

Processus aléatoires à temps discret

Erick Herbin

26 janvier 2024

Table des matières

0.1	Espérance conditionnelle sachant une sous-tribu	6
0.2	Définition de l'espérance conditionnelle dans L^2	7
0.3	Définition de l'espérance conditionnelle dans L^1	8
0.4	Espérance conditionnelle par rapport à des v. a.	12
I	Martingales (4 séances)	15
1	Quelques généralités sur les processus aléatoires	16
1.1	Rappels de théorie de la mesure : lemme des classes monotones	16
1.2	Définitions et propriétés	19
1.3	Temps d'arrêt	20
1.4	Loi d'un processus	22
1.4.1	Tribu de Kolmogorov	23
1.4.2	Théorème de Kolmogorov	24
2	Martingales	27
2.1	Définitions et exemples	27
2.2	Décomposition de Doob	29
2.3	Théorème d'arrêt	30
2.3.1	Exemple : stratégie de jeu	32
2.4	Inégalités de martingales	34
3	Convergence d'une martingale	40
3.1	Convergence presque sûre d'une martingale	40
3.2	Martingales bornées dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ ($p > 1$)	43
3.3	Martingales uniformément intégrables	44

II	Chaînes de Markov à temps discret (3 séances)	50
4	Chaînes de Markov	51
4.1	Matrices de transition et chaînes de Markov	51
4.2	Loi d'une chaîne de Markov et probabilités de transition	55
4.3	La chaîne de Markov canonique	60
4.3.1	Définition	60
4.3.2	Propriété de Markov	61
5	Classification des états des chaînes de Markov	64
5.1	Opérateur potentiel	64
5.2	Communication entre les états d'une chaîne de Markov	65
5.2.1	Réurrence et transience	65
5.2.2	Relation d'équivalence $x \rightsquigarrow y$	67
5.3	Mesures invariantes	71
6	Comportement asymptotique	77
6.1	Cas récurrent irréductible	77
6.2	Cas apériodique	82
III	Quelques processus à temps continu (2 séances)	84
7	Processus gaussiens	85
7.1	Définition et existence	85
7.2	Bruit blanc de \mathbb{R}	87
7.3	Processus isonormal de \mathbb{R}	89
8	Introduction au mouvement brownien	90
8.1	Définition comme processus gaussien	90
8.2	Quelques problématiques de changement d'échelles	91
8.2.1	Convergence faible	91
8.2.2	Principe d'invariance de Donsker	93
8.3	Trajectoire du mouvement brownien	93
8.3.1	Existence d'une modification continue	94

8.4	Premières propriétés du mouvement brownien	96
-----	------------------------------------------------------	----

Introduction

L'accroissement des capacités de simulation numérique a fait émerger de nouvelles problématiques dans le spectre de l'ingénieur. Après 30 ans de développement en modélisation numérique, les probabilités fournissent un cadre adapté à la prise en compte des incertitudes inhérentes aux erreurs de modélisation et à la variabilité de l'environnement.

Sur la base de la théorie moderne des probabilités, les processus aléatoires permettent de modéliser des phénomènes fluctuants ou stochastiques dont l'évolution dans le temps revêt une grande importance. Par exemple, l'ajout de termes aléatoires dans les EDP régissant les phénomènes physiques, tels que la turbulence, conduit à des solutions qui sont des processus stochastiques. Dans ce cadre, les dépendances avec le passé peuvent être décrites par des fonctions de corrélation (processus gaussiens), par leurs espérances conditionnelles (martingales) ou par leurs lois conditionnelles (processus de Markov).

Pour maîtriser ces concepts modernes, il est nécessaire de bien comprendre le cas où les processus sont indexés par des éléments d'ensembles discrets avant d'aborder les outils complexes issus du calcul stochastique intervenant en conception robuste, modélisation physique et financière ou traitement du signal et d'images.

Avertissement : Ce document n'a pas pour objectif de constituer un support de cours complet. Il s'apparente à un plan détaillé contenant les renvois précis à des références accessibles sur Internet.

La première partie de ce cours s'appuie largement sur les références [6, 13, 11].

Rappels d'espérance conditionnelle

Référence : Lecture Notes du cours CIP de 1ère année ([5]).

Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, si A et B sont dans \mathcal{F} et $\mathbf{P}(B) > 0$, la probabilité conditionnelle de A sachant B est définie par

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

Si X et Y sont des variables aléatoires à valeurs respectivement dans \mathbb{R} et dans un espace discret E . Il est intéressant de connaître la distribution ou l'espérance de X , connaissant la réalisation de Y . Pour tout $y \in E$, on a $\{Y = y\} \in \mathcal{F}$ donc si $\mathbf{P}(Y = y) > 0$, alors on peut considérer la mesure de probabilité

$$\begin{aligned} \mathbf{Q} : \mathcal{F} &\rightarrow [0, 1] \\ A &\mapsto \mathbf{P}(A | Y = y). \end{aligned}$$

Si $X \in L^1(\Omega, \mathcal{F}, \mathbf{Q})$, l'espérance conditionnelle de X sachant $\{Y = y\}$ est définie par

$$\mathbf{E}[X | Y = y] = \mathbf{E}_{\mathbf{Q}}[X],$$

où $\mathbf{E}_{\mathbf{Q}}[X]$ désigne l'espérance de X sous la probabilité \mathbf{Q} .

Par exemple, dans le cas où X est également à valeurs dans un espace discret \tilde{E} ,

$$\mathbf{E}[X | Y = y] = \sum_{x \in \tilde{E}} x \mathbf{P}(X = x | Y = y),$$

si la somme est convergente.

Pour définir l'espérance de X sachant la variable aléatoire Y , on considère

$$\psi : y \mapsto \begin{cases} \mathbf{E}[X | Y = y] & \text{si } \mathbf{P}(Y = y) > 0 \\ 0 & \text{sinon.} \end{cases}$$

L'espérance de X sachant Y est alors définie par

$$\mathbf{E}[X | Y] = \psi(Y).$$

La construction moderne décrite dans ce chapitre permet d'étendre la définition de l'espérance conditionnelle de X sachant une variable aléatoire Y qui n'est pas nécessairement discrète, et pour laquelle les événements $\{Y = y\}$ sont de probabilité nulle.

0.1 Espérance conditionnelle sachant une sous-tribu

De manière plus générale que ci-dessus, on cherche à définir l'espérance conditionnelle par rapport à une sous-tribu \mathcal{G} de \mathcal{F} .

Définition 0.1.1

Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, soit X une variable aléatoire réelle dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F} . On appelle espérance conditionnelle de X sachant \mathcal{G} , la variable aléatoire réelle Y vérifiant les deux conditions suivantes :

- (i) Y est dans $L^1(\Omega, \mathcal{G}, \mathbf{P})$;
- (ii) Pour tout $A \in \mathcal{G}$,

$$\int_A X.d\mathbf{P} = \int_A Y.d\mathbf{P}. \quad (1)$$

L'espérance conditionnelle est unique presque sûrement et notée $\mathbf{E}[X | \mathcal{G}]$.

On définit alors la probabilité conditionnelle $\mathbf{P}(A | \mathcal{G}) = \mathbf{E}[\mathbb{1}_A | \mathcal{G}]$.

Notons qu'en écrivant l'égalité (1) sous la forme $\mathbf{E}[X.\mathbb{1}_A] = \mathbf{E}[Y.\mathbb{1}_A]$, le point (ii) est équivalent à

(ii') Pour toute v. a. Z \mathcal{G} -mesurable bornée, $\mathbf{E}[XZ] = \mathbf{E}[YZ]$.

Preuve Montrons l'unicité (presque sure) de l'espérance conditionnelle.

Soit Y et Y' sont deux v. a. réelles dans $L^1(\Omega, \mathcal{G}, \mathbf{P})$ vérifiant (1). On fixe $\epsilon > 0$ et on considère

$$A = \{\omega : Y(\omega) - Y'(\omega) \geq \epsilon\} \in \mathcal{G}.$$

On a alors

$$0 = \mathbf{E}[Y.\mathbb{1}_A] - \mathbf{E}[Y'.\mathbb{1}_A] = \mathbf{E}[(Y - Y')\mathbb{1}_A] \geq \epsilon \mathbf{P}(A).$$

Comme par définition de la probabilité $\mathbf{P}(A) \geq 0$, on en déduit $\mathbf{P}(A) = 0$.

Pour tout $\epsilon > 0$, $\mathbf{P}\{\omega : Y(\omega) - Y'(\omega) \geq \epsilon\} = 0$. En prenant $\epsilon = \frac{1}{n}$, on obtient une suite croissante d'événements. En passant à la limite $n \rightarrow \infty$, on obtient

$$\mathbf{P}\{\omega : Y(\omega) \geq Y'(\omega)\} = 0,$$

c'est-à-dire $Y \leq Y'$ p.s.

Par symétrie, on a aussi $Y' \leq Y$ p.s. et donc $Y = Y'$ p.s.

L'existence de l'espérance conditionnelle est une conséquence du théorème de Radon-Nikodym, en théorie de la mesure. Dans ce cours, on montre l'existence dans l'espace L^2 qui est un espace de Hilbert, puis on passe dans L^1 par densité. \square

Exemple On étudie les deux exemples extrêmes :

- Si toutes les informations sur X sont contenues dans \mathcal{G} , i.e. si $X \in L^1(\Omega, \mathcal{G}, \mathbf{P})$, alors on a $\mathbf{E}[X | \mathcal{G}] = X$ p.s.
- Si la connaissance de \mathcal{G} n'apporte aucune information sur X , i.e. si X et \mathcal{G} sont indépendantes, alors on a $\mathbf{E}[X | \mathcal{G}] = \mathbf{E}[X]$ p.s.

0.2 Définition de l'espérance conditionnelle dans L^2

Rappelons que l'espace $L^2(\Omega, \mathcal{F}, \mathbf{P})$ muni du produit scalaire

$$\begin{aligned} L^2(\Omega, \mathcal{F}, \mathbf{P}) \times L^2(\Omega, \mathcal{F}, \mathbf{P}) &\rightarrow \mathbb{R} \\ (X, Y) &\mapsto \mathbf{E}[XY] \end{aligned}$$

est un espace de Hilbert (cf. cours d'Analyse).

Si \mathcal{G} est une sous-tribu de \mathcal{F} , alors $L^2(\Omega, \mathcal{G}, \mathbf{P})$ est un sous-espace fermé de $L^2(\Omega, \mathcal{F}, \mathbf{P})$ (pourquoi?).

Toute v.a. X de $L^2(\Omega, \mathcal{F}, \mathbf{P})$ admet donc une projection orthogonale \hat{X} sur $L^2(\Omega, \mathcal{G}, \mathbf{P})$ définie par

$$\begin{aligned} \forall Z \in L^2(\Omega, \mathcal{G}, \mathbf{P}); \quad \mathbf{E}[(X - \hat{X})Z] &= 0 \\ \Leftrightarrow \mathbf{E}[XZ] &= \mathbf{E}[\hat{X}Z]. \end{aligned}$$

Dans le cas particulier où $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, l'espérance conditionnelle $\mathbf{E}[X | \mathcal{G}]$ est donc la projection orthogonale de X sur le sous-espace $L^2(\Omega, \mathcal{G}, \mathbf{P})$. L'existence et l'unicité découle donc des propriétés de la projection orthogonale.

Définition 0.2.1

Soit X une variable aléatoire réelle dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F} . L'espérance conditionnelle de X sachant \mathcal{G} est définie comme la variable aléatoire $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ telle que

$$\forall Z \in L^2(\Omega, \mathcal{G}, \mathbf{P}); \quad \mathbf{E}[XZ] = \mathbf{E}[YZ].$$

L'espérance conditionnelle est unique (modulo égalité dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$) et notée $\mathbf{E}[X | \mathcal{G}]$.

$\mathbf{E}[X | \mathcal{G}]$ peut être vu comme la v. a. de $L^2(\Omega, \mathcal{G}, \mathbf{P})$ qui minimise l'erreur quadratique $\mathbf{E}[(X - Z)^2]$, $Z \in L^2(\Omega, \mathcal{G}, \mathbf{P})$.

Proposition 0.2.2

Soit X une variable aléatoire réelle dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F}

1. L'application $X \mapsto \mathbf{E}[X | \mathcal{G}]$ est linéaire dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$.
2. Si $X \geq 0$ p.s. alors $\mathbf{E}[X | \mathcal{G}] \geq 0$ p.s.
3. $\mathbf{E}(\mathbf{E}[X | \mathcal{G}]) = \mathbf{E}[X]$.

Preuve

1. Pour toutes v.a. réelles X, Y dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$, $\lambda \in \mathbb{R}$ et $Z \in L^2(\Omega, \mathcal{G}, \mathbf{P})$, on a

$$\begin{aligned}\mathbf{E}[(X + \lambda Y)Z] &= \mathbf{E}[XZ] + \lambda \mathbf{E}[YZ] \\ &= \mathbf{E}(\mathbf{E}[X | \mathcal{G}]Z) + \lambda \mathbf{E}(\mathbf{E}[Y | \mathcal{G}]Z) \\ &= \mathbf{E}[(\mathbf{E}[X | \mathcal{G}] + \lambda \mathbf{E}[Y | \mathcal{G}])Z],\end{aligned}$$

d'où $\mathbf{E}[X + \lambda Y | \mathcal{G}] = \mathbf{E}[X | \mathcal{G}] + \lambda \mathbf{E}[Y | \mathcal{G}]$.

2. On applique la définition avec $Z = \mathbb{1}_{\{\mathbf{E}[X | \mathcal{G}] < 0\}}$ et $X \geq 0$ p.s. On a alors $\mathbf{E}[XZ] \geq 0$, comme produit de v.a. positives p.s. D'autre part, si $\mathbf{P}(\mathbf{E}[X | \mathcal{G}] < 0) > 0$, alors on a

$$\mathbf{E}(\mathbf{E}[X | \mathcal{G}]Z) < 0.$$

On conclut donc $\mathbf{P}(\mathbf{E}[X | \mathcal{G}] < 0) = 0$.

3. On applique la définition pour $Z = 1$.

□

0.3 Définition de l'espérance conditionnelle dans L^1

Dans le cas où X est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ (c'est-à-dire dans le cas général de la définition 0.1.1), on n'a pas de structure d'espace de Hilbert permettant de définir l'espérance conditionnelle comme une projection orthogonale. On utilise alors un argument de densité pour passer d'une définition dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ à $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

1ère étape :

On considère le sous-ensemble $L^+(\Omega, \mathcal{F}, \mathbf{P})$ constitué des variables aléatoires qui sont positives presque sûrement.

Pour toutes variables aléatoires réelles X et Z dans $L^+(\Omega, \mathcal{F}, \mathbf{P})$ où Z est \mathcal{G} -mesurable, on considère les variables aléatoires arrêtées $X_n = X \wedge n$ et $Z_n = Z \wedge n$ pour $n \in \mathbb{N}$. Comme X_n et Z_n sont bornées par n , elles sont respectivement dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ et $L^2(\Omega, \mathcal{G}, \mathbf{P})$

Par le théorème de convergence monotone, on a pour tout n fixé,

$$\begin{aligned}\mathbf{E}[X_n Z] &= \lim_{m \rightarrow \infty} \mathbf{E}[X_n Z_m] \\ &= \lim_{m \rightarrow \infty} \mathbf{E}(\mathbf{E}[X_n | \mathcal{G}] Z_m) \\ &= \mathbf{E}(\mathbf{E}[X_n | \mathcal{G}] Z).\end{aligned}$$

On écrit ensuite, grâce au théorème de convergence monotone

$$\begin{aligned}\mathbf{E}[XZ] &= \lim_{n \rightarrow \infty} \mathbf{E}[X_n Z] \\ &= \lim_{n \rightarrow \infty} \mathbf{E}(\mathbf{E}[X_n | \mathcal{G}] Z).\end{aligned}$$

D'après la proposition 0.2.2, la suite des variables aléatoires $(\mathbf{E}[X_n | \mathcal{G}])_{n \in \mathbb{N}}$ est croissante. On peut donc appliquer le théorème de convergence monotone et définir

$$\mathbf{E}[X | \mathcal{G}] = \lim_{n \rightarrow \infty} \mathbf{E}[X_n | \mathcal{G}].$$

L'objet ainsi défini est \mathcal{G} -mesurable, comme la limite des $\mathbf{E}[X_n | \mathcal{G}]$ qui sont \mathcal{G} -mesurables et par convergence monotone,

$$\mathbf{E}[XZ] = \mathbf{E}(\mathbf{E}[X | \mathcal{G}] Z).$$

Notons que la définition de $\mathbf{E}[X | \mathcal{G}]$ pour $X \in L^+(\Omega, \mathcal{F}, \mathbf{P})$ et la Proposition 0.2.2 conduit à

$$\forall X \in L^+(\Omega, \mathcal{F}, \mathbf{P}); \quad \mathbf{E}[X | \mathcal{G}] \geq 0.$$

Dans le cas particulier où X est intégrable, *i.e.* $X \in L^+(\Omega, \mathcal{F}, \mathbf{P}) \cap L^1(\Omega, \mathcal{F}, \mathbf{P})$, le théorème de convergence dominée implique $\mathbf{E}[X | \mathcal{G}] \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ et $\mathbf{E}(\mathbf{E}[X | \mathcal{G}]) = \mathbf{E}[X]$.

2ème étape :

Pour passer de la définition dans $L^+(\Omega, \mathcal{F}, \mathbf{P})$ à $L^1(\Omega, \mathcal{F}, \mathbf{P})$, on décompose toute variable aléatoire $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ en

$$X = X^+ - X^-$$

où $X^+ = \max(X, 0)$ et $X^- = -\min(X, 0)$ sont des variables aléatoires dans $L^+(\Omega, \mathcal{F}, \mathbf{P})$.

Comme $|X| = X^+ + X^-$, on a $X^+, X^- \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Ainsi dans la construction précédente, les variables $\mathbf{E}[X^+ | \mathcal{G}]$ et $\mathbf{E}[X^- | \mathcal{G}]$ sont dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ donc finies presque sûrement.

On définit alors

$$\mathbf{E}[X | \mathcal{G}] = \mathbf{E}[X^+ | \mathcal{G}] - \mathbf{E}[X^- | \mathcal{G}].$$

A partir de la Proposition 0.2.2, la construction de l'espérance conditionnelle dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ implique les propriétés suivantes :

Proposition 0.3.1

Soit X une variable aléatoire réelle dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F}

1. L'application $X \mapsto \mathbf{E}[X | \mathcal{G}]$ est linéaire dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.
2. Si $X \geq 0$ p.s. alors $\mathbf{E}[X | \mathcal{G}] \geq 0$ p.s.
3. $\mathbf{E}(\mathbf{E}[X | \mathcal{G}]) = \mathbf{E}[X]$.

Le résultat suivant montre que l'espérance conditionnelle permet de caractériser les v.a. qui sont \mathcal{G} -mesurables.

Théorème 0.3.2

Soit X une variable aléatoire positive ou intégrable sur $(\Omega, \mathcal{F}, \mathbf{P})$. Soit \mathcal{G} une sous-tribu de \mathcal{F} . Alors, $\mathbf{E}[X | \mathcal{G}] = X$ p.s. si et seulement si X est \mathcal{G} -mesurable.

Preuve Par définition de l'espérance conditionnelle dans $L^+(\Omega, \mathcal{F}, \mathbf{P})$ ou dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, $\mathbf{E}[X | \mathcal{G}]$ est \mathcal{G} -mesurable.

Inversement, si X est \mathcal{G} -mesurable, alors X est dans $L^+(\Omega, \mathcal{G}, \mathbf{P})$ ou dans $L^1(\Omega, \mathcal{G}, \mathbf{P})$ (suivant le cas X positive ou X intégrable). Donc la caractérisation de $\mathbf{E}[X | \mathcal{G}]$ par

$$\forall Z \in L^+(\Omega, \mathcal{G}, \mathbf{P}) \text{ ou } L^1(\Omega, \mathcal{G}, \mathbf{P}); \quad \mathbf{E}[XZ] = \mathbf{E}(\mathbf{E}[X | \mathcal{G}] Z)$$

montre que $\mathbf{E}[X | \mathcal{G}] = X$ p.s. \square

Théorème 0.3.3

Soit X et Y des variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F} . On suppose que X est \mathcal{G} -mesurable.

Si les v.a. X , Y et XY sont intégrables (ou positives), alors on a

$$\mathbf{E}[XY | \mathcal{G}] = X \mathbf{E}[Y | \mathcal{G}] \quad \text{p.s.}$$

Preuve Dans un premier temps, on suppose les v.a. X et Y positives. Pour toute v.a. $Z \in L^+(\Omega, \mathcal{G}, \mathbf{P})$, on a

$$\mathbf{E}[XYZ] = \mathbf{E}(\mathbf{E}[Y | \mathcal{G}] XZ) \tag{2}$$

car le produit XZ est dans $L^+(\Omega, \mathcal{G}, \mathbf{P})$.

Comme le produit $X \mathbf{E}[Y | \mathcal{G}]$ est également dans $L^+(\Omega, \mathcal{G}, \mathbf{P})$, l'égalité (2) implique

$$\mathbf{E}[XY | \mathcal{G}] = X \mathbf{E}[Y | \mathcal{G}] \quad \text{p.s.}$$

On retourne au cas général en décomposant $X = X^+ - X^-$ et $Y = Y^+ - Y^-$, où X^+, X^-, Y^+, Y^- sont des variables aléatoires positives. On écrit alors

$$\begin{aligned} XY &= (X^+ - X^-)(Y^+ - Y^-) \\ &= X^+Y^+ - X^+Y^- - X^-Y^+ + X^-Y^- \end{aligned}$$

et on applique le résultat à chacun des 4 termes. \square

L'espérance conditionnelle possède les propriétés de convergence de l'intégrale par rapport à une mesure : théorème de convergence monotone, théorème de convergence dominée, lemme de Fatou.

Théorème 0.3.4

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles sur $(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F} .

- Convergence monotone : Si $X_n \geq 0$ pour tout n et X_n croît vers X p.s., alors

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_n | \mathcal{G}] = \mathbf{E}[X | \mathcal{G}] \quad p.s.$$

- Lemme de Fatou : Si $X_n \geq 0$ pour tout n , alors

$$\mathbf{E} \left[\liminf_{n \rightarrow \infty} X_n | \mathcal{G} \right] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[X_n | \mathcal{G}] \quad p.s.$$

- Convergence dominée : Si $\lim_{n \rightarrow \infty} X_n = X$ p.s. et s'il existe $Z \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ telle que $|X_n| \leq Z$ pour tout n , alors

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_n | \mathcal{G}] = \mathbf{E}[X | \mathcal{G}] \quad p.s.$$

Notons que ces résultats pour l'espérance sont des propriétés connues de l'intégrale par rapport à une mesure (cf. cours d'Analyse).

