantes :

- (symétrie axiale) $-B$ est un mouvement brownien.
- (changement d'échelle) $\left(\frac{1}{\sqrt{\sigma}}B_{\sigma t}\right)_{t \geq 0}$ est un mouvement brownien.
- (propriété de Markov au temps s) $(B_{s+t} - B_s)_{t \geq 0}$ est un mouvement brownien, indépendant de $\mathcal{F}_s = \sigma((B_u)_{u \leq s})$.
- (inversion du temps) $\left(tB_{1/t}\right)_{t > 0}$, prolongé par la valeur 0 en 0, est un mouvement brownien.
- (retournement du temps) $(B_{s-t} - B_s)_{t \in [0, s]}$ est un mouvement brownien (sur $[0, s]$).

Preuve : Il faut vérifier les points (i) à (iv) dans chaque cas. Le point (i) est toujours évident. Vérifier (ii) et (iii) revient à vérifier que le processus défini \tilde{B} est gaussien et a pour covariance $\text{Cov}(\tilde{B}_s, \tilde{B}_t) = s \wedge t$. Ce calcul est laissé en exercice. Le point (iv) est évident sauf pour d) en $t = 0$; on peut remarquer que l'existence d'une limite nulle en 0 s'exprime à l'aide des valeurs en $t > 0$ (et même $t \in \mathbf{Q}_+$ par continuité sur $]0, \infty[$), et utiliser l'identité en loi vérifiée sur ces valeurs avec le fait qu'il existe un mouvement brownien (donc continu en 0). \square

Vu la continuité en 0 de B , le point d) donne alors $tB_{1/t} \rightarrow 0$ quand $t \rightarrow 0^+$ donc, pour $u = 1/t$:

Corollaire 5.10 (Loi des grands nombres pour B). Presque sûrement, $\frac{B_u}{u} \xrightarrow{u \rightarrow \infty} 0$.

5.3 Propriétés de Markov et de martingale, et conséquences

5.3.1 Propriété de Markov

Pour tout $x \in \mathbb{R}$, notons \mathbb{P}_x la loi du mouvement brownien issu de x , c'est-à-dire simplement du processus $(x + B_t)_{t \geq 0}$.

On a déjà énoncé : pour tout $s > 0$, $(B_{s+t} - B_s)_{t \geq 0}$ est un mouvement brownien, indépendant de $\mathcal{F}_s = \sigma((B_u)_{u \leq s})$. Ainsi, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée (ou positive),

$$\mathbb{E}_0(f(B_{s+t}) | \mathcal{F}_s) = \mathbb{E}_0(f(B_s + (B_{s+t} - B_s)) | \mathcal{F}_s) = g(B_s),$$

en utilisant la proposition 2.10, où $g(x) = \mathbb{E}_0(f(x + B_t)) = \mathbb{E}_x(f(B_t))$. Autrement écrit,

$$\mathbb{E}_0(f(B_{s+t}) | \mathcal{F}_s) = \mathbb{E}_0(f(B_s + (B_{s+t} - B_s)) | \mathcal{F}_s) = \mathbb{E}_{B_s}(f(B_t)) = P_t f(B_s),$$

où

$$P_t f(x) = \int p_t(x, y) f(y) dy,$$

avec

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \quad (\text{densité de } \mathcal{N}(x, t) \text{ en } y).$$

Cette écriture est à rapprocher des formules suivantes que l'on connaît pour les chaînes de Markov :

$$\mathbb{E}(f(X_{k+n}) | \mathcal{F}_k) = \mathbb{E}_{X_k}(f(X_n)) = P^n f(X_k) \quad \text{avec} \quad P^n f(x) = \sum_y P^n(x, y) f(y).$$

L'analogie de la famille $(P^n)_{n \geq 0}$ de puissances (ou d'itérés) d'une matrice (infinie) est une famille $(P_t)_{t \geq 0}$ d'opérateurs s'appliquant aux fonctions mesurables bornées, qui vérifie aussi une relation de semi-groupe :

$$P_t P_s f(x) = P_{s+t} f(x).$$

Cela vient de la propriété de Markov : en intégrant la relation plus haut,

$$P_{s+t} f(0) = \mathbb{E}_0(f(B_{s+t})) = \mathbb{E}(P_t f(B_s)) = P_s(P_t f)(0),$$

et cela vaut pour tout x en translatant.