Preuve On utilise ces résultats connus pour l'espérance, dans la définition de l'espérance conditionnelle. \square

Théorème 0.3.5 (Inégalité de Jensen)

Soit X une variable aléatoire réelle de $(\Omega, \mathcal{F}, \mathbf{P})$ et soit \mathcal{G} une sous-tribu de \mathcal{F} .

Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ une application convexe. Si X et $\varphi(X)$ sont dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, alors

$$\varphi(\mathbf{E}[X | \mathcal{G}]) \leq \mathbf{E}[\varphi(X) | \mathcal{G}].$$

Preuve L'application φ étant convexe, elle admet une dérivée à droite φ'_d . On a

$$\forall x, y \in \mathbb{R}; \quad \varphi'_d(x)(y - x) \leq \varphi(y) - \varphi(x).$$

Soit $n \in \mathbb{N}^*$ et posons $Y = \mathbf{E}[X | \mathcal{G}]$. On a

$$\varphi'_d(Y \cdot \mathbb{1}_{\{|Y| \leq n\}})(X - Y \cdot \mathbb{1}_{\{|Y| \leq n\}}) \leq \varphi(X) - \varphi(Y \cdot \mathbb{1}_{\{|Y| \leq n\}}).$$

On prend l'espérance conditionnelle de cette inégalité par rapport à \mathcal{G} . En utilisant le fait que Y , $Y \cdot \mathbb{1}_{\{|Y| \leq n\}}$ et $\varphi'_d(Y \cdot \mathbb{1}_{\{|Y| \leq n\}})$ sont \mathcal{G} -mesurables, on obtient

$$\varphi'_d(Y \cdot \mathbb{1}_{\{|Y| \leq n\}})(Y - Y \cdot \mathbb{1}_{\{|Y| \leq n\}}) \leq \mathbf{E}[\varphi(X) | \mathcal{G}] - \varphi(Y \cdot \mathbb{1}_{\{|Y| \leq n\}}).$$

En faisant tendre n vers ∞ , on obtient

$$0 \leq \mathbf{E}[\varphi(X) \mid \mathcal{G}] - \varphi(Y). \quad \square$$

La dernière propriété de l'espérance conditionnelle, mentionnée dans cette section, concerne l'enchaînement des espérances conditionnelles.

Proposition 0.3.6

Soient \mathcal{G} et \mathcal{H} deux sous-tribus de \mathcal{F} telles que $\mathcal{G} \subset \mathcal{H}$. Alors, pour toute variable aléatoire X dans $L^+(\Omega, \mathcal{F}, \mathbf{P})$ ou $L^1(\Omega, \mathcal{F}, \mathbf{P})$, on a

$$\mathbf{E}(\mathbf{E}[X \mid \mathcal{H}] \mid \mathcal{G}) = \mathbf{E}[X \mid \mathcal{G}] \quad p.s.$$

Preuve On démontre le cas où $X \in L^+(\Omega, \mathcal{F}, \mathbf{P})$.

Soit Z une variable aléatoire \mathcal{G} -mesurable positive. Par définition de $\mathbf{E}[\bullet \mid \mathcal{G}]$, on a

$$\mathbf{E}[Z \mathbf{E}(\mathbf{E}[X \mid \mathcal{H}] \mid \mathcal{G})] = \mathbf{E}(Z \mathbf{E}[X \mid \mathcal{H}]).$$

D'autre part, Z étant \mathcal{H} -mesurable (car $\mathcal{G} \subset \mathcal{H}$),

$$\begin{aligned} \mathbf{E}(Z \mathbf{E}[X \mid \mathcal{H}]) &= \mathbf{E}(\mathbf{E}[ZX \mid \mathcal{H}]) \\ &= \mathbf{E}[ZX]. \end{aligned}$$

Il résulte que

$$\mathbf{E}[Z \mathbf{E}(\mathbf{E}[X \mid \mathcal{H}] \mid \mathcal{G})] = \mathbf{E}[ZX],$$

ce qui implique le résultat, grâce à la caractérisation de $\mathbf{E}[X \mid \mathcal{G}]$. \square

Remarque 0.3.7. Avec les mêmes hypothèses que dans la proposition précédente, on a également $\mathbf{E}(\mathbf{E}[X \mid \mathcal{G}] \mid \mathcal{H}) = \mathbf{E}[X \mid \mathcal{G}]$ presque sûrement, car $\mathbf{E}[X \mid \mathcal{G}]$ est \mathcal{H} -mesurable.

0.4 Espérance conditionnelle par rapport à des v. a.

L'espérance conditionnelle par rapport à une sous-tribu permet de définir l'espérance conditionnelle par rapport à une variable aléatoire Y , en considérant la tribu engendrée $\mathcal{G} = \sigma(Y)$.

Définition 0.4.1

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{F}, \mathbf{P})$. Si \mathcal{G} est la tribu engendrée par une variable aléatoire Y , alors on note $\mathbf{E}[X \mid Y]$ à la place de $\mathbf{E}[X \mid \mathcal{G}]$.

De même, si $\mathcal{G} = \sigma(Y_1, \dots, Y_n)$, on note l'espérance conditionnelle $\mathbf{E}[X \mid Y_1, \dots, Y_n]$.

Théorème 0.4.2

Soit X une variable aléatoire réelle sur (Ω, \mathcal{F}, P) et soit \mathcal{G} la sous-tribu de \mathcal{F} engendrée par une v.a. $Y : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$.

Alors il existe une application borélienne $h : E \rightarrow \mathbb{R}$ telle que

$$\mathbf{E}[X \mid Y] = h(Y).$$

Preuve Soit $A \in \mathcal{G} = \sigma(Y)$. Par définition de la tribu engendrée, il existe $B \in \mathcal{E}$ tel que $A = Y^{-1}(B) = \{\omega \in \Omega : Y(\omega) \in B\}$.

On a donc $\mathbb{1}_A = \mathbb{1}_B(Y)$. Par linéarité, toute fonction étagée \mathcal{G} -mesurable s'écrit comme fonction étagée \mathcal{E} -mesurable de Y . Par densité, si Z est une v. a. réelle dans $L^1(\Omega, \mathcal{G}, \mathbf{P})$, alors Z s'écrit $Z = h(Y)$ où $h : E \rightarrow \mathbb{R}$ est une application borélienne.

En prenant $Z = \mathbf{E}[X \mid Y]$, on obtient le résultat. \square

Dans le cas où (X, Y) est un vecteur aléatoire dont la loi est absolument continue par rapport à la mesure de Lebesgue, l'espérance conditionnelle $\mathbf{E}[X \mid Y]$ se calcule facilement.

Proposition 0.4.3

Soit (X, Y) un vecteur aléatoire à valeur dans \mathbb{R}^2 , admettant une densité $f_{(X,Y)}$. On suppose $\mathbf{E}[|X|] < \infty$ et $f_Y(y) = \int f_{(X,Y)}(x, y) dx > 0$ pour tout y . On pose pour $x \in \mathbb{R}$,

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

et pour $y \in \mathbb{R}$,

$$h(y) = \int_{\mathbb{R}} x f_{Y=y}(x) dx$$

Alors, on a

$$\mathbf{E}[X \mid Y] = h(Y) \quad p.s.$$

Preuve Soit $A \in \sigma(Y)$. Par définition de $\sigma(Y)$, il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $A = Y^{-1}(B)$. Montrons que $\mathbf{E}[h(Y)\mathbb{1}_A] = \mathbf{E}[X\mathbb{1}_A]$.

$$\begin{aligned}
\mathbf{E}[X\mathbb{1}_A] &= \mathbf{E}[X\mathbb{1}_B(Y)] \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \mathbb{1}_B(y) f_{(X,Y)}(x,y) dx dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \mathbb{1}_B(y) f_Y(y) f_{X|Y=y}(x) dx dy \\
&= \int_{\mathbb{R}} \mathbb{1}_B(y) f_Y(y) \left(\int_{\mathbb{R}} x f_{X|Y=y}(x) dx \right) dy \\
&= \int_{\mathbb{R}} \mathbb{1}_B(y) f_Y(y) h(y) dy \\
&= \mathbf{E}[h(Y)\mathbb{1}_B(Y)] = \mathbf{E}[h(Y)\mathbb{1}_A].
\end{aligned}$$

La caractérisation de l'espérance conditionnelle $\mathbf{E}[X \mid Y]$ permet alors de conclure que $\mathbf{E}[X \mid Y] = h(Y)$ p.s. \square

Première partie

Martingales (4 séances)

Chapitre 1

Quelques généralités sur les processus aléatoires

1.1 Rappels de théorie de la mesure : lemme des classes monotones

Supposons qu'une tribu \mathcal{A} de Ω soit générée par une collection \mathcal{C} de sous-ensembles de Ω . Existe-t-il des conditions sur \mathcal{C} telles qu'une mesure sur $\mathcal{A} = \sigma(\mathcal{C})$ soit caractérisée par ses valeurs sur \mathcal{C} ?

Nous verrons que l'une des conditions sur \mathcal{C} est la stabilité par intersections finies. Les définitions suivantes précisent les propriétés de stabilité qui sont pertinentes pour répondre à la question.

Définition 1.1.1 (π -système)

Une collection \mathcal{C} d'ensembles est appelée π -système si elle est stable par intersections finies. Cela peut s'énoncer comme suit : pour tout $A, B \in \mathcal{C}$, $A \cap B \in \mathcal{C}$.

Définition 1.1.2

Une collection \mathcal{C} d'ensembles est dite

- *stable par limites croissantes si pour tous ensembles $(A_n)_{n \in \mathbb{N}}$ in \mathcal{C} tels que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$, l'ensemble $\bigcup_{n \in \mathbb{N}} A_n$ est dans \mathcal{C} .*
- *stable par différences si pour tous ensembles A, B dans \mathcal{C} tels que $B \subset A$, l'ensemble $A \setminus B$ est dans \mathcal{C} .*

Définition 1.1.3 (*classes monotone*)

Une collection \mathcal{C} de parties de Ω est appelée classe monotone (ou λ -système dans la terminologie de Dynkin) si les trois conditions suivantes sont vérifiées :

- (i) Ω est dans \mathcal{C} ;
- (ii) \mathcal{C} est stable par limites croissantes ;
- (iii) \mathcal{C} est stable par différences.

Lemme 1.1.4

Soit \mathcal{C} une classe monotone. Si la classe \mathcal{C} est également un π -système, alors \mathcal{C} est une tribu.

Preuve Since $\Omega \in \mathcal{C}$ and \mathcal{C} is closed under differences, we can claim that \mathcal{C} is closed under complementation and that $\emptyset \in \mathcal{C}$.

To prove that \mathcal{C} is a σ -algebra, it remains to prove that \mathcal{C} is closed under countable unions.

Let $(A_n)_{n \in \mathbb{N}}$ be in \mathcal{C} .

For all $n \in \mathbb{N}$, the set $B_n = \bigcup_{k \leq n} A_k$ is in \mathcal{C} (since $\Omega \setminus B_n = \bigcap_{k \leq n} (\Omega \setminus A_k)$ is in \mathcal{C}).

The sequence $(B_n)_{n \in \mathbb{N}}$ is non-decreasing, then $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} B_n$ is in \mathcal{C} . \square

Théorème 1.1.5 (*Lemme des classes monotones*)

Si \mathcal{C} est un π -système, alors la plus petite classe monotone contenant \mathcal{C} est $\mathcal{D} = \sigma(\mathcal{C})$.

Preuve • The intersection of any family of monotone classes is a monotone class : if $(\mathcal{C}_t)_{t \in \mathcal{T}}$ are monotone classes, then $\bigcap_{t \in \mathcal{T}} \mathcal{C}_t$ is a monotone class.

The intersection \mathcal{D} of all the monotone classes that contain \mathcal{C} is the smallest one. The σ -algebra $\sigma(\mathcal{C})$ is obviously a monotone class which contains \mathcal{C} . Then, we have $\mathcal{D} \subset \sigma(\mathcal{C})$. It just remains to prove that \mathcal{D} is a σ -algebra to conclude. According to Lemma 1.1.4, it suffices to prove that \mathcal{D} is closed under finite intersections.

In order to do that, for all $B \subset \Omega$, we define $\mathcal{D}_B = \{A \in \mathcal{D} : A \cap B \in \mathcal{D}\}$ and we prove that $\mathcal{D}_B = \mathcal{D}$ for all $B \in \mathcal{D}$.

• For any $B \subset \Omega$, we prove that the collection \mathcal{D}_B is closed under non-decreasing limits and under differences.

Let $(A_n)_{n \in \mathbb{N}}$ be a non-decreasing sequence of sets in \mathcal{D}_B . We have

$$\left(\bigcup_{n \in \mathbb{N}} A_n \right) \cap B = \bigcup_{n \in \mathbb{N}} (A_n \cap B),$$

where $(A_n \cap B)_{n \in \mathbb{N}}$ is a non-decreasing sequence and, for all $n \in \mathbb{N}$, $A_n \cap B$ is in the monotone class \mathcal{D} . Then, $\bigcup_{n \in \mathbb{N}} (A_n \cap B)$ is in \mathcal{D} . This leads to $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}_B$.

Let A_1, A_2 in \mathcal{D}_B with $A_2 \subset A_1$. We have

$$(A_2 \setminus A_1) \cap B = (A_2 \cap B) \setminus (A_1 \cap B),$$

where $A_1 \cap B$ and $A_2 \cap B$ are in the monotone class \mathcal{D} and $A_2 \cap B \subset A_1 \cap B$. Then, $(A_2 \cap B) \setminus (A_1 \cap B)$ is in \mathcal{D} . This leads to $(A_2 \setminus A_1) \cap B \in \mathcal{D}$.

When $B \in \mathcal{D}$, we also have $\Omega \in \mathcal{D}_B$ and we can conclude that \mathcal{D}_B is a monotone class.

- We now prove that for $B \in \mathcal{C}$, $\mathcal{D}_B = \mathcal{D}$.

For all $C \in \mathcal{C}$, we have $B \cap C$ is in the π -system $\mathcal{C} \subset \mathcal{D}$.

Since $C \in \mathcal{D}$, we can claim that $C \in \mathcal{D}_B$.

The monotone class \mathcal{D}_B contains \mathcal{C} and is included in \mathcal{D} . Then $\mathcal{D}_B = \mathcal{D}$.

- We now prove that for $B \in \mathcal{D}$, $\mathcal{D}_B = \mathcal{D}$.

For all $C \in \mathcal{C}$, we have $B \in \mathcal{D}_C$ because $\mathcal{D}_C = \mathcal{D}$. Then, $B \cap C$ is in \mathcal{D} and thus $C \in \mathcal{D}_B$.

As previously, the monotone class \mathcal{D}_B contains \mathcal{C} and is included in \mathcal{D} . Then $\mathcal{D}_B = \mathcal{D}$.

- The fact that $\mathcal{D}_B = \mathcal{D}$ for all $B \in \mathcal{D}$ implies that \mathcal{D} is closed under finite intersections. By Lemma 1.1.4, the fact that the monotone class \mathcal{D} is a π -system implies that \mathcal{D} is a σ -algebra. \square

Théorème 1.1.6

Soient μ et ν deux mesures de probabilités sur l'espace mesurable (Ω, \mathcal{A}) .
Supposons que μ et ν coïncident sur un π -système $\mathcal{C} \subset \mathcal{A}$ tel que $\sigma(\mathcal{C}) = \mathcal{A}$.
Alors, $\mu = \nu$.

Le résultat peut également être formulé en : Si $\mu(C) = \nu(C)$ pour tout C dans le π -système \mathcal{C} , alors $\mu(A) = \nu(A)$ pour tout A dans $\mathcal{A} = \sigma(\mathcal{C})$.

Preuve Let us define $\mathcal{B} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$. Our goal is to prove that $\mathcal{B} = \mathcal{A}$.

First, we show that \mathcal{B} is a monotone class :

- Since $\mu(\Omega) = \nu(\Omega) = 1$, we have $\Omega \in \mathcal{B}$.
- We show that \mathcal{B} is closed under non-decreasing limits.

Let $(A_n)_{n \in \mathbb{N}}$ be elements of \mathcal{B} such that $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$ and set $A = \bigcup_{n \in \mathbb{N}} A_n$.

The monotone convergence theorem for measures implies $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$ and $\nu(A) = \lim_{n \rightarrow \infty} \nu(A_n)$. By definition of \mathcal{B} , we have $\mu(A_n) = \nu(A_n)$ for all $n \in \mathbb{N}$.

This leads to $\mu(A) = \nu(A)$ and $A \in \mathcal{B}$.

- We show that \mathcal{B} is closed under differences.

Let A, B be elements of \mathcal{B} such that $B \subset A$. We have $\mu(A) = \nu(A)$ and $\mu(B) = \nu(B)$. This implies $\mu(A \setminus B) = \mu(A) - \mu(B) = \nu(A) - \nu(B) = \nu(A \setminus B)$ and then $A \setminus B \in \mathcal{B}$.

Finally \mathcal{B} is a monotone class that contains \mathcal{C} .

Then Theorem 1.1.5 implies $\sigma(\mathcal{C}) \subset \mathcal{B} \subset \mathcal{A}$. We conclude that $\mathcal{B} = \mathcal{A}$. \square

1.2 Définitions et propriétés

Définition 1.2.1 (*Processus stochastique*)

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

On appelle processus aléatoire (ou stochastique) à valeurs dans (E, \mathcal{E}) , une collection de variables aléatoires $(X_t)_{t \in \mathcal{T}}$ à valeurs dans (E, \mathcal{E}) .

Les ensembles d'indices les plus couramment considérés sont les cas $\mathcal{T} = \mathbb{N}, \mathbb{R}_+$ ou \mathbb{R}_+^N . On verra dans le cadre de ce cours que suivant l'ensemble d'indices considéré, l'étude des processus stochastiques revêt une complexité variable.

Exemple Signal temporel, image, champ aérodynamique, ...

Définition 1.2.2 (*Trajectoire d'un processus*)

Soit $X = \{X_t; t \in \mathcal{T}\}$ un processus aléatoire sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. Alors pour tout $\omega \in \Omega$ fixé, l'application $t \mapsto X_t(\omega)$ est appelée trajectoire de X .

Les trajectoires du processus X sont des éléments de $E^{\mathcal{T}}$, l'ensemble des applications de \mathcal{T} dans E .

De manière usuelle, on omet le paramètre ω en gardant à l'esprit que les X_t sont bien des variables aléatoires. Ainsi, l'application $t \mapsto X_t$ peut être appelée *fonction aléatoire*.

Définition 1.2.3

Le processus $\{X_t; t \in \mathcal{T}\}$ est appelé modification du processus $\{Y_t; t \in \mathcal{T}\}$ si pour tout $t \in \mathcal{T}$, $\mathbf{P}(X_t = Y_t) = 1$.

Définition 1.2.4

Les processus $\{X_t; t \in \mathcal{T}\}$ et $\{Y_t; t \in \mathcal{T}\}$ sont dits indistingables si leurs trajectoires sont égales presque sûrement, i.e.

$$\mathbf{P}(\forall t \in \mathcal{T}, X_t = Y_t) = 1.$$

La condition précédente s'écrit $\mathbf{P}\left(\bigcap_{t \in \mathcal{T}} (X_t - Y_t)^{-1}(\{0\})\right) = 1$. Elle requiert donc que $\bigcap_{t \in \mathcal{T}} (X_t - Y_t)^{-1}(\{0\})$ appartienne à la tribu \mathcal{F} , ce qui n'est pas toujours vrai.

Pour les notions suivantes, on considère des processus indexés par $\mathcal{T} = \mathbb{N}$.

Définition 1.2.5 (*Filtration*)

On appelle filtration de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} . On dit que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mathbf{P})$ est un espace de probabilité filtré.

Définition 1.2.6

Un processus $\{X_n; n \in \mathbb{N}\}$ est dit

- adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si pour tout $n \in \mathbb{N}$, X_n est mesurable par rapport à la tribu \mathcal{F}_n ;
- prévisible pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si pour tout $n \geq 1$, X_n est mesurable par rapport à la tribu \mathcal{F}_{n-1} .

Exemple Si $\{X_n; n \in \mathbb{N}\}$ est un processus aléatoire, alors on peut définir pour tout $n \in \mathbb{N}$, la tribu

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \dots, X_n).$$

La collection $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$ est alors une filtration, appelée *filtration naturelle* de X .

1.3 Temps d'arrêt

Définition 1.3.1

Une variable aléatoire $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ est appelée temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si

$$\forall n \in \mathbb{N}, \quad \{T = n\} \in \mathcal{F}_n. \quad (1.1)$$

De manière équivalente, la condition (1.1) définissant un temps d'arrêt peut être remplacée par

$$\forall n \in \mathbb{N}, \quad \{T \leq n\} \in \mathcal{F}_n. \quad (1.2)$$

En effet, si (1.2) est vérifiée, pour tout $n \geq 1$,

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

Donc (1.1) est vérifiée.

Inversement, si (1.1) est vérifiée, pour tout $n \in \mathbb{N}$, on peut écrire

$$\{T \leq n\} = \bigcup_{k \leq n} \{T = k\}.$$

Comme pour tout $k \leq n$, $\{T = k\} \in \mathcal{F}_k$ et $\mathcal{F}_k \subset \mathcal{F}_n$, on a $\{T \leq n\} \in \mathcal{F}_n$, ce qui montre que (1.2) est vérifiée.

Moralement, un temps d'arrêt peut être vu comme un temps aléatoire dont la réalisation n'est fonction que de l'information connue jusqu'à sa valeur.

Exemple Soit $X = \{X_n; n \in \mathbb{N}\}$ un processus discret à valeurs dans \mathbb{R} . On définit le temps aléatoire T par

$$T(\omega) = \inf \{n \in \mathbb{N} : X_n(\omega) \geq 1\}.$$

T est un temps d'arrêt car il ne dépend que des valeurs des X_n pour $n \leq T$. De manière rigoureuse, on exprime l'événement $\{T \leq n\}$ sous la forme

$$\{T \leq n\} = \bigcup_{0 \leq k \leq n} \{X_k \geq 1\} \in \mathcal{F}_n$$

car chaque $\{X_k \geq 1\}$ est dans $\mathcal{F}_k \subset \mathcal{F}_n$. Ce qui prouve que T est un temps d'arrêt.

Par contre, le temps aléatoire $S = \sup \{n \in \mathbb{N} : X_n \geq 1\}$ n'est pas un temps d'arrêt. Il est nécessaire de connaître l'ensemble des réalisations des X_n pour définir la valeur de S .

Définition 1.3.2

Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration sur l'espace $(\Omega, \mathcal{F}, \mathbf{P})$, et soit T un temps d'arrêt. La collection

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

est une tribu appelée tribu du temps d'arrêt T .

On laisse la justification de cette définition en exercice.

Théorème 1.3.3

Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration sur l'espace $(\Omega, \mathcal{F}, \mathbf{P})$.

- Si S et T sont deux temps d'arrêt par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, alors $S \wedge T$ et $S \vee T$ sont aussi des temps d'arrêt par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$.
- Si S et T sont deux temps d'arrêt tels que $S \leq T$ p.s., alors $\mathcal{F}_S \subset \mathcal{F}_T$.
- Si $X = \{X_n; n \in \mathbb{N}\}$ est un processus adapté et T un temps d'arrêt, alors la variable aléatoire X_T , définie par $X_T(\omega) = X_{T(\omega)}(\omega)$ pour tout $\omega \in \Omega$, est \mathcal{F}_T -mesurable.

Preuve

- On ne traite que le cas de $S \wedge T$. Pour tout $n \in \mathbb{N}$,

$$\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}.$$

Or, $\{S \leq n\} \in \mathcal{F}_n$ et $\{T \leq n\} \in \mathcal{F}_n$ car S et T sont des temps d'arrêt. Donc $\{S \wedge T \leq n\} \in \mathcal{F}_n$, ce qui montre que $S \wedge T$ est un temps d'arrêt.

- Si $A \in \mathcal{F}_S$ alors

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

donc $A \in \mathcal{F}_T$.

- Pour tout $A \in \mathcal{B}(\mathbb{R})$, on veut montrer que $\{X_T \in A\} \in \mathcal{F}_T$.

$$\begin{aligned} \{X_T \in A\} \cap \{T \leq n\} &= \bigcup_{0 \leq k \leq n} \{X_T \in A\} \cap \{T = k\} \\ &= \bigcup_{0 \leq k \leq n} \{X_k \in A\} \cap \{T = k\} \in \mathcal{F}_n \end{aligned}$$

car $\{X_k \in A\} \cap \{T = k\}$ est dans $\mathcal{F}_k \subset \mathcal{F}_n$ pour tout $k \leq n$.

□

1.4 Loi d'un processus

Un processus stochastique $\{X_t; t \in \mathcal{T}\}$ à valeurs dans (E, \mathcal{E}) peut être considéré comme une application à valeurs dans l'espace des trajectoires

$$\begin{aligned} X : \Omega &\rightarrow E^{\mathcal{T}} \\ \omega &\mapsto (t \mapsto X_t(\omega)). \end{aligned}$$

Peut-on munir l'ensemble $E^{\mathcal{T}}$, des fonctions de \mathcal{T} dans E , d'une tribu rendant mesurable l'application X ?

L'existence d'une telle tribu permettrait de considérer le processus $\{X_t; t \in \mathcal{T}\}$ comme une variable aléatoire à valeurs dans $E^{\mathcal{T}}$ et par suite, de pouvoir considérer sa loi.

La motivation essentielle est d'être capable de considérer des probabilités d'événements faisant intervenir un nombre potentiellement indénombrable d'indices dans \mathcal{T} , tels que

$$\mathbf{P}(\text{Trajectoire de } X \in A) \quad \text{ou} \quad \mathbf{P}\left(\sup_{t \in \mathcal{T}} X_t \in B\right)$$

où A est un événement dans l'espace $E^{\mathcal{T}}$ des trajectoires et $B \in \mathcal{E}$.

Remarque 1.4.1. Lorsque \mathcal{T} est un ensemble fini, qu'on peut expliciter $\mathcal{T} = \{t_1, \dots, t_N\}$, l'ensemble $E^\mathcal{T}$ peut être identifié à l'ensemble produit E^N .

Ainsi, on peut munir $E^\mathcal{T}$ de la tribu produit $\mathcal{E}^{\otimes N}$.

1.4.1 Tribu de Kolmogorov

La tribu de Kolmogorov \mathcal{G} est définie comme la plus petite tribu rendant mesurables les applications

$$\begin{aligned}\varphi_t : E^\mathcal{T} &\rightarrow E \\ x &\mapsto \varphi_t(x) = x(t)\end{aligned}$$

pour tout $t \in \mathcal{T}$. On a donc $\mathcal{G} = \sigma(\varphi_t^{-1}(A); A \in \mathcal{E}, t \in \mathcal{T})$.

Proposition 1.4.2

Etant donné un processus stochastique $\{X_t; t \in \mathcal{T}\}$ à valeurs dans (E, \mathcal{E}) et \mathcal{G} la tribu de Kolmogorov sur l'espace $E^\mathcal{T}$, l'application

$$\begin{aligned}X : (\Omega, \mathcal{F}) &\rightarrow (E^\mathcal{T}, \mathcal{G}) \\ \omega &\mapsto (t \mapsto X_t(\omega)),\end{aligned}$$

est mesurable (et définit donc une variable aléatoire à valeurs dans $E^\mathcal{T}$).

Preuve Il s'agit de montrer que pour tout $B \in \mathcal{G}$, on a $X^{-1}(B) \in \mathcal{F}$.

Comme la tribu \mathcal{G} est engendrée par l'ensemble $\{\varphi_t^{-1}(A); t \in \mathcal{T}, A \in \mathcal{E}\}$, il suffit de montrer que pour tout $B = \varphi_t^{-1}(A)$ pour un certain $t \in \mathcal{T}$ et un certain $A \in \mathcal{E}$, on a $X^{-1}(B) \in \mathcal{F}$.

On a $X_t = \varphi_t \circ X$, donc

$$X^{-1}(B) = X^{-1}(\varphi_t^{-1}(A)) = (\varphi_t \circ X)^{-1}(A) = X_t^{-1}(A) \in \mathcal{F}$$

car X_t est une variable aléatoire (donc mesurable $(\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$). \square

Pour toute partie I de \mathcal{T} , on note également $\mathcal{G}(I)$ la tribu engendrée par les applications $x \mapsto x(t)$ pour tout $t \in I$, c'est-à-dire $\mathcal{G}(I) = \sigma(\varphi_t^{-1}(A); A \in \mathcal{E}, t \in I)$.

Proposition 1.4.3

$$\mathcal{G} = \bigvee_{I \subset \mathcal{T}} \mathcal{G}(I) = \bigvee_{I \text{ fini}, I \subset \mathcal{T}} \mathcal{G}(I),$$

en notant $\bigvee \mathcal{G}(I)$ la tribu engendrée par les $\mathcal{G}(I)$.

Preuve On a les inclusions suivantes

$$\bigvee_{I \subset \mathcal{T}} \mathcal{G}(I) \supset \bigvee_{I \text{ fini}, I \subset \mathcal{T}} \mathcal{G}(I) \supset \bigvee_{t \in \mathcal{T}} \mathcal{G}(\{t\}).$$

Or, $\bigvee_{I \subset \mathcal{T}} \mathcal{G}(I) = \mathcal{G}$ (car \mathcal{T} est une partie de \mathcal{T}) et $\bigvee_{t \in \mathcal{T}} \mathcal{G}(\{t\}) = \mathcal{G}$ (par définition de \mathcal{G}). On a donc nécessairement l'égalité énoncée. \square

En écrivant l'ensemble fini $I = \{t_1, \dots, t_N\}$, la tribu $\mathcal{G}(I)$ peut être identifiée à la tribu produit $\mathcal{E}^{\otimes N}$ (de l'espace produit E^N) par :

toute partie $A \subset E^{\mathcal{T}}$ est dans $\mathcal{G}(I)$ si et seulement si elle s'écrit $A = \{x : (x(t_1), \dots, x(t_N)) \in B\}$ pour un certain B dans la tribu produit $\mathcal{E}^{\otimes N}$.

La tribu de Kolmogorov est donc engendrée par les ensembles de la forme

$$C = \{\omega : \omega(t_1) \in B_1, \dots, \omega(t_n) \in B_n\}$$

où $n \geq 1$, $t_1, \dots, t_n \in \mathcal{T}$ et $B_1, \dots, B_n \in \mathcal{E}$. Ces ensembles sont appelés *cylindres*.

Un processus pouvant être considéré comme une variable aléatoire à valeurs dans $(E^{\mathcal{T}}, \mathcal{G})$, il est licite de définir :

Définition 1.4.4

La loi du processus $X = \{X_t; t \in \mathcal{T}\}$ sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ est la probabilité sur $(E^{\mathcal{T}}, \mathcal{G})$ définie comme l'image par X de la mesure de probabilité \mathbf{P} .

Inversement, toute probabilité μ sur $(E^{\mathcal{T}}, \mathcal{G})$ est également la loi d'un processus stochastique. Il suffit de considérer l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P}) = (E^{\mathcal{T}}, \mathcal{G}, \mu)$ et le *processus canonique* défini par

$$X_t(\omega) = \omega(t).$$

1.4.2 Théorème de Kolmogorov

Définition 1.4.5

La famille des lois fini-dimensionnelles du processus X est la famille $(\mu_I : I \subset \mathcal{T})$ des lois des variables aléatoires $(X_t)_{t \in I}$, pour l'ensemble des parties finies I de \mathcal{T} .

Les distributions fini-dimensionnelles $(\mu_I)_{I \subset \mathcal{T}}$ du processus X sont donc définies pour toute partie finie I de \mathcal{T} (noté $I \subset \subset \mathcal{T}$) par

$$\mu_I((B_t)_{t \in I}) = \mathbf{P}(\forall t \in I, X_t \in B_t)$$

pour toute famille $(B_t)_{t \in I}$ dans $\mathcal{E}^{\#I}$ (qui engendre la tribu produit $\mathcal{E}^{\otimes I}$).

Exemple Lorsque $\mathcal{T} = \mathbb{R}_+$, toute partie finie de \mathcal{T} est de la forme $I = \{t_1, t_2, \dots, t_n\}$, où $t_1 < t_2 < \dots < t_n$. Ainsi, les distributions fini-dimensionnelles de X s'écrivent

$$\mu_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = \mathbf{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n).$$

Toute mesure de probabilité μ sur $(E^{\mathcal{T}}, \mathcal{G})$ définit donc un système de distributions fini-dimensionnelles $(\mu_I)_{I \subset \mathcal{T}}$. Chaque μ_I est l'image de μ par la projection $x \mapsto (x(t), t \in I)$.

Le résultat suivant montre qu'une telle mesure est entièrement caractérisée par les lois fini-dimensionnelles $(\mu_I)_{I \subset \mathcal{T}}$.

Proposition 1.4.6

- Les lois fini-dimensionnelles d'un processus vérifient la condition de compatibilité suivante :

Si I est une partie finie de \mathcal{T} et si $t \in \mathcal{T} \setminus I$, alors

$$\forall A \in \mathcal{E}^{\otimes I}; \quad \mu_{I \cup \{t\}}(A \times E) = \mu_I(A).$$

- Si deux mesures de probabilité sur $(E^{\mathcal{T}}, \mathcal{G})$ admettent la même famille de distributions fini-dimensionnelles, alors elles sont égales.

Preuve

- Si $A \in \mathcal{E}^{\otimes I}$, on a

$$\mu_{I \cup \{t\}}(A \times E) = \mathbf{P}((X_s)_{s \in I \cup \{t\}} \in A \times E) = \mathbf{P}((X_s)_{s \in I} \in A) = \mu_I(A).$$

- Soient μ et ν deux mesures de probabilité sur $(E^{\mathcal{T}}, \mathcal{G})$, ayant les mêmes distributions fini-dimensionnelles $(\mu_I)_{I \subset \mathcal{T}}$. Elles coïncident sur tous les $\mathcal{G}(I)$ qui engendrent \mathcal{G} (argument de classes monotones).

□

Exemple En reprenant le cas particulier où $\mathcal{T} = \mathbb{R}_+$, la relation de compatibilité s'écrit

$$\begin{aligned} \mathbf{P}(X_{t_1} \in B_1, \dots, X_{t_{j-1}} \in B_{j-1}, X_{t_j} \in E, X_{t_{j+1}} \in B_{j+1}, \dots, X_{t_n} \in B_n) \\ = \mathbf{P}(X_{t_1} \in B_1, \dots, X_{t_{j-1}} \in B_{j-1}, X_{t_{j+1}} \in B_{j+1}, \dots, X_{t_n} \in B_n) \end{aligned}$$

pour toute partie finie $I = \{t_1, t_2, \dots, t_n\}$ de \mathcal{T} , avec $t_1 < t_2 < \dots < t_n$.

Grâce à la caractérisation de la loi de tout processus par ses distributions fini-dimensionnelles, l'égalité en loi de processus peut se définir par

Définition 1.4.7

Les processus $\{X_t; t \in \mathcal{T}\}$ et $\{Y_t; t \in \mathcal{T}\}$ sont dits identiques en loi si leurs distributions fini-dimensionnelles sont égales.

On note $\{X_t; t \in \mathcal{T}\} \stackrel{(d)}{=} \{Y_t; t \in \mathcal{T}\}$.

On a vu qu'une mesure de probabilité sur $(E^{\mathcal{T}}, \mathcal{G})$, c'est-à-dire la loi d'un processus stochastique $\{X_t; t \in \mathcal{T}\}$, définit des distributions fini-dimensionnelles. Réciproquement, le résultat suivant montre que partant d'un système de mesures de probabilité sur les fonctions d'un nombre fini de paramètres, E^I avec $I \subset \mathcal{T}$ et I fini, on peut construire une mesure de probabilité sur $E^{\mathcal{T}}$ sous contrainte d'une hypothèse de compatibilité.

Théorème 1.4.8 (Théorème d'extension de Kolmogorov)

Supposons que $E = \mathbb{R}^N$ muni de la tribu borélienne $\mathcal{E} = \mathcal{B}(\mathbb{R}^N)$.

Toute famille $(\mu_I)_{I \subset \subset \mathcal{T}}$ de probabilités sur $(E^I, \mathcal{E}^{\otimes I})$ indexée par les parties finies de \mathcal{T} vérifiant la relation de compatibilité, est la famille des distributions fini-dimensionnelles d'une (unique) probabilité sur $(E^{\mathcal{T}}, \mathcal{G})$.

Chapitre 2

Martingales

Le mot *martingale* fut historiquement employé pour désigner une stratégie gagnante aux jeux de hasard. Dans un contexte mathématique, il désigne un processus aléatoire discret, permettant de décrire un *jeu juste* à deux joueurs.

2.1 Définitions et exemples

Définition 2.1.1

On appelle martingale (resp. sous-martingale, surmartingale) un processus $X = \{X_n; n \in \mathbb{N}\}$ muni d'une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ vérifiant les trois conditions suivantes :

- (i) *mesurabilité* : X est adapté, i. e. X_n est \mathcal{F}_n -mesurable pour tout $n \in \mathbb{N}$;
- (ii) *intégrabilité* : Pour tout $n \in \mathbb{N}$, $X_n \in L^1(\Omega, \mathcal{F}, \mathbf{P})$;
- (iii) *relation de martingale* :

$$\forall n \in \mathbb{N}; \quad X_n = \mathbf{E}[X_{n+1} \mid \mathcal{F}_n] \quad p.s.$$

$$(resp. \ X_n \leq \mathbf{E}[X_{n+1} \mid \mathcal{F}_n] \ p.s., \ X_n \geq \mathbf{E}[X_{n+1} \mid \mathcal{F}_n] \ p.s.)$$

Notons que la 3ème condition est équivalente à

$$\forall n, m \in \mathbb{N} \text{ tels que } n \leq m; \quad X_n = \mathbf{E}[X_m \mid \mathcal{F}_n] \quad p.s.,$$

que si X est une surmartingale, alors $-X$ est une sous-martingale et qu'une **martingale est un processus qui est à la fois une sous-martingale et une surmartingale**.

D'autre part, il est clair qu'une **martingale est une martingale par rapport à sa filtration naturelle**. En effet, si $X = \{X_n; n \in \mathbb{N}\}$ est une martingale par rapport à une

filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, alors pour tout $n \in \mathbb{N}$,

$$\begin{aligned}\mathbf{E}[X_{n+1} \mid \sigma(X_0, \dots, X_n)] &= \mathbf{E}(\mathbf{E}[X_{n+1} \mid \mathcal{F}_n] \mid \sigma(X_0, \dots, X_n)) \\ &= \mathbf{E}(X_n \mid \sigma(X_0, \dots, X_n)) \\ &= X_n \quad \text{p.s.}\end{aligned}$$

en utilisant le fait que $\sigma(X_0, \dots, X_n) \subset \mathcal{F}_n$ pour tout n , par définition de la filtration engendrée.

Exemple Soient $(\xi_n)_{n \in \mathbb{N}}$ des variables aléatoires i.i.d., intégrables, de moyenne nulle.

On considère, pour tout $n \in \mathbb{N}$, $X_n = \sum_{k \leq n} \xi_k$ et $\mathcal{F}_n = \sigma(\xi_k; k \leq n)$.

Le processus $\{X_n; n \in \mathbb{N}\}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

En effet,

- (i) pour tout $n \in \mathbb{N}$, $X_n = \sum_{k \leq n} \xi_k$ est mesurable par rapport à $\mathcal{F}_n = \sigma(\xi_k; k \leq n)$;
- (ii) pour tout $n \in \mathbb{N}$, X_n est intégrable, comme somme finie de variables intégrables;
- (iii) pour tout $n \in \mathbb{N}$,

$$\begin{aligned}\mathbf{E}[X_{n+1} \mid \mathcal{F}_n] - X_n &= \mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] \\ &= \mathbf{E}[\xi_{n+1} \mid \mathcal{F}_n] = \mathbf{E}[\xi_{n+1}] = 0,\end{aligned}$$

en utilisant le fait que ξ_{n+1} est indépendante de $\mathcal{F}_n = \sigma(\xi_k; k \leq n)$.

Exemple Soit Y une variable aléatoire dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration.

En posant

$$\forall n \in \mathbb{N}, \quad X_n = \mathbf{E}[Y \mid \mathcal{F}_n],$$

on définit une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

En effet, pour tout $n \in \mathbb{N}$, $\mathbf{E}[Y \mid \mathcal{F}_n]$ est \mathcal{F}_n -mesurable et dans $L^1(\Omega, \mathcal{F}, P)$. De plus, comme $\mathcal{F}_n \subset \mathcal{F}_{n+1}$,

$$\begin{aligned}\mathbf{E}[X_{n+1} \mid \mathcal{F}_n] &= \mathbf{E}(\mathbf{E}[Y \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n) \\ &= \mathbf{E}[Y \mid \mathcal{F}_n] \quad \text{p.s.} \\ &= X_n \quad \text{p.s.}\end{aligned}$$

Proposition 2.1.2

Si $X = \{X_n; n \in \mathbb{N}\}$ est une martingale, alors on a $\mathbf{E}[X_n] = \mathbf{E}[X_0]$ pour tout $n \in \mathbb{N}$.

Preuve En prenant l'espérance de la relation de martingale $X_n = \mathbf{E}[X_{n+1} \mid \mathcal{F}_n]$ presque sûrement, on obtient

$$\forall n \in \mathbb{N}, \quad \mathbf{E}[X_n] = \mathbf{E}(\mathbf{E}[X_{n+1} \mid \mathcal{F}_n]) = \mathbf{E}[X_{n+1}],$$

ce qui implique le résultat. \square

Proposition 2.1.3

Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ une application convexe, et soit $\{X_n; n \in \mathbb{N}\}$ un processus adapté tel que $\mathbf{E}[\varphi(X_n)] < \infty$ pour tout $n \in \mathbb{N}$.

- Si $\{X_n; n \in \mathbb{N}\}$ est une martingale, $\{\varphi(X_n); n \in \mathbb{N}\}$ est une sous-martingale.
- Si $\{X_n; n \in \mathbb{N}\}$ est une sous-martingale et si φ est croissante, $\{\varphi(X_n); n \in \mathbb{N}\}$ est une sous-martingale.

Preuve Cf. [10] Proposition 4.2.2 p. 79 ou [13] Proposition 12.1.1 p. 166. \square

2.2 Décomposition de Doob

Théorème 2.2.1

Toute sous-martingale $\{X_n; n \in \mathbb{N}\}$ par rapport à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ se décompose de façon unique sous la forme

$$\forall n \in \mathbb{N}, \quad X_n = M_n + A_n,$$

où $\{M_n; n \in \mathbb{N}\}$ est une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et $\{A_n; n \in \mathbb{N}\}$ un processus croissant prévisible par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ issu de 0, i.e. tel que $A_0 = 0$.

Preuve • Supposons qu'une telle décomposition existe. On a

$$\forall n \in \mathbb{N}, \quad X_{n+1} - X_n = M_{n+1} - M_n + A_{n+1} - A_n,$$

ce qui implique

$$\mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = \mathbf{E}[M_{n+1} - M_n \mid \mathcal{F}_n] + \mathbf{E}[A_{n+1} - A_n \mid \mathcal{F}_n].$$

Comme $\{M_n; n \in \mathbb{N}\}$ est une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et $\{A_n; n \in \mathbb{N}\}$ est prévisible par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, on en déduit

$$\forall n \in \mathbb{N}, \quad \mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] = A_{n+1} - A_n.$$

Ainsi, le processus $\{A_n; n \in \mathbb{N}\}$ est totalement déterminé par $A_0 = 0$ et la relation de récurrence ci-dessus. Par suite, le processus $\{M_n; n \in \mathbb{N}\}$ est totalement déterminé par $M_n = X_n - A_n$ pour tout $n \in \mathbb{N}$.

- On définit $A_0 = 0$ et $A_{n+1} = A_n + \mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n]$ pour tout $n \geq 0$. $\{X_n; n \in \mathbb{N}\}$ étant une sous-martingale, le processus $\{A_n; n \in \mathbb{N}\}$ est croissant. De plus, pour tout $n \geq 0$, A_{n+1} est \mathcal{F}_n -mesurable, i.e. $\{A_n; n \in \mathbb{N}\}$ est prévisible.