On peut en fait montrer (sous certaines conditions) que $(P_s)_{s \geq 0}$ peut se voir comme la famille des "puissances" d'un opérateur : formellement, $P_t = e^{tD}$ pour un opérateur D , appelé générateur infinitésimal. L'opérateur D peut s'obtenir en dérivant en 0 :

$$Df(x) = \lim_{t \rightarrow 0^+} \frac{P_t f(x) - f(x)}{t} = \lim_{t \rightarrow 0^+} \frac{1}{t} (\mathbb{E}_x[f(X_t)] - f(x)),$$

c'est-à-dire que

$$\mathbb{E}_x[f(X_t)] = f(x) + tDf(x) + o_{t \rightarrow 0^+}(t).$$

Pour le mouvement brownien, avec la définition de P_t ci-dessus, on obtient que $D = \frac{1}{2}\Delta$, où Δ est l'opérateur laplacien ($= \frac{d^2}{dx^2}$ en dimension 1). On voit notamment que cette relation n'aura de sens qu'appliquée à des fonctions f assez régulières : l'opérateur D n'est pas défini sur autant de fonctions que P_t . Cette apparition du laplacien peut être mise en parallèle avec la marche aléatoire simple symétrique sur \mathbb{Z} : dans ce cas,

$$\mathbb{E}_x[f(X_1)] = \frac{1}{2}f(x-1) + \frac{1}{2}f(x+1) = f(x) + \frac{f(x+1) + f(x-1) - 2f(x)}{2},$$

ce qui fait apparaître une dérivée seconde discrète.

Plus généralement, on peut définir des **processus de Markov**, donnés par un semi-groupe $(P_t)_{t \geq 0}$ ou par un générateur D opérant sur un sous-domaine \mathcal{D} des fonctions mesurables bornées.

5.4 Propriété de martingale

Pour tout $t \geq 0$, on définit $\mathcal{F}_t = \sigma((B_u)_{u \in [0, t]})$. Alors la famille croissante $(\mathcal{F}_t)_{t \geq 0}$ est appelée une **filtration**.

On dira qu'un processus $(X_t)_{t \geq 0}$ est une **martingale** par rapport à $(\mathcal{F}_t)_{t \geq 0}$ si :

- (i) pour tout $t \geq 0$, X_t est \mathcal{F}_t -mesurable (X est adapté) ;
- (ii) pour tout $t \geq 0$, X_t est intégrable ;
- (iii) pour tous $0 \leq s \leq t$, $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$ p.s..

On a les exemples suivants :

Proposition 5.11. *Pour la filtration $(\mathcal{F}_t)_{t \geq 0}$ définie par $\mathcal{F}_t = \sigma((B_u)_{0 \leq u \leq t})$ pour tout t ,*

- a) $(B_t)_{t \geq 0}$ est une martingale ;
- b) $(B_t^2 - t)_{t \geq 0}$ est une martingale ;
- c) pour tout réel σ , $(\exp(\sigma B_t - \frac{\sigma^2}{2}t))_{t \geq 0}$ est une martingale.

Preuve : Les preuves sont similaires au cas discret. On traite a) par exemple : (i) est vérifié par définition de la filtration, (ii) est vérifié par intégrabilité des lois gaussiennes, et (iii) vient de la propriété de Markov pour B : pour tous $0 \leq s < t$,

$$\begin{aligned} \mathbb{E}(B_t | \mathcal{F}_s) &= \mathbb{E}(B_s + (B_t - B_s) | \mathcal{F}_s) = \mathbb{E}(B_s | \mathcal{F}_s) + \mathbb{E}(B_t - B_s | \mathcal{F}_s) \\ &= B_s + \mathbb{E}(B_t - B_s) = B_s, \end{aligned}$$

ce qui conclut la preuve que B est une martingale. □

De nombreuses propriétés des martingales s'étendent au cas continu. Par exemple :

Proposition 5.12 (Théorème de Doob). *Soit $(X_t)_{t \geq 0}$ une martingale telle que $\sup_{t \geq 0} \mathbb{E}[(X_t)_+] < \infty$. Alors presque sûrement $X_t \xrightarrow[t \rightarrow +\infty]{} X_\infty$ où $X_\infty \in L^1$.*

Preuve : On peut définir, pour tous $a < b$, le nombre de franchissements croissants de $[a, b]$ par $(X_t)_{t \geq 0}$, et observer que c'est la limite croissante du nombre de franchissements par $(X_t)_{t \in D_n}$ (où $D_n = 2^{-n}\mathbb{N}$), qui se majore par le lemme 3.30 vu dans le cas discret. Comme la majoration est uniforme, on peut conclure comme dans le cas discret. □

Notons de plus que le même résultat vaut pour des limites $t \rightarrow (t_0)^-$, ou $t \rightarrow (t_0)^+$ en tout point : ceci montre que l'hypothèse de martingale (et l'hypothèse de Doob, au voisinage de t_0) assure l'existence de limites à gauche et à droite.

On a aussi :

Proposition 5.13 (Inégalité de Doob dans L^2). *Si X est une martingale continue, alors pour tout $t > 0$,*

$$\mathbb{E}[(\sup_{s \leq t} X_s)^2] \leq 4\mathbb{E}[X_t^2].$$

Preuve : Soit $t > 0$. Notons, pour tout n , $D_n = \frac{t}{2^n}\mathbb{N}$. On constate que, pour tout n , la suite $(X_s)_{s \in D_n}$ est une martingale discrète (elle est extraite de la martingale continue X). En particulier, en vertu de l'inégalité de Doob dans L^2 ,

$$\mathbb{E} \left[\max_{\substack{0 \leq s \leq t, \\ s \in D_n}} X_s^2 \right] \leq 4\mathbb{E}[X_t^2].$$

La proposition s'en déduit par théorème de convergence monotone quand $n \rightarrow \infty$ (la continuité de X assure la convergence du maximum). □

Proposition 5.14. *Si $(X_t)_{t \geq 0}$ est une martingale continue de carré intégrable, et T est un temps d'arrêt, alors $X^T = (X_{t \wedge T})_{t \geq 0}$ est une martingale.*

Ici, un **temps d'arrêt** est une v.a. T à valeurs dans $\mathbb{R}_+ \cup +\infty$ telle que, pour tout $t \geq 0$, $\{T \leq t\} \in \mathcal{F}_t$.

Preuve : Notons, pour tout n , $\mathcal{D}_n = \frac{1}{2^n}\mathbb{N}$.

En notant, pour tout n , $T_n = \min\{t \in \mathcal{D}_n \mid t > T_n\}$, la suite $(T_n)_{n \geq 0}$ décroît vers T .

Tout d'abord, $X_{t \wedge T}$ est \mathcal{F}_t -mesurable et intégrable. Pour la mesurabilité, on peut noter que $X_{t \wedge T} = \lim_n X_{t \wedge T_n}$, et $X_{t \wedge T_n} = X_t \mathbf{1}_{\{t < T_n\}} + X_{T_n} \mathbf{1}_{\{T_n \leq t\}}$, et tous les termes sont \mathcal{F}_t -mesurables (pour $X_{T_n} \mathbf{1}_{\{T_n \leq t\}}$, en le décomposant comme une somme finie sur les valeurs possibles de $T_n \leq t$). Pour l'intégrabilité, on peut noter que $|X_{t \wedge T}| \leq \sup_{s \leq t} |X_s|$ et utiliser la proposition précédente.

Soit $0 \leq s < t$. On définit de même des suites $(s_n)_n$ et $(t_n)_n$ telles que, pour tout n , $0 \leq s \leq s_n \leq t \leq t_n$, $s_n, t_n \in \mathcal{D}_n$, et qui convergent vers s et t en décroissant. Pour tout n , par la propriété analogue pour la martingale discrète $(X_t)_{t \geq 0, t \in \mathcal{D}_n}$ arrêtée au temps T_n , on a

$$\mathbb{E}(X_{t_n \wedge T_n} \mid \mathcal{F}_{s_n}) = X_{s_n \wedge T_n}.$$

Alors, pour tout $A \in \mathcal{F}_s$ on a, pour tout n , $A \in \mathcal{F}_{s_n}$ et donc

$$\mathbb{E}(X_{t_n \wedge T_n} \mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X_{t_n \wedge T_n} \mid \mathcal{F}_{s_n}) \mathbf{1}_A) = \mathbb{E}(X_{s_n \wedge T_n} \mathbf{1}_A)$$

d'où à la limite (par théorème de convergence dominée, en dominant par $\sup_{u \leq t} |X_u|$ (intégrable d'après la proposition précédente)),

$$\mathbb{E}(X_{t \wedge T} \mathbf{1}_A) = \mathbb{E}(X_{s \wedge T} \mathbf{1}_A),$$

ce qui prouve la propriété de martingale : $\mathbb{E}(X_{t \wedge T} \mid \mathcal{F}_s) = X_{s \wedge T}$. \square

5.5 Applications de la propriété de martingale

Soit $a, b > 0$. On définit le temps d'arrêt

$$T = \inf\{t \geq 0 \mid B_t \notin [-a, b]\}.$$

Le temps T est fini p.s. En effet, la martingale arrêtée B^T converge p.s. par le théorème de Doob, ce qui n'est possible que si $T < \infty$ (en effet, les accroissements avant T sont gaussiens et donc p.s. non nuls). En particulier on observe que B est non borné : p.s., $\sup_{t > 0} |B_t| = \infty$.