On pose alors $M_n = X_n - A_n$ pour tout $n \geq 0$. On a alors

$$\begin{aligned}\mathbf{E}[M_{n+1} - M_n \mid \mathcal{F}_n] &= \mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] - \mathbf{E}[A_{n+1} - A_n \mid \mathcal{F}_n] \\ &= \mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] - \mathbf{E}(\mathbf{E}[X_{n+1} - X_n \mid \mathcal{F}_n] \mid \mathcal{F}_n) \\ &= 0.\end{aligned}$$

Comme $\{M_n; n \in \mathbb{N}\}$ est visiblement adapté (on peut le montrer facilement par récurrence), c'est donc une martingale. \square

Application à la définition du crochet d'une martingale :

Soit $\{N_n; n \in \mathbb{N}\}$ une martingale. Le processus $\{N_n^2; n \in \mathbb{N}\}$ est une sous-martingale (car $x \mapsto x^2$ est convexe). On peut appliquer la décomposition de Doob à N^2 . Il existe une martingale $\{M_n; n \in \mathbb{N}\}$ et un processus croissant prévisible $\{A_n; n \in \mathbb{N}\}$ issus de 0, tels que

$$\forall n \in \mathbb{N}, \quad N_n^2 = M_n + A_n.$$

Le processus $\{A_n; n \in \mathbb{N}\}$ est appelé *crochet de la martingale* $\{N_n; n \in \mathbb{N}\}$, et est noté $\langle N \rangle = \{\langle N \rangle_n; n \in \mathbb{N}\}$. Il est défini par :

- $\{\langle N \rangle_n; n \in \mathbb{N}\}$ est l'unique processus croissant prévisible issu de 0 tel que

$$\{N_n^2 - \langle N \rangle_n; n \in \mathbb{N}\} \quad \text{est une martingale.}$$

- De manière équivalente, $\langle N \rangle$ est défini par

$$\begin{cases} \forall n \geq 0, & \mathbf{E}[N_{n+1}^2 - N_n^2 \mid \mathcal{F}_n] = \langle N \rangle_{n+1} - \langle N \rangle_n, \\ \text{et} & \langle N \rangle_0 = 0. \end{cases}$$

- De manière équivalente, $\langle N \rangle$ est défini par

$$\begin{cases} \forall n \geq 0, & \mathbf{E}[(N_{n+1} - N_n)^2 \mid \mathcal{F}_n] = \langle N \rangle_{n+1} - \langle N \rangle_n, \\ \text{et} & \langle N \rangle_0 = 0. \end{cases}$$

2.3 Théorème d'arrêt

Théorème 2.3.1

Si $X = \{X_n; n \in \mathbb{N}\}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -martingale et T un $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -temps d'arrêt borné presque sûrement, alors $\mathbf{E}[X_T] = \mathbf{E}[X_0]$.

Preuve On écrit

$$X_T(\omega) = \sum_{n=0}^{\infty} X_n(\omega) \mathbb{1}_{\{T(\omega)=n\}}.$$

En supposant $\mathbf{P}(T \leq M) = 1$, où $M \in \mathbb{N}^*$, par linéarité de l'espérance on a

$$\mathbf{E}[X_T] = \sum_{n=0}^M \mathbf{E}[X_n \mathbb{1}_{\{T=n\}}].$$

Comme $\{T = n\} \in \mathcal{F}_n$, on a

$$\begin{aligned} \mathbf{E}[X_T] &= \sum_{n=0}^M \mathbf{E}[\mathbf{E}[X_M | \mathcal{F}_n] \mathbb{1}_{\{T=n\}}] = \sum_{n=0}^M \mathbf{E}[X_M \mathbb{1}_{\{T=n\}}] \\ &= \mathbf{E}\left[X_M \sum_{n=0}^M \mathbb{1}_{\{T=n\}}\right] = \mathbf{E}[X_M] = \mathbf{E}[X_0]. \end{aligned}$$

□

Attention : Le caractère borné du temps d'arrêt T est obligatoirement requis pour que le Théorème 2.3.1 soit vrai.

Par la suite, on verra que ce résultat peut être étendu aux sous-martingales et surmartingales en :

Théorème 2.3.2

Si $X = \{X_n; n \in \mathbb{N}\}$ est une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -sous-martingale (resp. sur-martingale) et T un $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -temps d'arrêt borné presque sûrement, alors $\mathbf{E}[X_T] \geq \mathbf{E}[X_0]$ (resp. $\mathbf{E}[X_T] \leq \mathbf{E}[X_0]$).

Nous verrons en TD, une certaine forme de réciproque au Théorème 2.3.1. Une amélioration de ce résultat constitue le théorème d'arrêt de Doob.

Théorème 2.3.3 (Théorème d'arrêt)

Soit $X = \{X_n; n \in \mathbb{N}\}$ une martingale par rapport à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Si S et T sont deux temps d'arrêt par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, bornés presque sûrement et tels que $S \leq T$ p.s. alors

$$\mathbf{E}[X_T | \mathcal{F}_S] = X_S \quad \text{p.s.}$$

Preuve Supposons $\mathbf{P}(S \leq T \leq M) = 1$, où $M \in \mathbb{N}^*$.

On a alors $|X_T| \leq |X_0| + |X_1| + \dots + |X_M|$ qui est intégrable. Donc $X_T \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ et de même $X_S \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Il faut démontrer que pour tout $A \in \mathcal{F}_S$, $\mathbf{E}[X_T \mathbb{1}_A] = \mathbf{E}[X_S \mathbb{1}_A]$.

Pour tout $A \in \mathcal{F}_S \subset \mathcal{F}_T$, on définit

$$R = S\mathbb{1}_A + T\mathbb{1}_{A^c}.$$

La variable aléatoire R ainsi définie est un temps d'arrêt car

$$\{R \leq n\} = (A \cap \{S \leq n\}) \cup (A^c \cap \{T \leq n\})$$

or, $A \cap \{S \leq n\} \in \mathcal{F}_n$ (car $A \in \mathcal{F}_S$) et $A^c \cap \{T \leq n\} \in \mathcal{F}_n$ (car $A^c \in \mathcal{F}_T$).

De plus, R est borné par M , donc on peut appliquer le Théorème 2.3.1 qui implique $\mathbf{E}[X_R] = \mathbf{E}[X_0] = \mathbf{E}[X_T]$. Or,

$$\begin{aligned}\mathbf{E}[X_R] &= \mathbf{E}[X_S\mathbb{1}_A + X_T\mathbb{1}_{A^c}] \\ \mathbf{E}[X_T] &= \mathbf{E}[X_T\mathbb{1}_A + X_T\mathbb{1}_{A^c}].\end{aligned}$$

On en déduit $\mathbf{E}[X_T\mathbb{1}_A] = \mathbf{E}[X_S\mathbb{1}_A]$. \square

2.3.1 Exemple : stratégie de jeu

On considère le cadre d'un jeu à deux joueurs, où chaque mise perdue par l'un des joueurs est gagnée par l'autre. On note X_n le gain du joueur 1 par unité de pari lors du jeu n .

On définit, pour tout $n \geq 1$, $\mathcal{F}_n = \sigma(X_k; k \leq n)$.

Le jeu est dit *juste* si $\mathbf{E}[X_n | \mathcal{F}_{n-1}] = 0$, c'est-à-dire si le processus $\{S_n; n \in \mathbb{N}\}$, défini par $S_n = \sum_{k=1}^n X_k$ pour tout $n \in \mathbb{N}$, est une martingale, et injuste sinon.

L'unique marge de manoeuvre du joueur 1, au cours de la successions des jeux $n \geq 1$, est d'ajuster ses mises en fonction des résultats des jeux précédents. On note C_n la mise du joueur 1 lors du jeu n .

Une *stratégie de jeu* consiste en un procédé pour choisir la mise C_n , en fonction des résultats des mises précédentes jusqu'au temps $n - 1$. Ainsi, une stratégie consiste en un processus $C = \{C_n; n \in \mathbb{N}\}$ prévisible par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ (l'histoire des résultats des jeux).

Le gain du joueur lors du jeu n vaut alors $C_n X_n = C_n(S_n - S_{n-1})$ et la somme totale des gains jusqu'au jeu n est

$$Y_n = \sum_{k=1}^n C_k(S_k - S_{k-1}).$$

On note $C \bullet S$ le processus Y et on l'appelle *transformation de martingale* de S par C (c'est l'analogue discret de l'intégrale stochastique $\int C \cdot dS$).

Proposition 2.3.4

Soient $\{S_n; n \in \mathbb{N}\}$ un processus adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et $\{C_n; n \in \mathbb{N}\}$ un processus borné prévisible par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

On définit le processus $C \bullet S = \{(C \bullet S)_n; n \in \mathbb{N}\}$ par

$$(C \bullet S)_0 = 0 \quad \text{et} \quad (C \bullet S)_n = \sum_{k=1}^n C_k(S_k - S_{k-1}) \quad \text{pour tout } n \geq 1.$$

- (i) Si S est une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, alors $C \bullet S$ est une martingale par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$.
- (ii) Si S est une sous-martingale (resp. surmartingale) par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et si $C_n \geq 0$ pour tout $n \geq 1$, alors $C \bullet S$ est une sous-martingale (resp. surmartingale) par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

Preuve Pour tout $n \in \mathbb{N}$, $(C \bullet S)_n$ est mesurable par rapport à \mathcal{F}_n (car pour tout $k \leq n$, C_k et $S_k - S_{k-1}$ sont \mathcal{F}_n -mesurables). De plus, $(C \bullet S)_n$ est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ puisque pour tout $k \leq n$, C_k est borné et $S_k - S_{k-1}$ est intégrable.

Pour tout $n \geq 0$, C_{n+1} est borné et \mathcal{F}_n -mesurable. Donc

$$\begin{aligned} \mathbf{E}[(C \bullet S)_{n+1} - (C \bullet S)_n \mid \mathcal{F}_n] &= \mathbf{E}[C_{n+1}(S_{n+1} - S_n) \mid \mathcal{F}_n] \\ &= C_{n+1} \mathbf{E}[S_{n+1} - S_n \mid \mathcal{F}_n]. \end{aligned}$$

Ainsi, si S est une martingale alors $\mathbf{E}[(C \bullet S)_{n+1} - (C \bullet S)_n \mid \mathcal{F}_n] = 0$ pour tout $n \geq 1$, i.e. $C \bullet S$ est une martingale.

Si S est une sous-martingale (resp. surmartingale) et $C_n \geq 0$ pour tout $n \geq 1$, alors $\mathbf{E}[(C \bullet S)_{n+1} - (C \bullet S)_n \mid \mathcal{F}_n] \geq 0$ (resp. $\mathbf{E}[(C \bullet S)_{n+1} - (C \bullet S)_n \mid \mathcal{F}_n] \leq 0$) pour tout $n \geq 1$, i.e. $C \bullet S$ est une sous-martingale (resp. surmartingale). \square

En conséquence, l'espérance de gain après chaque jeu n est $\mathbf{E}[(C \bullet S)_n] = 0$.

Dans le cadre du jeu à deux joueurs, une stratégie définit non seulement le processus des mises C_n , $n \geq 1$, mais également la définition du moment où le joueur va cesser de jouer en fonction de ses gains jusque là.

Il s'agit alors de définir un temps d'arrêt par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, correspondant à une espérance de gain strictement positive. Le budget du joueur étant fini, ce temps d'arrêt doit être borné (la somme des budgets des deux joueurs est finie). Or, pour tout temps d'arrêt T borné, le théorème d'arrêt montre que $\mathbf{E}[(C \bullet S)_T] = 0$.

Ceci montre qu'il n'existe pas de stratégie gagnante en temps fini pour un jeu juste à deux joueurs.

Retour sur le Théorème d'arrêt On reprend les hypothèses du Théorème 2.3.1.

On pose $H_n = \mathbb{1}_{\{T \geq n\}} = 1 - \mathbb{1}_{\{T \leq n-1\}}$ pour tout $n \geq 1$. Le processus $\{H_n; n \in \mathbb{N}\}$ est borné et prévisible par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

On a

$$\forall n \geq 1, \quad (H \bullet X)_n = X_{n \wedge T} - X_0.$$

Donc le processus $\{X_{n \wedge T}; n \in \mathbb{N}\}$ est une martingale (resp. sous-martingale) si X est une martingale (resp. sous-martingale).

Ainsi, si T est un temps d'arrêt borné par $M \in \mathbb{N}^*$ et si X est une martingale (resp. sous-martingale), on a

$$\mathbf{E}[X_T] = \mathbf{E}[X_{M \wedge T}] = \mathbf{E}[X_0] \quad (\text{resp. } \geq \mathbf{E}[X_0]).$$

Ceci prouve donc les Théorèmes 2.3.1 et 2.3.2.

2.4 Inégalités de martingales

Dans le cadre de l'étude d'une martingale $\{M_n; n \geq 0\}$, il est usuel d'employer la notation

$$\forall n \geq 0, \quad M_n^* = \sup_{k \leq n} |M_k|.$$

Les deux premiers résultats de cette section s'intéressent à la loi et finalement l'intégrabilité de M_n^* . Ils permettront d'étudier la convergence dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ des martingales.

Lemme 2.4.1

Soit $\{X_n; n \in \mathbb{N}\}$ une sous-martingale par rapport à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, et soient S et T deux temps d'arrêt par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$, qui sont bornés presque sûrement et vérifient $S \leq T$ presque sûrement.

Alors,

$$\mathbf{E}[X_S] \leq \mathbf{E}[X_T].$$

Preuve On sait que X_S et X_T sont dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ (par un argument utilisé précédemment, par exemple dans la preuve du théorème d'arrêt).

On définit alors le processus $\{H_n; n \geq 1\}$ par

$$\forall n \geq 0; \quad H_n = \mathbb{1}_{\{S < n \leq T\}} = \mathbb{1}_{\{S \leq n-1\}} - \mathbb{1}_{\{T \leq n-1\}}.$$

Le processus H est alors borné et prévisible par rapport à $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et on peut considérer le processus $(H \bullet X)$ défini par $(H \bullet X)_0 = 0$ et

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

Pour tout $n \in \mathbb{N}$ tel que $S \leq T < n$, on a

$$(H \bullet X)_n = \sum_{k=1}^n \mathbb{1}_{\{S \leq k-1\}}(X_k - X_{k-1}) - \sum_{k=1}^n \mathbb{1}_{\{T \leq k-1\}}(X_k - X_{k-1}) = X_T - X_S$$

et $\mathbf{E}[(H \bullet X)_n] \geq 0$ car $H \bullet X$ est une sous-martingale issue de 0. \square

On aurait également pu utiliser la décomposition de Doob de la sous-martingale $\{X_n; n \in \mathbb{N}\}$ et appliquer le théorème d'arrêt à la partie martingale de la décomposition.

Théorème 2.4.2 (*Première inégalité de Doob*)

Soit $X = \{X_n; n \geq 0\}$ une sous-martingale et soit $a > 0$. Alors pour tout $n \geq 0$,

$$a \mathbf{P} \left(\sup_{k \leq n} X_k \geq a \right) \leq \mathbf{E} \left[X_n \cdot \mathbb{1}_{\{\sup_{k \leq n} X_k \geq a\}} \right] \leq \mathbf{E}[(X_n)_+].$$

Si X est une martingale, alors on a

$$a \mathbf{P} \left(\sup_{k \leq n} |X_k| \geq a \right) \leq \mathbf{E}[|X_n|].$$

Preuve On considère le temps d'arrêt $T = \inf\{n \geq 0 : X_n \geq a\}$.

On a alors

$$A = \left\{ \sup_{0 \leq k \leq n} X_k \geq a \right\} = \{T \leq n\}$$

et la minoration

$$X_{T \wedge n} \geq a \mathbb{1}_A + X_n \mathbb{1}_{A^c},$$

car sur A , $X_{T \wedge n} = X_T \geq a$ et sur A^c , $X_{T \wedge n} = X_n$.

En utilisant le lemme précédent, on déduit

$$T \wedge n \leq n \Rightarrow \mathbf{E}[X_{T \wedge n}] \leq \mathbf{E}[X_n].$$

Des deux inégalités précédentes, on déduit

$$a \mathbf{P}(A) + \mathbf{E}[X_n \cdot \mathbb{1}_{A^c}] \leq \mathbf{E}[X_n]$$

d'où

$$a \mathbf{P} \left(\sup_{0 \leq k \leq n} X_k \geq a \right) \leq \mathbf{E}[X_n \cdot \mathbb{1}_A].$$

Pour la deuxième partie de l'inégalité, il suffit de remarquer que

$$\mathbf{E}[X_n \cdot \mathbb{1}_A] \leq \mathbf{E}[(X_n \cdot \mathbb{1}_A)_+] \leq \mathbf{E}[(X_n)_+].$$

La deuxième partie du théorème découle du fait que si $\{X_n; n \geq 0\}$ est une martingale, alors $\{|X_n|; n \geq 0\}$ est une sous-martingale. \square

Théorème 2.4.3 (Inégalité L^p de martingale de Doob)

Soit $M = \{M_n; n \geq 0\}$ une sous-martingale positive ou une martingale. Soit $p > 1$. Alors, pour tout $n \geq 0$, on a

$$\left(\mathbf{E} \left[\left(\sup_{k \leq n} |M_k| \right)^p \right] \right)^{\frac{1}{p}} \leq \frac{p}{p-1} \left(\mathbf{E}[|M_n|^p] \right)^{\frac{1}{p}},$$

qu'on peut écrire, en posant $M_n^* = \sup_{k \leq n} |M_k|$,

$$\|M_n^*\|_{L^p} \leq \frac{p}{p-1} \|M_n\|_{L^p}.$$

Preuve • On montre d'abord que si M est une sous-martingale positive, on a

$$\mathbf{E} \left[\left(\sup_{k \leq n} M_k \right)^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbf{E}[(M_n)^p].$$

Le théorème en découlera en remarquant que si M est une martingale, alors $|M|$ est une sous-martingale positive.

On peut supposer $\mathbf{E}[(M_n)^p] < \infty$ (sinon, le résultat est évident).

• Le processus $\{(M_n)^p; n \in \mathbb{N}\}$ est une sous-martingale (car $x \mapsto x^p$ est convexe et croissante) donc pour tout $k \leq n$, $\mathbf{E}[(M_k)^p] \leq \mathbf{E}[(M_n)^p]$ (car la suite $(\mathbf{E}[(M_n)^p])_{n \in \mathbb{N}}$ est croissante).

En écrivant

$$(M_n^*)^p = \sup_{k \leq n} (M_k)^p \leq \sum_{k \leq n} (M_k)^p,$$

on en déduit que $\mathbf{E}[(M_n^*)^p] \leq (n+1) \mathbf{E}[(M_n)^p] < \infty$.

• La première inégalité de Doob (Théorème 2.4.2) implique

$$\forall a > 0, \quad a \mathbf{P}(M_n^* \geq a) \leq \mathbf{E} [X_n \cdot \mathbb{1}_{\{M_n^* \geq a\}}].$$

En multipliant l'inégalité par a^{p-2} et en intégrant par rapport à a pour la mesure de Lebesgue λ sur $[0, +\infty[$, le membre de gauche devient

$$\int_0^{+\infty} a^{p-1} \mathbf{P}(M_n^* \geq a) \lambda(da) = \mathbf{E} \left[\int_0^{M_n^*} a^{p-1} \lambda(da) \right] = \frac{1}{p} \mathbf{E}[(M_n^*)^p]$$

en utilisant le théorème de Tonelli-Fubini pour intervertir les deux intégrations.

Pour le membre de droite, on obtient

$$\int_0^{+\infty} a^{p-2} \mathbf{E} [M_n \cdot \mathbb{1}_{\{M_n^* \geq a\}}] \lambda(da) = \mathbf{E} \left[M_n \int_0^{M_n^*} a^{p-2} \lambda(da) \right] = \frac{1}{p-1} \mathbf{E} [M_n (M_n^*)^{p-1}].$$

Or, d'après l'inégalité de Hölder, en considérant le conjugué q de p (c'est-à-dire tel que $\frac{1}{p} + \frac{1}{q} = 1$, soit $q = p/(p-1)$), on a

$$\mathbf{E} [M_n (M_n^*)^{p-1}] \leq \mathbf{E} [(M_n)^p]^{1/p} \mathbf{E} [(M_n^*)^p]^{1-1/p}.$$

On en déduit

$$\frac{1}{p} \mathbf{E} [(M_n^*)^p] \leq \frac{1}{p-1} \mathbf{E} [(M_n)^p]^{1/p} \mathbf{E} [(M_n^*)^p]^{1-1/p}$$

qui implique l'inégalité recherchée. \square

Inégalité du nombre de montées (de Doob) L'inégalité suivante concerne le nombre de montées d'une martingale, notion introduite par Doob. Celle-ci s'avèrera très utile dans le chapitre suivant pour montrer le résultat de convergence presque sure des martingales.

Etant donnée une sous-martingale $X = \{X_n; n \geq 0\}$ à valeurs dans \mathbb{R} , le nombre de montées de X le long de l'intervalle $[a, b]$ avant le temps n est défini comme : le plus grand nombre de couples (S_k, T_k) ($1 \leq k \leq N$) tels que

$$\begin{cases} 0 \leq S_1 < T_1 < S_2 < T_2 < \dots < S_N < T_N \leq n \\ \forall k = 1, \dots, N, \quad X_{S_k} \leq a \quad \text{et} \quad X_{T_k} \geq b. \end{cases}$$

Les entiers S_k et T_k peuvent être définis de la manière suivante : on pose $T_0 = 0$ et par récurrence, pour tout $k \geq 1$,

$$S_k = \min\{l > T_{k-1} : X_l \leq a\}, \quad T_k = \min\{l > S_k : X_l \geq b\}.$$

Les variables aléatoires S_k et T_k ($\forall k \geq 1$) sont des temps d'arrêt. En effet, on peut écrire par exemple pour T_k ,

$$\{T_k \leq n\} = \bigcup_{0 \leq m_1 < n_1 < \dots < m_k < n_k \leq n} \{X_{m_1} \leq a, X_{n_1} \geq b, \dots, X_{m_k} \leq a, X_{n_k} \geq b\},$$

ce qui montre que $\{T_k \leq n\} \in \mathcal{F}_n$.

Le nombre de montées de X le long de l'intervalle $[a, b]$ avant le temps n est alors défini comme

$$U_n([a, b]) = \max\{k : T_k \leq n\} = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \leq n\}}.$$

Comme les variables aléatoires T_k sont des temps d'arrêt, la variable aléatoire $U_n([a, b])$ est \mathcal{F}_n -mesurable.

Théorème 2.4.4 (Inégalité du nombre de montées, de Doob)

Soient $X = \{X_n; n \geq 0\}$ une sous-martingale, $a < b$ et $U_n([a, b])$ le nombre de montées de X le long de l'intervalle $[a, b]$ avant le temps n . Alors,

$$\mathbf{E}[U_n([a, b])] \leq \frac{1}{b-a} \mathbf{E}[(X_n - a)_+]$$

où $(X_n - a)_+ = \max(X_n - a, 0)$.