Comme $(B_{t \wedge T})_{t \geq 0}$ est une martingale, on a pour tout $t \geq 0$,

$$\mathbb{E}(B_{t \wedge T}) = \mathbb{E}(B_{0 \wedge T}) = 0.$$

De plus $B_{t \wedge T} \xrightarrow[t \rightarrow \infty]{} B_T$, et $|B_{t \wedge T}| \leq \max(a, b)$, donc on peut appliquer le théorème de convergence dominée pour obtenir :

$$\mathbb{E}(B_T) = 0.$$

Or B_T ne prend que $-a$ et b pour valeurs, donc

$$\mathbb{E}(B_T) = -a\mathbb{P}(B_T = -a) + b\mathbb{P}(B_T = b).$$

En comparant, vu que $\mathbb{P}(B_T = b) = 1 - \mathbb{P}(B_T = -a)$, il vient

$$\mathbb{P}(B_T = -a) = \frac{b}{a+b}.$$

Comme $(B_{t \wedge T}^2 - t \wedge T)_{t \geq 0}$ est une martingale, on a pour tout $t \geq 0$

$$\mathbb{E}(B_{t \wedge T}^2 - t \wedge T) = 0,$$

d'où

$$\mathbb{E}(B_{t \wedge T}^2) = \mathbb{E}(t \wedge T).$$

Or le terme de gauche converge vers $\mathbb{E}(B_T^2)$ par convergence dominée (on a $|B_{t \wedge T}| \leq \max(a, b)$) et le terme de droite converge vers $\mathbb{E}(T)$ par convergence monotone. Il en résulte

$$\mathbb{E}(B_T^2) = E(T).$$

Or le terme de gauche se calcule grâce au résultat précédent :

$$\mathbb{E}(B_T^2) = a^2 \mathbb{P}(B_T = a) + b^2(1 - \mathbb{P}(B_T = a))$$

d'où

$$\mathbb{E}(T) = \frac{a^2 b}{a + b} + \frac{ab^2}{a + b} = ab.$$

Soit $a > 0$. Considérons le temps d'arrêt

$$T_a = \inf\{t \geq 0 \mid B_t = a\}.$$

On a montré que p.s. $T_a < \infty$ ou $T_{-a} < \infty$, mais on n'a pas prouvé que p.s. $T_a < \infty$.

Soit $\sigma > 0$. Comme $X_t = e^{\sigma B_t - \frac{\sigma^2}{2}t}$ définit une martingale, $(X_{t \wedge T_a})_{t \geq 0}$ est aussi une martingale. En particulier on en déduit pour tout t ,

$$\mathbb{E}(e^{\sigma B_{t \wedge T_a} - \frac{\sigma^2}{2}(t \wedge T_a)}) = \mathbb{E}(X_0) = 1.$$

Si $T_a = \infty$, alors $\sigma B_t - \frac{\sigma^2}{2}t \leq a - \frac{\sigma^2}{2}t \rightarrow -\infty$ quand $t \rightarrow +\infty$; et si $T_a < \infty$ alors $X_{T_a} = e^{\sigma a - \frac{\sigma^2}{2}T_a}$, donc dans tous les cas :

$$X_{t \wedge T_a} \xrightarrow[t \rightarrow \infty]{} e^{\sigma a - \frac{\sigma^2}{2}T_a} \mathbf{1}_{\{T_a < \infty\}}$$

Pour tout t , $B_{t \wedge T_a} \leq a$, donc vu que $\sigma > 0$ on a

$$|X_{t \wedge T_a}| \leq e^{\sigma a},$$

ce qui permet d'appliquer le théorème de convergence dominée à $X_{t \wedge T_a}$:

$$1 = \mathbb{E}(X_{t \wedge T_a}) \xrightarrow[t \rightarrow \infty]{} \mathbb{E}(e^{\sigma a - \frac{\sigma^2}{2}T_a} \mathbf{1}_{\{T_a < \infty\}}),$$

ce qui donne

$$\mathbb{E}(e^{-\frac{\sigma^2}{2}T_a} \mathbf{1}_{\{T_a < \infty\}}) = e^{-\sigma a}.$$

Pour $\sigma \rightarrow 0$, on en déduit en particulier par convergence monotone

$$\mathbb{P}(T_a < \infty) = 1.$$

Ceci montre que le mouvement brownien est récurrent : p.s., il visite tous les réels :

$$\liminf_{t \rightarrow \infty} B_t = -\infty, \quad \limsup_{t \rightarrow \infty} B_t = +\infty.$$

De plus on obtient la transformée de Laplace de T_a , qui caractérise sa loi :

$$\mathbb{E}(e^{-\frac{\sigma^2}{2}T_a}) = e^{-\sigma a}.$$

Bibliographie

- [1] P. Baldi, L. Mazliak et P. Priouret. *Martingales et chaînes de markov*, éditions Hermann.
- [2] Ph. Barbe, M. Ledoux. *Probabilités*, éditions EDP Sciences