Preuve Posons $Y_n = (X_n - a)_+$. Comme la fonction $x \mapsto (x - a)_+$ est convexe, le processus $\{Y_n; n \geq 0\}$ est également une sous-martingale.

On définit le processus $\{H_n; n \geq 0\}$ par

$$\forall n \geq 0; \quad H_n = \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k < n \leq T_k\}} \leq 1.$$

Le processus $\{H_n; n \geq 0\}$ est prévisible car pour tout $k \geq 1$, on a

$$\{S_k < n \leq T_k\} = \{S_k \leq n-1\} \setminus \{T_k \leq n-1\} \in \mathcal{F}_{n-1}.$$

On considère alors le processus $(H \bullet Y)$ défini par $(H \bullet Y)_0 = 0$ et

$$(H \bullet Y)_n = \sum_{k=1}^n H_k (Y_k - Y_{k-1}) = \sum_{l=1}^{+\infty} \sum_{k=1}^n \mathbb{1}_{\{S_l < k \leq T_l\}} (Y_k - Y_{k-1}).$$

Or, on a

$$\sum_{k=1}^n \mathbb{1}_{\{S_l < k \leq T_l\}} (Y_k - Y_{k-1}) = \begin{cases} Y_{T_l} - Y_{S_l} & \text{si } S_l < T_l \leq n \quad (\text{i.e. } l \leq U_n) \\ Y_n - Y_{S_l} & \text{si } S_l < n < T_l \quad (\text{i.e. } l = 1 + U_n \text{ et } S_l < +\infty) \\ 0 & \text{sinon.} \end{cases}$$

On en déduit

$$(H \bullet Y)_n = \sum_{l=1}^{U_n} (Y_{T_l} - Y_{S_l}) + (Y_n - Y_{S_{1+U_n}}) \cdot \mathbb{1}_{\{S_{1+U_n} < \infty\}}$$

pour tout $n \geq 1$.

Sur $\{S_{U_n+1} < \infty\}$, on a $Y_{S_{U_n+1}} = 0$ et $Y_n \geq 0$. En conséquence, on a la minoration

$$(H \bullet Y)_n \geq \sum_{k=1}^{U_n} (Y_{T_k} - Y_{S_k}) \geq (b-a) U_n.$$

En prenant l'espérance, on déduit

$$\mathbf{E}[(H \bullet Y)_n] \geq (b-a) \mathbf{E}[U_n].$$

On considère alors le processus $\{K_n; n \geq 0\}$ défini par $K_n = 1 - H_n$ pour tout $n \geq 0$. Le processus K est évidemment borné et prévisible. Donc $(K \bullet Y)$ est une sous-martingale. On en déduit $E[(K \bullet Y)_n] \geq E[(K \bullet Y)_0] = 0$.

De l'égalité

$$(K \bullet Y)_n + (H \bullet Y)_n = ((K + H) \bullet Y)_n = Y_n - Y_0,$$

on tire

$$(b - a) \mathbf{E}[U_n] \leq \mathbf{E}[(H \bullet Y)_n] \leq \mathbf{E}[(K \bullet Y)_n + (H \bullet Y)_n] = \mathbf{E}[Y_n - Y_0].$$

Le résultat suit en majorant $\mathbf{E}[Y_n - Y_0]$ par $\mathbf{E}[Y_n]$. \square

Chapitre 3

Convergence d'une martingale

Dans ce chapitre, on s'intéresse à la convergence de la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ lorsque $n \rightarrow \infty$, dans le cas où le processus $\{X_n; n \in \mathbb{N}\}$ est une martingale.

Les inégalités de martingales obtenues au chapitre précédent, permettent d'étudier les différents modes de convergence : presque sûre, dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ avec $p > 1$, puis dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

3.1 Convergence presque sûre d'une martingale

Dans ce 1er résultat de convergence, la majoration du nombre de montées d'une martingale (Théorème 2.4.4) permet de montrer la convergence presque sûre d'une martingale dont la partie positive est bornée dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Théorème 3.1.1

Soit $\{X_n; n \in \mathbb{N}\}$ une sous-martingale telle que

$$\sup_{n \in \mathbb{N}} \mathbf{E}[(X_n)_+] < +\infty. \quad (3.1)$$

Alors la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement lorsque $n \rightarrow \infty$.

De plus, sa limite X_∞ vérifie $\mathbf{E}[|X_\infty|] < +\infty$.

On suit ici la preuve originale donnée par Doob. On pourra également consulter les références [6] p. 225 et [13] p. 171.

Preuve • Soient $a, b \in \mathbb{Q}$ tels que $a < b$. D'après le Théorème 2.4.4, on a pour tout $n \geq 1$,

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

Or, on a

$$(X_n - a)_+ \leq |a| + (X_n)_+$$

il suffit de distinguer les cas :

- $X_n \leq a$: On a alors $(X_n - a)_+ = 0$;
- $X_n > a$ et $X_n > 0$: On a alors $(X_n)_+ = X_n$ et $(X_n - a)_+ = X_n - a \leq |a| + X_n$;
- $a < X_n \leq 0$: On a alors $(X_n)_+ = 0$ et $(X_n - a)_+ = X_n - a \leq |a|$.

On en déduit alors les inégalités

$$(b - a) \mathbf{E}[U_n([a, b])] \leq |a| + \mathbf{E}[(X_n)_+] \leq |a| + \sup_{k \in \mathbb{N}} \mathbf{E}[(X_k)_+].$$

La suite $(U_n([a, b]))_{n \in \mathbb{N}}$ est croissante, donc on peut faire tendre n vers ∞ dans l'inégalité précédente. On obtient

$$(b - a) \mathbf{E}[U_\infty([a, b])] < \infty$$

ce qui entraîne $\mathbf{P}(U_\infty([a, b]) < \infty) = 1$ (d'après l'inégalité de Markov).

- Ce résultat implique que X_n croît le long de l'intervalle $[a, b]$ un nombre fini de fois presque sûrement.

En posant

$$\Lambda_{a,b} = \{\liminf_{n \rightarrow \infty} X_n \leq a; \limsup_{n \rightarrow \infty} X_n \geq b\},$$

on a alors $\mathbf{P}(\Lambda_{a,b}) = 0$.

On pose alors

$$\Lambda = \bigcup_{\substack{a < b \\ a, b \in \mathbb{Q}}} \Lambda_{a,b}$$

et on a $\mathbf{P}(\Lambda) = 0$, car l'ensemble des couples de rationnels est dénombrable.

En remarquant que

$$\Lambda = \{\liminf_{n \rightarrow \infty} X_n < \limsup_{n \rightarrow \infty} X_n\},$$

on déduit que

$$\mathbf{P}(\liminf_{n \rightarrow \infty} X_n = \limsup_{n \rightarrow \infty} X_n) = 1.$$

On a donc : $\lim_{n \rightarrow \infty} X_n$ existe presque sûrement.

- Il reste à montrer que cette limite, notée X_∞ , est finie. Pour cela, on écrit

$$\mathbf{E}[|X_n|] = \mathbf{E}[(X_n)_+] + \mathbf{E}[(X_n)_-] = 2\mathbf{E}[(X_n)_+] - \mathbf{E}[X_n].$$

Comme X est une sous-martingale, on a $\mathbf{E}[X_n] \geq \mathbf{E}[X_0]$ et

$$\mathbf{E}[|X_n|] \leq 2\mathbf{E}[(X_n)_+] - \mathbf{E}[X_0],$$

ce qui implique

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

D'après le Lemme de Fatou, on trouve

$$\mathbf{E}[|X_\infty|] \leq \liminf_{n \rightarrow \infty} \mathbf{E}[|X_n|] \leq \sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|] < \infty,$$

ce qui implique $X_\infty < +\infty$ presque sûrement. \square

Au cours de la preuve, on a montré que la condition (3.1) alliée au fait que $\{X_n; n \in \mathbb{N}\}$ est une sous-martingale implique

$$\sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|] < +\infty, \quad (3.2)$$

c'est-à-dire que X est borné dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Inversement, si $\sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|] < +\infty$ alors la condition (3.1) est vérifiée puisque $(X_n)_+ \leq |X_n|$.

Finalement, la condition (3.1) peut être remplacée par (3.2) dans l'énoncé du théorème.

Remarque 3.1.2. *Attention, le Théorème 3.1.1 est un résultat de convergence presque sûre d'une martingale $\{X_n; n \geq 0\}$, qui assure le fait que la limite X_∞ est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.*

Il n'est absolument pas dit que la convergence de X_n a lieu dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$!!! Ceci est faux en général (cf. [13] p. 172 exemple (1)).

Une autre preuve du Théorème 3.1.1 est donnée dans [10] (cf. p. 86). Elle est basée sur la décomposition de Doob et un résultat de convergence des martingales dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Le passage de $L^2(\Omega, \mathcal{F}, \mathbf{P})$ à $L^1(\Omega, \mathcal{F}, \mathbf{P})$ est réalisé en arrêtant les martingales, par un procédé déjà utilisé lors de la définition de l'espérance conditionnelle dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Corollaire 3.1.3

Soit $\{X_n; n \in \mathbb{N}\}$ une $(\mathcal{F}_n)_{n \in \mathbb{N}}$ -surmartingale positive.

Alors X_n converge presque sûrement lorsque $n \rightarrow \infty$, vers une variable aléatoire X_∞ qui est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

De plus, pour tout $n \in \mathbb{N}$, on a $X_n \geq \mathbf{E}[X_\infty | \mathcal{F}_n]$ p.s.

Preuve Le processus stochastique $\{X'_n; n \in \mathbb{N}\}$ défini par $X'_n = -X_n$ pour tout $n \geq 0$, est visiblement une sous-martingale. L'hypothèse (3.1) est trivialement vérifiée pour X' , puisque $(X'_n)_+ = 0$ pour tout $n \in \mathbb{N}$. On peut donc appliquer le Théorème 3.1.1 qui implique la convergence p.s. de X'_n et donc de X_n vers une v.a. qui est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

On applique ensuite le Lemme de Fatou pour les espérances conditionnelles

$$X_n \geq \liminf_{m \rightarrow \infty} \mathbf{E}[X_m | \mathcal{F}_n] \geq \mathbf{E}[\liminf_{m \rightarrow \infty} X_m | \mathcal{F}_n] = \mathbf{E}[X_\infty | \mathcal{F}_n] \quad \text{p.s.}$$

\square

Pour terminer cette section, après avoir insisté sur la différence entre convergence presque sûre vers une variable aléatoire qui est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et convergence dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, mentionnons que nous verrons dans la Section 3.3 que les martingales qui convergent dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ sont exactement les martingales $\{M_n; n \in \mathbb{N}\}$ qui s'écrivent comme $M_n = \mathbf{E}[Z \mid \mathcal{F}_n]$ p.s. pour une certaine variable aléatoire $Z \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.

3.2 Martingales bornées dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ ($p > 1$)

L'inégalité L^p de martingales de Doob (Théorème 2.4.3) permet de montrer la convergence dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ d'une martingale bornée dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$.

Théorème 3.2.1

Soit $\{X_n; n \in \mathbb{N}\}$ une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, telle que

$$\sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|^p] < \infty \quad \text{pour un certain } p > 1.$$

Alors, X_n converge presque sûrement et dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ vers une variable aléatoire X_∞ telle que

$$\mathbf{E}[|X_\infty|^p] = \sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|^p].$$

De plus, on a

$$\forall n \in \mathbb{N}, \quad X_n = \mathbf{E}[X_\infty \mid \mathcal{F}_n] \quad \text{p.s.}$$

Preuve • Comme $\{X_n; n \in \mathbb{N}\}$ est bornée dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ avec $p > 1$, elle est également bornée dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$. Grâce au Théorème 3.1.1, on sait que X_n converge presque sûrement vers une variable aléatoire X_∞ .

De plus, l'inégalité sur les martingales L^p (Théorème 2.4.3) implique

$$\forall n \in \mathbb{N}, \quad \mathbf{E}[(X_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}[|X_n|^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{k \in \mathbb{N}} \mathbf{E}[|X_k|^p]$$

en notant $X_n^* = \sup_{k \leq n} |X_k|$.

La suite $(X_n^*)_{n \in \mathbb{N}}$ est croissante et converge donc (pour tout ω) vers $X_\infty^* = \sup_{k \in \mathbb{N}} |X_k|$. En appliquant le théorème de convergence monotone, on obtient grâce à l'inégalité précédente

$$\mathbf{E}[(X_\infty^*)^p] \leq \left(\frac{p}{p-1}\right)^p \sup_{k \in \mathbb{N}} \mathbf{E}[|X_k|^p] < \infty,$$

ce qui implique $X_\infty^* \in L^p(\Omega, \mathcal{F}, \mathbf{P})$.

• Comme toutes les $|X_n|$ sont majorées par X_∞^* , on peut appliquer le théorème de convergence dominée pour montrer que X_n converge dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ vers X_∞

$$\mathbf{E}[|X_n - X_\infty|^p] \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty.$$

- Comme l'application $x \mapsto |x|^p$ est convexe, le processus $\{|X_n|^p; n \in \mathbb{N}\}$ est une sous-martingale. On en déduit que la suite $\mathbf{E}[|X_n|^p]$ est croissante.

Par convergence monotone, on obtient donc

$$\mathbf{E}[|X_\infty|^p] = \lim_{n \rightarrow \infty} \mathbf{E}[|X_n|^p] = \sup_{n \in \mathbb{N}} \mathbf{E}[|X_n|^p].$$

- Pour montrer la dernière assertion du théorème, on remarque que la convergence dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ implique celle dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ (par l'inégalité de Hölder).

Pour tout événement $A \in \mathcal{F}_n$, le théorème de convergence dominée (par exemple) implique

$$\mathbf{E}[(X_n - X_\infty) \cdot \mathbb{1}_A] = \lim_{k \rightarrow \infty} \mathbf{E}[(X_n - X_k) \cdot \mathbb{1}_A].$$

Or, pour tout $k \geq n + 1$, $X_n = \mathbf{E}[X_k | \mathcal{F}_n] \Leftrightarrow \mathbf{E}[X_n - X_k | \mathcal{F}_n] = 0$.

On a donc $\mathbf{E}[X_n \cdot \mathbb{1}_A] = \mathbf{E}[X_\infty \cdot \mathbb{1}_A]$ pour tout $A \in \mathcal{F}_n$, c'est-à-dire $X_n = \mathbf{E}[X_\infty | \mathcal{F}_n]$ p.s. \square

3.3 Martingales uniformément intégrables

Définition 3.3.1

Une martingale $\{M_n; n \in \mathbb{N}\}$ est dite régulière ou fermée s'il existe une variable aléatoire $Z \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ telle que pour tout $n \in \mathbb{N}$, $M_n = \mathbf{E}[Z | \mathcal{F}_n]$ p.s.

Le but de cette section est d'étudier à quelle condition une martingale donnée est une martingale régulière. La réponse à cette question est étroitement liée à la notion d'*uniforme intégrabilité*.

Définition 3.3.2

Une famille de variables aléatoires $(X_i)_{i \in I}$ dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ est dite uniformément intégrable (u.i.) si

$$\lim_{a \rightarrow +\infty} \left(\sup_{i \in I} \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \right) = 0.$$

Le constat immédiat est que la propriété d'*uniforme intégrabilité* implique d'être *borné dans* $L^1(\Omega, \mathcal{F}, \mathbf{P})$. En effet, si $(X_i)_{i \in I}$ est une famille de variables aléatoires uniformément intégrable, on peut choisir a assez grand pour que

$$\sup_{i \in I} \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \leq 1.$$

On en déduit la majoration

$$\mathbf{E}[|X_i|] = \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| \leq a\}}] + \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \leq a + 1.$$

La réciproque est fautive : une famille bornée dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ n'est pas nécessairement uniformément intégrable.

Proposition 3.3.3

Soit $(X_i)_{i \in I}$ une famille de v.a. dans $(\Omega, \mathcal{F}, \mathbf{P})$.

- (i) Si $(X_i)_{i \in I}$ est bornée dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ pour un certain $p > 1$, alors elle est uniformément intégrable.
- (ii) S'il existe une variable aléatoire positive Z telle que $|X_i| \leq Z$ p.s. pour tout $i \in I$ et $\mathbf{E}[Z] < +\infty$, alors la famille $(X_i)_{i \in I}$ est uniformément intégrable.

Preuve

- (i) Pour tout $a \geq 1$, on a

$$\forall x > a, \quad x \leq x^p$$

et on pose $\varphi(a) = \sup_{x > a} \left(\frac{x}{x^p} \right)$. On a $\lim_{a \rightarrow +\infty} \varphi(a) = 0$.

On en déduit pour tout $i \in I$,

$$\mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \leq \varphi(a) \mathbf{E} [|X_i|^p \mathbb{1}_{\{|X_i| > a\}}] \leq \varphi(a) \mathbf{E} [|X_i|^p].$$

Comme $(X_i)_{i \in I}$ est bornée dans $L^p(\Omega, \mathcal{F}, \mathbf{P})$ par $\sup_{j \in I} \mathbf{E} [|X_j|^p]$, l'inégalité

$$\mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \leq \varphi(a) \sup_{j \in I} \mathbf{E} [|X_j|^p],$$

permet de conclure que la famille $(X_i)_{i \in I}$ est uniformément intégrable.

- (ii) Pour tout $i \in I$, comme $|X_i| \leq Z$ p.s., on a

$$\mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \leq \mathbf{E} [|Z| \mathbb{1}_{\{|Z| > a\}}].$$

Le théorème de convergence dominée permet alors de conclure que la famille $(X_i)_{i \in I}$ est uniformément intégrable.

□

Le résultat suivant fournit une caractérisation de l'uniforme intégrabilité en terme d'équi-continuité.

Proposition 3.3.4

Soit $(X_i)_{i \in I}$ une famille bornée dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$. Alors, les deux assertions suivantes sont équivalentes :

- (i) La famille $(X_i)_{i \in I}$ est uniformément intégrable.
- (ii) Pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout événement $A \in \mathcal{F}$ de probabilité $\mathbf{P}(A) < \delta$, on a

$$\forall i \in I; \quad \mathbf{E} [|X_i| \mathbb{1}_A] < \epsilon.$$

Preuve Pour montrer que (i) \Rightarrow (ii), on considère $\epsilon > 0$. L'uniforme intégrabilité de la famille $(X_i)_{i \in I}$ implique qu'il existe $a > 0$ tel que

$$\sup_{i \in I} \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] < \frac{\epsilon}{2}.$$

On pose $\delta = \frac{\epsilon}{2a}$. Pour tout $A \in \mathcal{F}$ tel que $\mathbf{P}(A) < \delta$, on a

$$\begin{aligned} \mathbf{E} [|X_i| \mathbb{1}_A] &= \mathbf{E} [|X_i| \mathbb{1}_{A \cap \{|X_i| \leq a\}}] + \mathbf{E} [|X_i| \mathbb{1}_{A \cap \{|X_i| > a\}}] \\ &\leq a \mathbf{E} [\mathbb{1}_{A \cap \{|X_i| \leq a\}}] + \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] \\ &\leq a \mathbf{P}(A) + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Inversement, pour montrer que (ii) \Rightarrow (i), on pose $C = \sup_{i \in I} \mathbf{E} [|X_i|]$. L'inégalité de Markov implique que

$$\forall i \in I, \quad \mathbf{P}(|X_i| > a) \leq \frac{\mathbf{E} [|X_i|]}{a} \leq \frac{C}{a}.$$

On considère alors $\epsilon > 0$ et $\delta > 0$ (comme dans l'énoncé du (ii)). Pour tout $a > C/\delta$, on a $\frac{C}{a} < \delta$, puis en appliquant (ii) à $A = \{|X_i| > a\}$,

$$\forall i \in I, \quad \mathbf{E} [|X_i| \mathbb{1}_{\{|X_i| > a\}}] < \epsilon.$$

On conclut alors que la famille $(X_i)_{i \in I}$ est uniformément intégrable. \square

Proposition 3.3.5

Soient $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ et $(\mathcal{G}_i)_{i \in I}$ une famille de sous-tribus de \mathcal{F} . Alors, la famille des $X_i = \mathbf{E}[X \mid \mathcal{G}_i]$ ($i \in I$) est uniformément intégrable.

Preuve Par le théorème de convergence dominée, le singleton $\{X\}$ est uniformément intégrable. Ainsi, d'après la proposition précédente, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout $A \in \mathcal{F}$,

$$\mathbf{P}(A) < \delta \Rightarrow \mathbf{E} [|X| \mathbb{1}_A] < \epsilon. \quad (3.3)$$

En utilisant l'inégalité de Jensen, on vérifie que

$$\forall i \in I, \quad |X_i| = |\mathbf{E}[X \mid \mathcal{G}_i]| \leq \mathbf{E} [|X| \mid \mathcal{G}_i], \quad (3.4)$$

ce qui entraîne $\mathbf{E} [|X_i|] \leq \mathbf{E} [|X|]$ pour tout $i \in I$.

Par l'inégalité de Markov, pour tout $a > 0$,

$$\forall i \in I, \quad \mathbf{P}(|X_i| > a) \leq \frac{\mathbf{E} [|X_i|]}{a} \leq \frac{\mathbf{E} [|X|]}{a}.$$

Pour tout $a > \mathbf{E}[|X|]/\delta$, en utilisant l'inégalité (3.4), on a

$$\forall i \in I, \quad \mathbf{E}[|X_i| \mathbb{1}_{\{|X_i|>a\}}] \leq \mathbf{E}[\mathbf{E}(|X| \mid \mathcal{G}_i) \mathbb{1}_{\{|X_i|>a\}}] = \mathbf{E}[|X| \mathbb{1}_{\{|X_i|>a\}}].$$

Or, comme $\mathbf{P}(|X_i| > a) < \delta$, l'assertion (3.3) entraîne $\mathbf{E}[|X| \mathbb{1}_{\{|X_i|>a\}}] < \epsilon$.

On en déduit

$$\forall i \in I, \quad \mathbf{E}[|X_i| \mathbb{1}_{\{|X_i|>a\}}] < \epsilon,$$

puis que la famille $(X_i)_{i \in I}$ est uniformément intégrable. \square

Alors qu'on a insisté sur la différence entre convergence p.s. vers une v.a. qui est dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et convergence dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, le résultat suivant permet de caractériser la convergence dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ pour les martingales.

D'autre part, ce résultat principal de cette section répond à la question de la caractérisation des martingales régulières.

Théorème 3.3.6

Soit $\{M_n; n \in \mathbb{N}\}$ une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. Alors les trois assertions suivantes sont équivalentes :

- (i) *La famille $(M_n)_{n \in \mathbb{N}}$ est uniformément intégrable.*
- (ii) *M_n converge presque sûrement et dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ lorsque $n \rightarrow \infty$ vers une v.a. M_∞ .*
- (iii) *La martingale $\{M_n; n \in \mathbb{N}\}$ est régulière.*

Lorsque ces conditions sont vérifiées, on a pour tout $n \in \mathbb{N}$, $M_n = \mathbf{E}[M_\infty \mid \mathcal{F}_n]$ p.s.

Preuve • Pour montrer que (i) \Rightarrow (ii), on remarque que la famille $(M_n)_{n \in \mathbb{N}}$ est bornée dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ puisqu'elle est uniformément intégrable.

D'après le théorème de convergence presque sûre des sous-martingales (Théorème 3.1.1), on en déduit que la martingale $\{M_n; n \in \mathbb{N}\}$ converge presque sûrement vers une variable aléatoire $M_\infty \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Pour montrer que la convergence a lieu dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, on fixe $\epsilon > 0$ et on considère

$$\begin{aligned} \mathbf{E}[|M_n - M_\infty|] &= \mathbf{E}[|M_n - M_\infty| \mathbb{1}_{\{|M_n - M_\infty| \leq \epsilon\}}] + \mathbf{E}[|M_n - M_\infty| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}] \\ &\leq \epsilon + \mathbf{E}[|M_\infty| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}] + \mathbf{E}[|M_n| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}], \end{aligned}$$

en utilisant l'inégalité triangulaire $|M_n - M_\infty| \leq |M_n| + |M_\infty|$.

Or, comme $M_\infty \in L^1(\Omega, \mathcal{F}, \mathbf{P})$, le théorème de convergence dominée implique que $\mathbf{E}[|M_\infty| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}]$ tend vers 0 lorsque $n \rightarrow \infty$. Donc il existe $n_1 \in \mathbb{N}$ tel que

$$\forall n \geq n_1, \quad \mathbf{E}[|M_\infty| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}] < \epsilon.$$

Pour majorer l'autre terme, on utilise la caractérisation de l'uniforme intégrabilité de la Proposition 3.3.4. On considère alors le $\delta > 0$ de la Proposition 3.3.4, relatif à ϵ .

Comme $\mathbf{P}(|M_n - M_\infty| > \epsilon)$ tend vers 0 lorsque $n \rightarrow \infty$ (M_n converge vers M_∞ en probabilité), il existe $n_2 \in \mathbb{N}$ tel que

$$\forall n \geq n_2, \quad \mathbf{P}(|M_n - M_\infty| > \epsilon) < \delta.$$

On a alors (Proposition 3.3.4)

$$\forall n \geq n_2, \quad \mathbf{E} [|M_n| \mathbb{1}_{\{|M_n - M_\infty| > \epsilon\}}] < \epsilon.$$

En rassemblant ces résultats, on conclut

$$\forall n \geq n_1 \vee n_2, \quad \mathbf{E} [|M_n - M_\infty|] < 3\epsilon,$$

puis que M_n converge dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ vers M_∞ .

- Pour montrer que (ii) \Rightarrow (iii), la propriété de martingales peut s'écrire : pour tout $n \in \mathbb{N}$,

$$\forall A \in \mathcal{F}_n, \forall k \in \mathbb{N}, \quad \mathbf{E}[M_n \cdot \mathbb{1}_A] = \mathbf{E}[M_{n+k} \cdot \mathbb{1}_A],$$

car $M_n = \mathbf{E}[M_{n+k} \mid \mathcal{F}_n]$ p.s.

Comme M_{n+k} converge dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ vers M_∞ lorsque $k \rightarrow \infty$, on a

$$\mathbf{E}[M_{n+k} \cdot \mathbb{1}_A] \rightarrow \mathbf{E}[M_\infty \cdot \mathbb{1}_A] \quad \text{lorsque } k \rightarrow \infty.$$

Donc, en faisant tendre k vers ∞ , on obtient

$$\forall A \in \mathcal{F}_n, \quad \mathbf{E}[M_n \cdot \mathbb{1}_A] = \mathbf{E}[M_\infty \cdot \mathbb{1}_A],$$

i.e. $M_n = \mathbf{E}[M_\infty \mid \mathcal{F}_n]$ p.s. La martingale $\{M_n; n \in \mathbb{N}\}$ est donc régulière.

- L'implication (iii) \Rightarrow (i) est une conséquence directe de la Proposition 3.3.5. \square

Corollaire 3.3.7

Soient Z une variable aléatoire dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ et $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration de $(\Omega, \mathcal{F}, \mathbf{P})$. La martingale $\{M_n; n \in \mathbb{N}\}$ définie par $M_n = \mathbf{E}[Z \mid \mathcal{F}_n]$ converge presque sûrement et dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$ vers la variable aléatoire $M_\infty = \mathbf{E}[Z \mid \mathcal{F}_\infty]$ où $\mathcal{F}_\infty = \bigvee_{n=1}^{\infty} \mathcal{F}_n$.

Preuve Comme le Théorème 3.3.6 s'applique à la martingale régulière $\{M_n; n \in \mathbb{N}\}$, il ne reste qu'à démontrer que $M_\infty = \mathbf{E}[Z \mid \mathcal{F}_\infty]$ p.s.

Pour tout $n \in \mathbb{N}$, M_n est \mathcal{F}_∞ -mesurable, donc sa limite p.s. M_∞ l'est également.

Pour tout $n \in \mathbb{N}$ et pour tout $A \in \mathcal{F}_n$, on a $\mathbf{E}[Z \cdot \mathbb{1}_A] = \mathbf{E}[M_n \cdot \mathbb{1}_A]$ par définition de M_n . D'autre part, comme M_n converge vers M_∞ dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, $\mathbf{E}[M_n \cdot \mathbb{1}_A] = \mathbf{E}[M_\infty \cdot \mathbb{1}_A]$.

La relation $\mathbf{E}[Z \cdot \mathbb{1}_A] = \mathbf{E}[M_\infty \cdot \mathbb{1}_A]$ pour tout $A \in \mathcal{F}_n$ est vraie pour tout $n \in \mathbb{N}$, donc elle reste vraie pour tout $A \in \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right) = \mathcal{F}_\infty$ (argument de classe monotone).

Ainsi, on conclut que $M_\infty = \mathbf{E}[Z \mid \mathcal{F}_\infty]$ p.s. \square

Deuxième partie

Chaînes de Markov à temps discret (3 séances)

Chapitre 4

Chaînes de Markov

A l'instar des martingales, les chaînes de Markov revêtent une importance considérable dans la théorie générale des processus stochastiques.

Exemple Marche aléatoire dans \mathbb{Z}^2 .

4.1 Matrices de transition et chaînes de Markov

Définition 4.1.1

Un processus stochastique $X = \{X_n; n \in \mathbb{N}\}$ de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ à valeurs (E, \mathcal{E}) est une chaîne de Markov par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si

- (i) Pour tout $n \in \mathbb{N}$, X_n est \mathcal{F}_n -mesurable;
- (ii) Pour tout $B \in \mathcal{E}$ et pour tout $n \in \mathbb{N}$,

$$\mathbf{P}(X_{n+1} \in B \mid \mathcal{F}_n) = \mathbf{P}(X_{n+1} \in B \mid X_n) \quad p.s.$$

Dans le cas où E est au plus dénombrable, la définition d'une chaîne de Markov est étroitement liée à celle de matrice de transition. Dans la suite de cette section, on supposera toujours que l'espace E est fini ou dénombrable.

Définition 4.1.2

On appelle matrice de transition (ou matrice stochastique) toute famille de réels $(Q(i, j))_{i, j \in E}$ vérifiant les deux conditions suivantes :

- (i) Pour tous $i, j \in E$, $0 \leq Q(i, j) \leq 1$;
- (ii) Pour tout $i \in E$, $\sum_{j \in E} Q(i, j) = 1$.

Rappelons que lorsque E est un espace discret, on peut écrire

$$\mathbf{P}(X_{n+1} \in B \mid X_n) = h(X_n), \tag{4.1}$$

où $h : E \rightarrow \mathbb{R}_+$ est l'application mesurable définie par $h(i) = \mathbf{P}(X_{n+1} \in B \mid X_n = i)$ pour tout $i \in E$.

En utilisant la caractérisation d'une mesure de probabilité $\mathbf{P}(\bullet \mid X_n = i)$ sur un espace discret, la définition d'une chaîne de Markov sur un espace E au plus dénombrable, est équivalente à celle d'une loi initiale μ et d'une matrice de transition $(Q(i, j))_{i, j \in E}$ telles que

(i) Pour tout $B \in \mathcal{E}$,

$$\mathbf{P}(X_0 \in B) = \mu(B);$$

(ii) et

$$\forall n \in \mathbb{N}; \quad \mathbf{P}(X_{n+1} = j \mid \mathcal{F}_n) = Q(X_n, j). \quad (4.2)$$

Autrement dit, la valeur de X_{n+1} ne dépend que de l'état présent X_n et non du passé. Ce deuxième point de la définition est appelé *propriété de Markov*.

Remarque 4.1.3. La fonction $Q(i, \bullet)$ donnant la loi conditionnelle de X_{n+1} sachant que $X_n = i$ ne dépend pas de l'entier n : on dit que la chaîne de Markov X est homogène. Il est possible de généraliser la définition des chaînes de Markov, en écrivant

$$\forall i, j \in E; \quad \mathbf{P}(X_{n+1} = j \mid X_n = i) = Q_n(i, j),$$

où le mécanisme de transition entre les instants n et $n + 1$ dépend de n . On parle alors de chaînes de Markov inhomogènes.

Dans la suite de cette section, on considère le cas particulier où la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ est la filtration naturelle de $\{X_n; n \in \mathbb{N}\}$: $\mathcal{F}_n = \sigma(X_k; 0 \leq k \leq n)$, pour tout $n \in \mathbb{N}$.

Proposition 4.1.4

Le processus stochastique $\{X_n; n \in \mathbb{N}\}$ de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ à valeurs dans (E, \mathcal{E}) est une chaîne de Markov (homogène) par rapport à sa filtration naturelle si et seulement s'il existe une mesure de probabilité μ sur (E, \mathcal{E}) et une matrice de transition $Q(i, j)_{i, j \in E}$ telles que

(i) Pour tout $B \in \mathcal{E}$, $\mathbf{P}(X_0 \in B) = \mu(B)$;

(ii) Pour tout $n \in \mathbb{N}$ et pour tous $i, j, x_0, \dots, x_{n-1} \in E$,

$$\begin{aligned} \mathbf{P}(X_{n+1} = j \mid X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i) \\ = \mathbf{P}(X_{n+1} = j \mid X_n = i) = Q(i, j). \end{aligned} \quad (4.3)$$

μ et Q sont appelées respectivement la loi initiale et la matrice de transition de la chaîne X .

Preuve Si $\{X_n; n \in \mathbb{N}\}$ est une chaîne de Markov, alors la condition (i) est vérifiée pour la loi μ de la variable aléatoire X_0 .

De plus, pour tout $n \in \mathbb{N}$ et pour tous $i, x_0, \dots, x_{n-1} \in E$, on considère l'événement $B = \{X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i\}$. Pour tout $j \in E$, l'expression (4.1) permet d'écrire

$$\begin{aligned} \mathbf{E} [\mathbb{1}_B \mathbb{1}_{\{X_{n+1}=j\}}] &= \mathbf{E} (\mathbb{1}_B \mathbf{E} [\mathbb{1}_{\{X_{n+1}=j\}} \mid X_0, \dots, X_n]) \\ &= \mathbf{E} (\mathbb{1}_B \mathbf{E} [\mathbb{1}_{\{X_{n+1}=j\}} \mid X_n]) \\ &= \mathbf{E} [\mathbb{1}_B h(X_n)] = h(i) \mathbf{P}(B) = Q(i, j) \mathbf{P}(B). \end{aligned}$$

Or,

$$\begin{aligned} \mathbf{E} [\mathbb{1}_B \mathbb{1}_{\{X_{n+1}=j\}}] &= \mathbf{P}(X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i; X_{n+1} = j) \\ &= \mathbf{P}(X_{n+1} = j \mid X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i) \mathbf{P}(B). \end{aligned}$$

On déduit de ces deux égalités

$$\mathbf{P}(X_{n+1} = j \mid X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i) = Q(i, j).$$

Réciproquement, supposons que le processus $\{X_n; n \in \mathbb{N}\}$ vérifie les conditions (i) et (ii) de l'énoncé.

Pour tout événement $B = \{X_0 = x_0; \dots; X_{n-1} = x_{n-1}; X_n = i\}$, en remontant le calcul ci-dessus, on obtient

$$\mathbf{E} [\mathbb{1}_B \mathbb{1}_{\{X_{n+1}=j\}}] = \mathbf{E} (\mathbb{1}_B \mathbf{E} [\mathbb{1}_{\{X_{n+1}=j\}} \mid X_n]),$$

ce qui montre que $\mathbf{E} [\mathbb{1}_{\{X_{n+1}=j\}} \mid X_0, \dots, X_n] = \mathbf{E} [\mathbb{1}_{\{X_{n+1}=j\}} \mid X_n]$ presque sûrement, car les événements de la forme de B engendrent la tribu $\sigma(X_0, \dots, X_n)$. \square

On voit ainsi que l'évolution de la chaîne de Markov après l'instant n ne dépend que de son état présent X_n et non du passé représenté par les valeurs de X_0, \dots, X_{n-1} . C'est l'interprétation de la *propriété de Markov*.

Exemple On considère une marche aléatoire simple $S = \{S_n; n \in \mathbb{N}\}$ symétrique sur \mathbb{Z} , dont on note $X_n = S_n - S_{n-1}$ les accroissements pour $n \geq 1$ et $X_0 = S_0$. Par définition, on a $\mathbf{P}(X_n = 1) = \mathbf{P}(X_n = -1) = 1/2$. Comme dans le chapitre 2, on définit la filtration $(\mathcal{F}_n; n \in \mathbb{N})$ engendrée par X_0, X_1, \dots, X_n (c'est-à-dire par S_0, S_1, \dots, S_n). On calcule

$$\begin{aligned} \mathbf{P}(S_{n+1} = j \mid S_0 = s_0; \dots; S_{n-1} = s_{n-1}; S_n = i) \\ &= \mathbf{P}(S_n + X_{n+1} = j \mid S_0 = s_0; \dots; S_{n-1} = s_{n-1}; S_n = i) \\ &= \mathbf{P}(X_{n+1} = j - i \mid S_0 = s_0; \dots; S_{n-1} = s_{n-1}; S_n = i) \\ &= \mathbf{P}(X_{n+1} = j - i), \end{aligned}$$

en utilisant le fait que X_{n+1} est indépendant des variables aléatoires X_0, \dots, X_n , et donc de S_0, \dots, S_n .

Ainsi, S est une chaîne de Markov sur \mathbb{Z} , avec pour tout $i \in \mathbb{Z}$,

$$Q(i, j) = \begin{cases} 1/2 & \text{si } |j - i| = 1 \\ 0 & \text{sinon.} \end{cases}$$

On verra dans la suite du chapitre que lorsque E est un espace au plus dénombrable, la loi d'une chaîne de Markov est complètement déterminée par les nombres $\mu(\{i\})$ et $Q(i, j)$ où $i, j \in E$, où μ est sa loi initiale et Q sa matrice de transition.

Théorème 4.1.5

Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, soit $X = \{X_n; n \in \mathbb{N}\}$ une chaîne de Markov par rapport à sa filtration naturelle. Soient μ la loi initiale et $(Q(i, j))_{i, j \in E}$ la matrice de transition de X .

Alors pour tout $(x_0, x_1, \dots, x_n) \in E^{n+1}$,

$$\mathbf{P}(X_0 = x_0; X_1 = x_1; \dots; X_n = x_n) = \mu(x_0) Q(x_0, x_1) Q(x_1, x_2) \dots Q(x_{n-1}, x_n). \quad (4.4)$$

Réciproquement, si un processus stochastique $\{X_n; n \in \mathbb{N}\}$ vérifie (4.4) pour une mesure de probabilité μ et une matrice de transition $(Q(i, j))_{i, j \in E}$, alors X est une chaîne de Markov de loi initiale μ et de matrice de transition $(Q(i, j))_{i, j \in E}$ pour sa filtration naturelle.

Preuve On écrit la décomposition

$$\begin{aligned} \mathbf{P}(X_0 = x_0; X_1 = x_1; \dots; X_n = x_n) \\ &= \mathbf{P}(X_0 = x_0; X_1 = x_1; \dots; X_{n-1} = x_{n-1}) \times \mathbf{P}(X_n = x_n \mid X_0 = x_0; \dots; X_{n-1} = x_{n-1}) \\ &= \mathbf{P}(X_0 = x_0; X_1 = x_1; \dots; X_{n-1} = x_{n-1}) \times Q(x_{n-1}, x_n). \end{aligned}$$

L'expression recherchée s'obtient alors par récurrence.

Réciproquement, si la relation (4.4) est vérifiée, on a

$$\mathbf{P}(X_{n+1} = j \mid X_0 = x_0; \dots; X_n = x_n) = \frac{\mu(x_0) Q(x_0, x_1) \dots Q(x_n, j)}{\mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n)} = Q(x_n, j).$$

□

Dans le cas fini, la loi d'une chaîne de Markov (par rapport à sa filtration naturelle) est déterminée par les puissances de sa matrice de transition.

Corollaire 4.1.6

Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, soit $X = \{X_n; n \in \mathbb{N}\}$ une chaîne de Markov sur l'espace fini $E = \{e_1, \dots, e_N\}$. On considère la matrice de transition $Q = [Q(i, j)]_{1 \leq i, j \leq N}$ et ses puissances entières $Q^n = Q \times \dots \times Q$.

Pour tout couple $(i, j) \in E \times E$, on a

$$\mathbf{P}(X_n = j \mid X_0 = i) = (Q^n)_{ij}.$$

Preuve En utilisant le Théorème 4.1.5, on écrit

$$\begin{aligned}
\mathbf{P}(X_n = j \mid X_0 = i) &= \sum_{x_1} \dots \sum_{x_{n-1}} \mathbf{P}(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = j \mid X_0 = i) \\
&= \sum_{x_1} \dots \sum_{x_{n-1}} Q(i, x_1) Q(x_1, x_2) \dots Q(x_{n-1}, j) \\
&= (Q^n)_{ij}.
\end{aligned}$$

□

Soient $x_0, \dots, x_n, \dots, x_{n+k}$ des points quelconques de E . On écrit

$$\begin{aligned}
&\mathbf{P}(X_{n+1} = x_{n+1}; \dots; X_{n+k} = x_{n+k} \mid X_0 = x_0; \dots; X_n = x_n) \\
&= \frac{\mathbf{P}(X_0 = x_0; \dots; X_n = x_n; X_{n+1} = x_{n+1}; \dots; X_{n+k} = x_{n+k})}{\mathbf{P}(X_0 = x_0; \dots; X_n = x_n)} \\
&= \frac{\mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) Q(x_n, x_{n+1}) \dots Q(x_{n+k-1}, x_{n+k})}{\mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n)} \\
&= Q(x_n, x_{n+1}) \dots Q(x_{n+k-1}, x_{n+k}) \\
&= \mathbf{P}(X_1 = x_{n+1}; \dots; X_k = x_{n+k} \mid X_0 = x_n).
\end{aligned}$$

L'égalité obtenue montre que, sachant que la chaîne de Markov est en x_n à l'instant n , le processus obtenu par translation temporelle $\{X_{n+k}; k \in \mathbb{N}^*\}$ a les mêmes lois finies-dimensionnelles (et donc la même loi) que la chaîne de Markov initiale X partant de x_n . De plus, il est indépendant de la trajectoire de X avant l'instant n .

4.2 Loi d'une chaîne de Markov et probabilités de transition

Dans cette section, on s'intéresse à la détermination de la loi d'une chaîne de Markov en fonction de sa probabilité de transition et de sa loi initiale.

On a vu que dans le cas où E est fini ou dénombrable, la transition entre deux états successifs de la chaîne de Markov est définie par la matrice de transition. Dans le cas où E n'est plus dénombrable, la notion de matrice de transition peut être remplacée par celle de noyau de transition.

Définition 4.2.1

Soient (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) deux espaces mesurables. On appelle probabilité de transition (ou noyau de transition) de E_1 dans E_2 une application

$$\nu : E_1 \times \mathcal{E}_2 \rightarrow [0, 1]$$

vérifiant les deux conditions suivantes :

- (i) Pour tout $B \in \mathcal{E}_2$, l'application $x \mapsto \nu(x, B)$ est \mathcal{E}_1 -mesurable ;
- (ii) Pour tout $x \in E_1$, l'application $B \mapsto \nu(x, B)$ est une mesure de probabilité sur (E_2, \mathcal{E}_2) .

Exemple Soit λ une mesure positive σ -finie sur (E_2, \mathcal{E}_2) et soit $f : E_1 \times E_2 \rightarrow \mathbb{R}_+$ une application mesurable telle que

$$\forall x \in E_1; \quad \int_{E_2} f(x, y) \lambda(dy) = 1.$$

Alors l'application

$$\begin{aligned} E_1 \times \mathcal{E}_2 &\rightarrow [0, 1] \\ (x, B) &\mapsto \nu(x, B) = \int_B f(x, y) \cdot \lambda(dy) \end{aligned}$$

définit une probabilité de transition de E_1 dans E_2 .

Proposition 4.2.2

Soit ν une probabilité de transition de (E_1, \mathcal{E}_1) dans (E_2, \mathcal{E}_2) .

- Si h est une fonction mesurable positive (ou bornée) sur (E_2, \mathcal{E}_2) , alors

$$\varphi : x \mapsto \varphi(x) = \int_{E_2} h(y) \cdot \nu(x, dy)$$

définit une fonction mesurable positive (ou bornée) sur E_1 .

- Si λ est une mesure de probabilité sur (E_1, \mathcal{E}_1) , alors

$$\mu : B \mapsto \int_{E_1} \nu(x, B) \cdot \lambda(dx)$$

définit une mesure de probabilité sur (E_2, \mathcal{E}_2) .

Preuve Laissée en exercice. \square

Dans le cadre des chaînes de Markov sur un espace E au plus dénombrable, la notion de matrice de transition est équivalente à celle de probabilité de transition de E dans E . En

effet, les relations

$$\begin{aligned}\forall i \in E, \forall B \in \mathcal{E}; \quad \nu(i, B) &= \sum_{j \in B} Q(i, j) \\ \forall i, j \in E; \quad Q(i, j) &= \nu(i, \{j\})\end{aligned}$$

définissent une relation biunivoque entre matrices de transition et probabilités de transition. Ceci justifie la notation usuelle $Q(i, B)$ pour la probabilité de transition induite par Q .

Dans toute la suite de la section, l'espace E est supposé fini ou dénombrable.

Les deux points de la Proposition 4.2.2 se réduisent dans le cas particulier où $E_1 = E_2 = E$ est dénombrable.

Pour toute fonction $f : E \rightarrow \mathbb{R}_+$ et pour toute mesure positive μ sur E , on définit la fonction Qf et la mesure positive μQ par

$$\begin{aligned}i \in E \mapsto Qf(i) &= \sum_{j \in E} Q(i, j)f(j) \\ B \in \mathcal{E} \mapsto \mu Q(B) &= \sum_{i \in E} \mu(\{i\})Q(i, B) = \sum_{i \in E} \sum_{j \in B} \mu(\{i\})Q(i, j).\end{aligned}$$

Les applications $f \mapsto Qf$ et $\mu \mapsto \mu Q$ sont visiblement linéaires, et on peut considérer le crochet de dualité

$$(\mu, f) \mapsto \langle \mu, f \rangle = \int_E f(x) \cdot \mu(dx) = \sum_{i \in E} f(i) \mu(\{i\}).$$

On a alors, pour toute fonction $f : E \rightarrow \mathbb{R}_+$ et toute mesure positive μ sur E ,

$$\langle \mu, Qf \rangle = \langle \mu Q, f \rangle.$$

D'autre part, la probabilité de transition ν peut s'exprimer comme

$$\forall i \in E, \forall B \in \mathcal{E}; \quad \nu(i, B) = Q\mathbb{1}_B(i).$$

Proposition 4.2.3

Une matrice à coefficients positifs $(Q(i, j))_{i, j \in E}$ est une matrice de transition si et seulement si $Q\mathbb{1} = 1$.

Le produit de deux matrices de transition est encore une matrice de transition.

Preuve Par définition du produit $Q\mathbb{1}$, on a

$$\forall i \in E, \quad Q\mathbb{1}(i) = \sum_{j \in E} Q(i, j).$$

Ainsi, Q est une matrice de transition si et seulement si $Q\mathbb{1}(i) = 1$ pour tout $i \in E$.

Si Q et R sont deux matrices de transition, alors $Q1 = 1$ et $R1 = 1$ ce qui implique que $(QR)1 = Q(R1) = 1$. Donc QR est une matrice de transition. \square

Les notations Qf et μQ permettent d'exprimer facilement les distributions fini-dimensionnelles d'une chaîne de Markov $\{X_n; n \in \mathbb{N}\}$.

On considère μ et $(Q(i, j))_{i, j \in E}$ la loi initiale et la matrice de transition de la chaîne de Markov $\{X_n; n \in \mathbb{N}\}$.

On sait que pour tous $x_1, \dots, x_n \in E$, on a

$$\mathbf{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n).$$

Comme première conséquence de cette égalité, on peut en tirer les lois marginales P_{X_n} du processus X . Pour tous $x, y \in E$, on a

$$\mathbf{P}(X_0 = x, X_2 = y) = \mu(x) \sum_{z \in E} Q(x, z) Q(z, y) = \mu(x) Q^2(x, y)$$

et plus généralement, pour tout $n \geq 0$,

$$\mathbf{P}(X_0 = x, X_n = y) = \mu(x) Q^n(x, y).$$

On en déduit

$$\mathbf{P}(X_n = y) = \sum_{x \in E} \mu(x) Q^n(x, y) = \mu Q^n(y).$$

D'autre part, pour toute fonction $f : E^{n+1} \rightarrow \mathbb{R}$ positive ou bornée, on peut écrire

$$\mathbf{E}[f(X_0, \dots, X_n)] = \sum_{x_0, \dots, x_n \in E} f(x_0, \dots, x_n) \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n).$$

En considérant le cas particulier $f(x_0, \dots, x_n) = \mathbb{1}_{\{A_0\}}(x_0) \dots \mathbb{1}_{\{A_n\}}(x_n)$ où A_0, \dots, A_n sont des boréliens de \mathbb{R} , on obtient

$$\mathbf{P}(X_0 \in A_0, \dots, X_n \in A_n) = \sum_{x_0 \in A_0, \dots, x_n \in A_n} \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n).$$

Lorsque $A_i = \{x_i\}$ ou $A_i = E$, on en déduit pour tous $l_0, \dots, l_k \in \mathbb{N}^*$,

$$\mathbf{P}(X_{l_0} = x_0, X_{l_0+l_1} = x_1, \dots, X_{l_0+\dots+l_k} = x_k) = \mu Q^{l_0}(x_0) Q^{l_1}(x_0, x_1) \dots Q^{l_k}(x_{k-1}, x_k). \quad (4.5)$$

La définition d'une chaîne de Markov de loi initiale μ et de probabilité de transition Q peut s'exprimer comme un processus $X = \{X_n; n \in \mathbb{N}\}$ vérifiant,

(i) Pour tout $B \in \mathcal{E}$,

$$\mathbf{P}(X_0 \in B) = \mu(B);$$

(ii) et pour tout $B \in \mathcal{E}$,

$$\forall n \in \mathbb{N}; \quad \mathbf{P}(X_{n+1} \in B \mid \mathcal{F}_n) = Q(X_n, B).$$

Ainsi, le formalisme des probabilités de transition permet d'étendre la définition des chaînes de Markov au cas où l'espace E n'est pas dénombrable.

La proposition rassemble sous une forme plus rigoureuse que précédemment les résultats déjà énoncés.

Proposition 4.2.4

Dans l'espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, soit $\{X_n; n \in \mathbb{N}\}$ une chaîne de Markov par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, à valeurs dans (E, \mathcal{E}) , de matrice de transition $(Q(i, j))_{i, j \in E}$.

(i) Pour tout $n \in \mathbb{N}$ et toute fonction mesurable $f : E \rightarrow \mathbb{R}_+$,

$$\mathbf{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \mathbf{E}[f(X_{n+1}) \mid X_n] = Qf(X_n).$$

(ii) Supposons que $(\mathcal{F}_n)_{n \in \mathbb{N}}$ est la filtration naturelle de X . Pour tous $n \in \mathbb{N}$ et $p \in \mathbb{N}^*$, et pour tous $y_1, \dots, y_p \in E$,

$$\mathbf{P}(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0, \dots, X_n) = Q(X_n, y_1)Q(y_1, y_2) \dots Q(y_{p-1}, y_p),$$

ce qui implique

$$\mathbf{P}(X_{n+p} = y_p \mid X_n) = Q^p(X_n, y_p).$$

Si on pose $Y_p = X_{n+p}$ pour tout $p \in \mathbb{N}$, le processus $\{Y_p; p \in \mathbb{N}\}$ est une chaîne de Markov de matrice de transition Q .

Preuve

(i) Par définition de la chaîne de Markov, on a

$$\mathbf{P}(X_{n+1} \in B \mid \mathcal{F}_n) = \mathbf{P}(X_{n+1} \in B \mid X_n) \quad \text{p.s.}$$

On en déduit pour toute fonction mesurable $f : E \rightarrow \mathbb{R}_+$,

$$\mathbf{E}[f(X_{n+1}) \mid \mathcal{F}_n] = \mathbf{E}[f(X_{n+1}) \mid X_n] \quad \text{p.s.}$$

Dans le cas particulier $f = \mathbb{1}_B$ pour $B \in \mathcal{E}$, on a

$$\mathbf{E}[\mathbb{1}_B(X_{n+1}) \mid \mathcal{F}_n] = \mathbf{P}(X_{n+1} \in B \mid X_n) = \sum_{j \in B} Q(X_n, j) = Q\mathbb{1}_B(X_n).$$

Par combinaison linéaire et un argument de classe monotone, on en déduit le résultat pour les fonctions étagées puis les fonctions mesurables.

(ii) Pour tous $y_1, \dots, y_p, x_0, \dots, x_n \in E$, on a

$$\mathbf{P}(X_{n+1} = y_1, \dots, X_{n+p} = y_p \mid X_0 = x_0, \dots, X_n = x_n) = Q(x_n, y_1)Q(y_1, y_2) \dots Q(y_{p-1}, y_p).$$

En sommant sur tous les choix possibles de y_1, \dots, y_{p-1} , on obtient

$$\mathbf{P}(X_{n+p} = y_p \mid X_n) = \sum_{y_1, \dots, y_{p-1}} Q(x_n, y_1)Q(y_1, y_2) \dots Q(y_{p-1}, y_p) = Q^p(x_n, y_p).$$

Pour la dernière assertion, on écrit

$$\begin{aligned} \mathbf{P}(Y_0 = y_0, Y_1 = y_1, \dots, Y_p = y_p) &= \mathbf{P}(X_n = y_0, X_{n+1} = y_1, \dots, X_{n+p} = y_p) \\ &= \mathbf{P}(X_n = y_0)Q(y_0, y_1) \dots Q(y_{p-1}, y_p). \end{aligned}$$

Cette relation caractérise le fait que $\{Y_p; p \in \mathbb{N}\}$ est une chaîne de Markov de loi initiale P_{X_n} et de matrice de transition Q .

□

4.3 La chaîne de Markov canonique

4.3.1 Définition

Dans cette section, on s'intéresse à l'existence d'une chaîne de Markov à valeurs dans un espace fini ou dénombrable E , de loi initiale μ et de matrice de transition Q données.

On reprend ici l'espace de probabilité $\Omega = E^{\mathbb{N}}$ et le processus canonique $X_n(\omega) = \omega(n)$ ($\omega \in \Omega, n \in \mathbb{N}$) introduits dans le chapitre 1.

Théorème 4.3.1

Soit E un espace fini ou dénombrable. On considère l'espace canonique

$$\begin{aligned} \Omega &= E^{\mathbb{N}}, \quad X_n(\omega) = \omega(n), \quad \forall \omega \in \Omega \\ \mathcal{F}_n &= \sigma(X_k, k \leq n), \quad \mathcal{F} = \sigma(X_k, k \geq 0). \end{aligned}$$

Soient μ une mesure de probabilité et Q une matrice de transition sur l'espace E .

Alors, il existe une unique mesure de probabilité \mathbf{P} sur (Ω, \mathcal{F}) telle que $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, (X_n)_{n \in \mathbb{N}}, \mathbf{P})$ soit une chaîne de Markov de loi initiale μ et de matrice de transition Q .

Preuve Pour tout $n \in \mathbb{N}$, la fonction $\mu_n : E^{n+1} \rightarrow \mathbb{R}_+$ telle que

$$\forall x_0, \dots, x_n \in E, \quad \mu_n(x_0, \dots, x_n) = \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n)$$

définit une mesure de probabilité sur E^{n+1} .

On vérifie que les mesures de probabilité ainsi définies vérifient les conditions de compatibilité de Kolmogorov. D'après le théorème d'extension de Kolmogorov, il existe une unique mesure de probabilité \mathbf{P} sur $(\Omega = E^{\mathbb{N}}, \mathcal{F} = \sigma(X_n; n \in \mathbb{N}))$ telle que

$$\begin{aligned} \forall x_0, \dots, x_n \in E, \quad \mathbf{P}\{X_0 = x_0, \dots, X_n = x_n\} &= \mu_n(x_0, \dots, x_n) \\ &= \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n). \end{aligned}$$

D'après le Théorème 4.1.5, le processus $\{X_n; n \in \mathbb{N}\}$ est une chaîne de Markov. \square

La mesure de probabilité du théorème précédent, relative à la loi initiale μ , est notée \mathbf{P}_μ . Dans le cas particulier où $\mu = \delta_x$, c'est-à-dire lorsqu'on impose à la chaîne de Markov de partir de $x \in E$, on note $\mathbf{P}_x = \mathbf{P}_{\delta_x}$.

On a alors

$$\forall A \in \mathcal{F}, \quad \mathbf{P}_\mu(A) = \sum_{x \in E} \mu(x) \mathbf{P}_x(A).$$

La chaîne de Markov canonique $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, (X_n)_{n \in \mathbb{N}}, \mathbf{P}_\mu)$ vérifie alors pour tout $n \in \mathbb{N}$, toute application mesurable $g : E \rightarrow \mathbb{R}_+$ et toute mesure μ

$$\begin{aligned} \mathbf{P}_x(X_n = y) &= Q^n(x, y), & \mathbf{E}_x[g(X_n)] &= Q^n g(x) \\ \mathbf{P}_\mu(X_n = y) &= \mu Q^n(y), & \mathbf{E}_\mu[g(X_n)] &= \sum_{x \in E} \mu(x) Q^n g(x) = \mu Q^n g. \end{aligned}$$

4.3.2 Propriété de Markov

Le principal intérêt de considérer l'espace canonique $\Omega = E^{\mathbb{N}}$ est de pouvoir utiliser l'opérateur de translation $\theta : \Omega \rightarrow \Omega$ défini par

$$\forall \omega \in \Omega; \quad \theta(\omega) = [n \mapsto \omega(n+1)].$$

L'opérateur θ décale donc de +1 la fonction ω . On définit également les itérés

$$\theta_0 = \text{Id}, \quad \theta_n = \theta_{n-1} \circ \theta = \theta \circ \theta_{n-1}, \quad \forall n \geq 1.$$

On a donc $X_k \circ \theta_n = X_{k+n}$ pour tout $k \geq 0$ et

$$\theta_n^{-1}(\{X_0 = x_0, \dots, X_p = x_p\}) = \{X_n = x_0, \dots, X_{n+p} = x_p\}$$

pour tous p et n , ce qui montre que θ_n est une application mesurable de $(\Omega, \sigma(X_k, k \geq n))$ dans (Ω, \mathcal{F}) et que

$$\theta_n^{-1}(\mathcal{F}_p) = \sigma(X_n, \dots, X_{n+p}).$$

Théorème 4.3.2 (Propriété de Markov simple)

Soient Φ et Ψ deux applications mesurables positives sur Ω , avec Φ \mathcal{F}_n -mesurable. On a alors pour tout $n \geq 0$ et toute loi initiale μ ,

$$\mathbf{E}_\mu[\Phi \Psi \circ \theta_n] = \mathbf{E}_\mu[\Phi \mathbf{E}_{X_n}[\Psi]].$$

De manière équivalente,

$$\mathbf{E}_\mu[\Psi \circ \theta_n \mid \mathcal{F}_n] = \mathbf{E}_{X_n}[\Psi], \quad \mathbf{P}_\mu - p.s.$$

On peut traduire ce résultat en disant que la loi de $\theta_n(\omega)$ connaissant X_0, X_1, \dots, X_n est P_{X_n} .

Preuve Par définition de l'espérance conditionnelle $\mathbf{E}[\bullet \mid \mathcal{F}_n]$, il suffit de montrer la première égalité. On considère alors le cas $\Phi = \mathbb{1}_{\{X_0=x_0, \dots, X_n=x_n\}}$ et $\Psi = \mathbb{1}_{\{X_0=y_0, \dots, X_p=y_p\}}$, où $x_0, \dots, x_n, y_0, \dots, y_p \in E$.

On a

$$\begin{aligned} \mathbf{E}_\mu[\Phi \Psi \circ \theta_n] &= \mathbf{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_n = y_0, X_{n+1} = y_1, \dots, X_{n+p} = y_p) \\ &= \mu(x_0)Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \mathbb{1}_{\{y_0=x_n\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p). \end{aligned}$$

D'autre part, pour tout $y \in E$,

$$\mathbf{E}_y[\Psi] = \mathbb{1}_{\{y_0=y\}} Q(y_0, y_1) \dots Q(y_{p-1}, y_p),$$

ce qui entraîne

$$\mathbf{E}_\mu[\Phi \mathbf{E}_{X_n}[\Psi]] = \mathbf{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_n = y_0) Q(y_0, y_1) \dots Q(y_{p-1}, y_p)$$

et

$$\mathbf{E}_\mu[\Phi \Psi \circ \theta_n] = \mathbf{E}_\mu[\Phi \mathbf{E}_{X_n}[\Psi]].$$

On conclut par un argument de classe monotone. \square

Théorème 4.3.3 (Propriété de Markov forte)

Soit T un temps d'arrêt de la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Soient Φ et Ψ deux applications mesurables positives sur Ω . Supposons que Φ est \mathcal{F}_T -mesurable. Alors pour toute loi initiale μ ,

$$\mathbf{E}_\mu[\mathbb{1}_{\{T < \infty\}} \Phi \Psi \circ \theta_T] = \mathbf{E}_\mu[\mathbb{1}_{\{T < \infty\}} \Phi \mathbf{E}_{X_T}[\Psi]].$$

De manière équivalente,

$$\mathbf{E}_\mu[\mathbb{1}_{\{T < \infty\}} \Psi \circ \theta_T \mid \mathcal{F}_T] = \mathbb{1}_{\{T < \infty\}} \mathbf{E}_{X_T}[\Psi], \quad \mathbf{P}_\mu - p.s.$$

Preuve Pour tout $n \in \mathbb{N}$,

$$\mathbf{E}_\mu[\mathbb{1}_{\{T=n\}}\Phi \Psi \circ \theta_T] = \mathbf{E}_\mu[\mathbb{1}_{\{T=n\}}\Phi \Psi \circ \theta_n] = \mathbf{E}_\mu[\mathbb{1}_{\{T=n\}}\Phi \mathbf{E}_{X_n}[\Psi]]$$

en utilisant la propriété de Markov simple pour $\mathbb{1}_{\{T=n\}}\Phi$ qui est \mathcal{F}_n -mesurable car Φ est \mathcal{F}_T -mesurable. On somme alors l'égalité précédente pour $n \in \mathbb{N}$ et le résultat suit. \square

Chapitre 5

Classification des états des chaînes de Markov

5.1 Opérateur potentiel

Soit Q une matrice de transition et $X = \{X_n; n \in \mathbb{N}\}$ la chaîne de Markov canonique associée de matrice de transition Q .

Définition 5.1.1

On appelle matrice potentielle de Q (ou de X) la matrice

$$U = I + Q + Q^2 + Q^3 + \cdots = \sum_{k \geq 0} Q^k.$$

Soit $f : E \rightarrow \mathbb{R}_+$, la relation $\mathbf{E}_x[f(X_n)] = Q^n f(x)$ pour tout $x \in E$, implique

$$Uf(x) = \sum_{k \geq 0} Q^k f(x) = \sum_{k \geq 0} \mathbf{E}_x[f(X_k)] = \mathbf{E}_x \left[\sum_{k \geq 0} f(X_k) \right].$$

Dans le cas particulier où $f = \mathbb{1}_B$ pour $B \in \mathcal{E}$, on détermine la probabilité de transition

$$U(x, B) = U\mathbb{1}_B(x) = \mathbf{E}_x \left[\sum_{k \geq 0} \mathbb{1}_B(X_k) \right] = \mathbf{E}_x[N_B],$$

où $N_B = \sum_{k \geq 0} \mathbb{1}_B(X_k)$ représente le nombre de visites de la chaîne de Markov dans B , c'est-à-dire le nombre d'indices n tels que $X_n \in B$.

Dans le cas où $B = \{y\}$ pour $y \in E$, la probabilité de transition

$$U(x, y) = \mathbf{E}_x[N_y]$$

s'interprète comme le nombre moyen de visites en y de la chaîne de Markov partant de x . L'application $(x, y) \mapsto U(x, y)$ est également appelée *noyau potentiel* de X .

Proposition 5.1.2

Pour toute application $f : E \rightarrow \mathbb{R}_+$, la fonction $v = Uf$, appelée potentiel de f , est la plus petite solution positive de l'équation

$$v = f + Qv.$$

Preuve On vérifie aisément que

$$U = I + QU,$$

ce qui entraîne $v = f + Qv$.

Supposons maintenant que $w : E \rightarrow \mathbb{R}_+$ vérifie $w = f + Qw$. Il est clair que $w \geq f$ car Q et w sont à valeurs positives. On montre alors par récurrence sur n que $w \geq \sum_{k=0}^n Q^k f$.

- Pour $n = 0$, l'assertion se réduit à $w \geq f$.
- Si $w \geq \sum_{k=0}^n Q^k f$, alors

$$w = f + Qw \geq f + Q \sum_{k=0}^n Q^k f = \sum_{k=0}^{n+1} Q^k f.$$

En passant à la limite $n \rightarrow \infty$, on déduit $w \geq v$. \square

5.2 Communication entre les états d'une chaîne de Markov

5.2.1 Récurrence et transience

Dans cette section, $\{X_n; n \in \mathbb{N}\}$ désigne toujours la chaîne de Markov canonique sur $E^{\mathbb{N}}$.

Pour tout $x \in E$, on considère

$$N_x = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=x\}} \quad \text{le nombre de visites en } x$$

et $H_x = \inf\{n \geq 1 : X_n = x\}$ le temps de retour en x .

Attention : Si $\tau_x = \inf\{n \geq 0 : X_n = x\}$ désigne le temps d'atteinte de l'état $x \in E$, on a $H_x \neq \tau_x$ en général. En effet, si la chaîne de Markov canonique part de x (i.e. $X_0 = x$), on a évidemment $\tau_x = 0$ alors que $H_x \geq 1$. Cependant les deux temps d'arrêt sont reliés par

$$H_x = 1 + \tau_x \circ \theta,$$

si $\theta : E^{\mathbb{N}} \rightarrow E^{\mathbb{N}}$ désigne l'opérateur de translation défini par $\theta(\omega) = [n \mapsto \omega(n+1)]$.

Proposition 5.2.1 (et Définition)

Soit $x \in E$. L'une des deux assertions suivantes est vérifiée :

- $\mathbf{P}_x(H_x < \infty) = 1$: on a alors $N_x = \infty$ \mathbf{P}_x -p.s. et $\mathbf{E}_x[N_x] = \infty$.
L'état x est dit récurrent.
- $\mathbf{P}_x(H_x < \infty) < 1$: on a alors $N_x < \infty$ \mathbf{P}_x -p.s. et $\mathbf{E}_x[N_x] = \frac{1}{\mathbf{P}_x(H_x = \infty)} < \infty$.
L'état x est dit transitoire (ou transient).

De manière moins formelle, si x est récurrent alors la chaîne de Markov partant de x retourne en x une infinité de fois. Au contraire, si x est transitoire alors la chaîne partant de x ne visite x qu'un nombre fini de fois.

Preuve Pour tout entier $k \geq 1$, on a

$$\mathbb{1}_{\{N_x \geq k+1\}} = \mathbb{1}_{\{H_x < \infty\}} \mathbb{1}_{\{N_x \geq k\}} \circ \theta_{H_x},$$

ce qui entraîne

$$\mathbf{P}_x(N_x \geq k+1) = \mathbf{E}_x[\mathbb{1}_{\{H_x < \infty\}} \mathbb{1}_{\{N_x \geq k\}} \circ \theta_{H_x}].$$

D'après la propriété de Markov forte pour le temps d'arrêt H_x ,

$$\begin{aligned} \mathbf{E}_x[\mathbb{1}_{\{H_x < \infty\}} \mathbb{1}_{\{N_x \geq k\}} \circ \theta_{H_x}] &= \mathbf{E}_x[\mathbb{1}_{\{H_x < \infty\}} \mathbf{E}_x[\mathbb{1}_{\{N_x \geq k\}}]] \\ &= \mathbf{P}_x(H_x < \infty) \mathbf{P}_x(N_x \geq k). \end{aligned}$$

On en déduit

$$\mathbf{P}_x(N_x \geq k+1) = \mathbf{P}_x(H_x < \infty) \mathbf{P}_x(N_x \geq k).$$

Comme $\mathbf{P}_x(N_x \geq 1) = 1$, on trouve par récurrence

$$\forall k \geq 1, \quad \mathbf{P}_x(N_x \geq k) = \mathbf{P}_x(H_x < \infty)^{k-1}.$$

Ainsi, soit $\mathbf{P}_x(H_x < \infty) = 1$ ce qui entraîne $\mathbf{P}_x(N_x = \infty) = 1$, soit $\mathbf{P}_x(H_x < \infty) < 1$ et

$$\mathbf{E}_x[N_x] = \sum_{k=1}^{\infty} \mathbf{P}_x(N_x \geq k) = \frac{1}{\mathbf{P}_x(H_x = \infty)} < \infty.$$

□

Proposition 5.2.2

Soit U le noyau potentiel de la chaîne de Markov X .

- Pour tout $x \in E$, $U(x, x) = \infty$ si et seulement si x est récurrent.
- Pour tous $x, y \in E$ tels que $x \neq y$,

$$U(x, y) = \mathbf{P}_x(H_y < \infty) U(y, y).$$

Preuve Par définition du noyau potentiel, $U(x, x) = \infty$ si et seulement si $\mathbf{E}_x[N_x] = \infty$, c'est-à-dire si et seulement si x est récurrent (Proposition 5.2.1).

Le second point se montre en utilisant la propriété de Markov forte

$$\mathbf{E}_x[N_y] = \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} N_y \circ \theta_{H_y}] = \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbf{E}_y[N_y]] = \mathbf{P}_x(H_y < \infty) U(y, y).$$

□

La fonction $x \mapsto U(x, y)$ atteint son maximum au point y . Cette propriété est appelée *principe du maximum*.

Une conséquence de ce résultat est que si E est fini, alors la chaîne possède au moins un état récurrent.

Pour le montrer, on remarque d'abord que $N_E = \infty$ p.s. On en déduit pour $x \in E$,

$$\sum_{y \in E} U(x, y) = \mathbf{E}_x \left[\sum_{y \in E} N_y \right] = \mathbf{E}_x[N_E] = \infty.$$

La somme étant finie, il existe donc $y \in E$ tel que $U(x, y) = \infty$ p.s. Comme $U(y, y) \geq U(x, y)$, le point y est récurrent.

5.2.2 Relation d'équivalence $x \rightsquigarrow y$

Dans la suite, on s'intéresse aux communications entre les états de E .

On dit que x conduit à y , ce qu'on note $x \rightsquigarrow y$, si l'une des trois conditions équivalentes suivantes est vérifiée :

- (i) $\mathbf{P}_x(H_y < \infty) > 0$;
- (ii) Il existe $n \geq 0$ tel que $Q^n(x, y) > 0$;
- (iii) $U(x, y) > 0$.

Exercice : Vérifier que les conditions (i), (ii) et (iii) précédentes sont équivalentes.

Définition 5.2.3

La chaîne de Markov est dite irréductible si $x \rightsquigarrow y$ pour tous $x, y \in E$.

Lemme 5.2.4

Soient $x, y, z \in E$. Alors on a

$$\mathbf{P}_x(H_z < \infty) \geq \mathbf{P}_x(H_y < \infty) \mathbf{P}_y(H_z < \infty).$$

En conséquence, si $x \rightsquigarrow y$ et $y \rightsquigarrow z$ alors $x \rightsquigarrow z$.

Preuve On a l'implication

$$H_y + H_z \circ \theta_{H_y} < \infty \Rightarrow H_z < \infty.$$

Ceci implique

$$\mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbb{1}_{\{H_z < \infty\}} \circ \theta_{H_y}] \leq \mathbf{P}_x(H_z < \infty).$$

Or, en utilisant la propriété de Markov forte pour le temps d'arrêt H_y ,

$$\begin{aligned} \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbb{1}_{\{H_z < \infty\}} \circ \theta_{H_y}] &= \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbf{E}_y[\mathbb{1}_{\{H_z < \infty\}}]] \\ &= \mathbf{P}_x(H_y < \infty) \mathbf{P}_y(H_z < \infty). \end{aligned}$$

□

Lemme 5.2.5

Soit x un point récurrent et y tel que $x \rightsquigarrow y$. Alors y est récurrent et on a également $y \rightsquigarrow x$.

Preuve • On montre d'abord que $\mathbf{P}_y(H_x < \infty) = 1$.

Si la chaîne $\{X_n; n \in \mathbb{N}\}$ partant de x , atteint l'état y et si la chaîne translatée en y ne retourne plus en x , alors le nombre de visites de la chaîne en x est fini :

$$\begin{aligned} 0 &= \mathbf{P}_x(N_x < \infty) \geq \mathbf{P}_x(H_y < \infty, H_x \circ \theta_{H_y} = \infty) \\ &= \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbb{1}_{\{H_x = \infty\}} \circ \theta_{H_y}]. \end{aligned}$$

En utilisant la propriété de Markov forte pour le temps d'arrêt H_y , on obtient

$$\begin{aligned} 0 &= \mathbf{P}_x(N_x < \infty) \geq \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbf{P}_y(H_x = \infty)] \\ &= \mathbf{P}_x(H_y < \infty) \mathbf{P}_y(H_x = \infty). \end{aligned}$$

Comme $x \rightsquigarrow y$, on a $\mathbf{P}_x(H_y < \infty) > 0$ et donc $\mathbf{P}_y(H_x = \infty) = 0$, i.e. $y \rightsquigarrow x$.

- Pour montrer que y est récurrent, on montre que $U(y, y) = \infty$.

Comme $x \rightsquigarrow y$ et $y \rightsquigarrow x$, il existe deux entiers n et p tels que $Q^n(x, y) > 0$ et $Q^p(y, x) > 0$. Pour tout $k \in \mathbb{N}$, on a la minoration

$$Q^{n+k+p}(y, y) \geq Q^p(y, x)Q^k(x, x)Q^n(x, y)$$

ce qui entraîne

$$U(y, y) \geq \sum_{k=0}^{\infty} Q^{n+k+p}(y, y) \geq Q^p(y, x) \underbrace{\left(\sum_{k=0}^{\infty} Q^k(x, x) \right)}_{=U(x, x)=\infty} Q^n(x, y).$$

On en déduit que $U(y, y) = \mathbf{E}_y[N_y] = \infty$, ce qui implique $N_y = \infty$ \mathbf{P}_y -p.s., c'est-à-dire y est récurrent. \square

Une conséquence immédiate du Lemme 5.2.5 est que si x est récurrent et y transitoire, alors $U(x, y) = 0$.

Les Lemmes 5.2.4 et 5.2.5 montrent que la relation binaire

$$x\mathcal{R}y \iff x \rightsquigarrow y$$

est une relation d'équivalence sur l'ensemble des points récurrents (\mathcal{R} est réflexive, symétrique et transitive). Les classes d'équivalence de cette relation forment une partition de l'ensemble des points récurrents ; elles sont appelées *classes de récurrence* de la chaîne de Markov.

Théorème 5.2.6

Il existe une partition $(C_i)_{i \in I}$ des états récurrents de E telle que

- (i) Si $x, y \in C_i$, alors $U(x, y) = +\infty$ et $\mathbf{P}_x(N_y = +\infty) = 1$;
- (ii) Si $x \in C_i$ et $y \in C_j$, où $i \neq j$, alors $U(x, y) = 0$ et $\mathbf{P}_x(N_y = 0) = 1$;
- (iii) Si x est transitoire, alors $T = \inf\{n \geq 0 : X_n \in \bigcup_i C_i\}$ vérifie \mathbf{P}_x -p.s.

- Soit $T = \infty$, et $N_y < \infty$ pour tout $y \in E$.
- Soit $T < \infty$, et il existe $i \in I$ tel que $\forall n \geq T, X_n \in C_i$.

Preuve On considère les classes d'équivalence $(C_i)_{i \in I}$ de la relation \mathcal{R} sur les états récurrents de E .

- (i) Soient x et y deux états de la même classe C_i . On a alors $x \rightsquigarrow y$, c'est-à-dire $\mathbf{P}_x(H_y < \infty) = 1$. On en déduit, en utilisant la propriété de Markov forte pour le temps d'arrêt H_y ,

$$\begin{aligned} \mathbf{P}_x(N_y = \infty) &= \mathbf{E}_x[\mathbb{1}_{\{H_y < \infty\}} \mathbb{1}_{\{N_y = \infty\}} \circ \theta_{H_y}] \\ &= \mathbf{P}_x(H_y < \infty) \mathbf{P}_y(N_y = \infty) = 1, \end{aligned}$$

car y est un état récurrent.

(ii) Si $x \in C_i$ et $y \in C_j$ avec $i \neq j$, alors x ne conduit pas à y , c'est à dire $\mathbf{P}_x(N_y = 0) = 1$. Par définition, cela s'écrit $U(x, y) = 0$.

(iii) Soit x un état transitoire, i.e. $x \in E \setminus (\bigcup_{i \in I} C_i)$.

- Si $T(\omega) = \infty$, alors pour tout $y \in E$, l'état y n'est pas récurrent (donc il est transient). Donc $N_y < \infty$ \mathbf{P}_y -p.s. Grâce à la propriété de Markov forte appliquée au temps d'arrêt H_y , l'égalité

$$\{N_y < \infty\} = \{H_y < \infty\} \cap (\{N_y < \infty\} \circ \theta_{H_y}),$$

permet de conclure que $N_y < \infty$ \mathbf{P}_x -p.s.

- Si $T(\omega) < \infty$, il existe $i(\omega) \in I$ tel que $X_T(\omega) \in C_i$. D'après (i), $X_n \circ \theta_T \in C_i$ pour tout $n \geq 0$. Autrement dit, $X_n \in C_i$ pour tout $n \geq T$.

□

Corollaire 5.2.7

Soit X une chaîne de Markov irréductible. Alors, on a la dichotomie suivante :

- *Soit tous les états sont récurrents, il existe une seule classe de récurrence et on a, pour tout $x \in E$,*

$$\mathbf{P}_x(N_y = \infty, \forall y \in E) = 1.$$

La chaîne est alors dite récurrente irréductible.

- *Soit tous les états sont transitoires, et pour tout $x \in E$,*

$$\mathbf{P}_x(N_y < \infty, \forall y \in E) = 1.$$

La chaîne est alors dite transiente irréductible.

Comme on a vu que lorsque E est fini, il existe nécessairement un état récurrent, seul le premier cas peut alors se produire.

Preuve S'il existe un état récurrent, le Lemme 5.2.5 implique que tous les états sont récurrents. Comme on a alors $U(x, y) > 0$ pour tous $x, y \in E$, il n'y a qu'une seule classe de récurrence.

Le reste découle du théorème précédent. □

5.3 Mesures invariantes

Définition 5.3.1

Une chaîne de Markov $\{X_n; n \in \mathbb{N}\}$ est dite

- stationnaire si, pour tous $k, n \in \mathbb{N}$ les vecteurs aléatoires (X_0, \dots, X_n) et (X_k, \dots, X_{k+n}) ont même loi.
- réversible si pour tout entier $n \geq 1$, les vecteurs aléatoires (X_0, X_1, \dots, X_n) et $(X_n, X_{n-1}, \dots, X_0)$ ont même loi.

Définition 5.3.2

Soit μ une mesure positive sur E . On dit que

- μ est invariante pour la matrice de transition Q si $\mu = \mu Q$, i.e.

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

- μ est réversible par rapport à la matrice de transition Q si

$$\forall x, y \in E, \quad \mu(x) Q(x, y) = \mu(y) Q(y, x).$$

On notera qu'on utilise ici l'abus de notation $\mu(x)$ au lieu de $\mu(\{x\})$, qui est justifié par le fait qu'une mesure sur un espace dénombrable est déterminé par ses valeurs sur les atomes.

Proposition 5.3.3

Soient μ une mesure de probabilité et $Q = (Q(i, j))_{i, j \in E}$ une matrice de transition. Soit $\{X_n; n \in \mathbb{N}\}$ une chaîne de Markov de loi initiale μ et de matrice de transition Q .

- Si μ est invariante pour la matrice de transition Q , alors X est stationnaire.
- Si μ est réversible par rapport à la matrice de transition Q , alors X est réversible.

Preuve

- En utilisant le fait que μ est invariante pour Q , on écrit

$$\begin{aligned} \mathbf{P}(X_k = x_0, \dots, X_{k+n} = x_n) &= \mu Q^k(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= \mathbf{P}(X_0 = x_0, \dots, X_n = x_n). \end{aligned}$$

Ce qui montre que la chaîne de Markov X est stationnaire.

- Comme μ est réversible par rapport à Q , on a

$$\begin{aligned} \mathbf{P}(X_0 = x_0, \dots, X_n = x_n) &= \mu(x_0) Q(x_0, x_1) \dots Q(x_{n-1}, x_n) \\ &= Q(x_1, x_0) \mu(x_1) Q(x_1, x_2) \dots Q(x_{n-1}, x_n) \\ &= Q(x_1, x_0) Q(x_2, x_1) \dots Q(x_n, x_{n-1}) \mu(x_n) \\ &= \mathbf{P}(X_0 = x_n, \dots, X_n = x_0). \end{aligned}$$

Ce qui montre que la chaîne de Markov X est réversible. \square

Le lien entre réversibilité et invariance d'une mesure est décrit par le résultat suivant.

Proposition 5.3.4

Si μ est une mesure réversible par rapport à une matrice de transition $Q = (Q(i, j))_{i, j \in E}$, alors elle est invariante pour Q .

Preuve Si μ est réversible, alors

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

\square

Dans le reste de cette section, on s'intéresse à l'existence d'une mesure (pas forcément une mesure de probabilité) invariante par rapport à la matrice de transition d'une chaîne de Markov.

Théorème 5.3.5

Soit $X = \{X_n; n \in \mathbb{N}\}$ la chaîne de Markov canonique à valeurs dans E , de matrice de transition Q et soit $x \in E$ un point récurrent.

La fonction $\nu_x : E \rightarrow \mathbb{R}$ définie par

$$\forall y \in E, \quad \nu_x(y) = \mathbf{E}_x \left[\sum_{k=0}^{H_x-1} \mathbb{1}_{\{X_k=y\}} \right] \quad (5.1)$$

est une mesure invariante sur E pour la matrice Q .

De plus, $\nu_x(y) > 0$ si et seulement si y appartient à la classe de récurrence de x .

Preuve Si y n'est pas dans la classe de récurrence de x , alors $U(x, y) = \mathbf{E}_x[N_y] = 0$. On a alors $\nu_x(y) = 0$.

Pour tout $y \in E$, on calcule

$$\begin{aligned}\nu_x(y) &= \mathbf{E}_x \left[\sum_{k=1}^{H_x} \mathbb{1}_{\{X_k=y\}} \right] = \sum_{z \in E} \mathbf{E}_x \left[\sum_{k=1}^{H_x} \mathbb{1}_{\{X_{k-1}=z, X_k=y\}} \right] \\ &= \sum_{z \in E} \sum_{k=1}^{\infty} \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} \cdot \mathbb{1}_{\{X_k=y\}} \right].\end{aligned}$$

Or, l'événement $\{k \leq H_x, X_{k-1} = z\}$ est \mathcal{F}_{k-1} -mesurable. On a alors

$$\begin{aligned}\mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} \cdot \mathbb{1}_{\{X_k=y\}} \right] &= \mathbf{E}_x \left[\mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} \cdot \mathbb{1}_{\{X_k=y\}} \mid \mathcal{F}_{k-1} \right] \right] \\ &= \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} \mathbf{P}_x(X_k = y \mid \mathcal{F}_{k-1}) \right] \\ &= \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} Q(X_{k-1}, y) \right] \\ &= \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} Q(z, y) \right].\end{aligned}$$

On a donc

$$\begin{aligned}\nu_x(y) &= \sum_{z \in E} \sum_{k=1}^{\infty} \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x, X_{k-1}=z\}} \right] Q(z, y) \\ &= \sum_{z \in E} \mathbf{E}_x \left[\sum_{k=1}^{H_x} \mathbb{1}_{\{X_{k-1}=z\}} \right] Q(z, y) \\ &= \sum_{z \in E} \nu_x(z) Q(z, y) = \nu_x Q(y).\end{aligned}$$

La mesure ν_x est donc invariante pour Q . \square

Théorème 5.3.6

Soit $X = \{X_n; n \in \mathbb{N}\}$ une chaîne de Markov récurrente irréductible. Il existe une mesure invariante non triviale, unique à une constante multiplicative près.

Preuve Suite au Théorème 5.3.5, il suffit de montrer que toute mesure invariante μ de X est proportionnelle à la mesure ν_x définie par (5.1).

Par récurrence, on montre que pour tout $p \in \mathbb{N}$, pour tous $x, y \in E$, on a

$$\mu(y) \geq \mu(x) \mathbf{E}_x \left[\sum_{k=0}^{p \wedge (H_x-1)} \mathbb{1}_{\{X_k=y\}} \right]. \quad (5.2)$$

Dans la suite, on suppose que $x \neq y$ (l'inégalité est évidente lorsque $x = y$).

- $p = 0$: l'égalité (5.2) se réduit à $\mu(y) \geq 0$, qui est trivialement vérifiée.

- Supposons que l'inégalité (5.2) soit vérifiée pour $p \geq 0$. L'invariance de μ par rapport à Q permet d'écrire

$$\begin{aligned}\mu(y) &= \sum_{z \in E} \mu(z) Q(z, y) \\ &\geq \mu(x) \sum_{z \in E} \mathbf{E}_x \left[\sum_{k=0}^{p \wedge (H_x - 1)} \mathbb{1}_{\{X_k = z\}} \right] Q(z, y),\end{aligned}$$

en utilisant l'hypothèse de récurrence. Ainsi,

$$\begin{aligned}\mu(y) &\geq \mu(x) \sum_{z \in E} \sum_{k=0}^p \mathbf{E}_x \left[\mathbb{1}_{\{X_k = z, k \leq H_x - 1\}} \right] Q(z, y) \\ &\geq \mu(x) \sum_{z \in E} \sum_{k=0}^p \mathbf{E}_x \left[\mathbb{1}_{\{X_k = z, k \leq H_x - 1\}} \mathbb{1}_{\{X_{k+1} = y\}} \right],\end{aligned}$$

en utilisant la définition de $Q(z, y)$ et la propriété de Markov (simple) à l'instant k en notant que $\{X_k = z, k \leq H_x - 1\}$ est \mathcal{F}_k -mesurable. On en déduit

$$\begin{aligned}\mu(y) &\geq \mu(x) \sum_{k=0}^p \mathbf{E}_x \left[\mathbb{1}_{\{k \leq H_x - 1\}} \mathbb{1}_{\{X_{k+1} = y\}} \right] \\ &\geq \mu(x) \sum_{k=1}^{(p+1) \wedge H_x} \mathbf{E}_x \left[\mathbb{1}_{\{X_k = y\}} \right]\end{aligned}$$

Comme $x \neq y$, le terme correspondant à $k = 0$ ou $k = H_x$ est nul. La dernière égalité est donc le résultat recherché à l'ordre $p + 1$.

On fait tendre $p \rightarrow \infty$ dans l'inégalité (5.2). On obtient

$$\forall y \in E, \quad \mu(y) \geq \mu(x) \mathbf{E}_x \left[\sum_{k=0}^{H_x - 1} \mathbb{1}_{\{X_k = y\}} \right] = \mu(x) \nu_x(y).$$

Pour tout $n \geq 1$, l'invariance de μ et de ν_x permet d'écrire

$$\mu(x) = \sum_{z \in E} \mu(z) Q^n(z, x) \geq \sum_{z \in E} \mu(x) \nu_x(z) Q^n(z, x) = \mu(x) \nu_x(x) = \mu(x).$$

On en déduit qu'il y a égalité $\mu(z) = \mu(x) \nu_x(z)$ pour tout $z \in E$ tel que $Q^n(z, x) > 0$. La chaîne X étant irréductible, pour tout $z \in E$, il existe un entier n tel que $Q^n(z, x) > 0$. On peut donc conclure que la mesure μ peut s'écrire $\mu = \mu(x) \nu_x$, c'est-à-dire que μ est proportionnelle à ν_x . \square

Théorème 5.3.7

Soit X une chaîne de Markov récurrente irréductible. On a la dichotomie suivante :

- Soit il existe une unique mesure de probabilité invariante μ , et

$$\forall x \in E; \quad \mathbf{E}_x[H_x] = \frac{1}{\mu(x)}.$$

La chaîne est dite récurrente positive.

- Soit toute mesure invariante μ vérifie $\mu(E) = +\infty$, et

$$\forall x \in E; \quad \mathbf{E}_x[H_x] = +\infty.$$

La chaîne est dite récurrente nulle.

Preuve D'après le Théorème 5.3.6, toutes les mesures invariantes sont proportionnelles. Donc soit (i) elles sont toutes finies ; soit (ii) elles sont toutes infinies.

- (i) Dans ce cas, on normalise la mesure finie ν_x pour obtenir une mesure de probabilité invariante μ .

On détermine la constante de proportionnalité $C > 0$ telle que $\mu = C \nu_x$,

$$\mu(E) = 1 = C \nu_x(E).$$

On a alors $\mu(x) = \frac{\nu_x(x)}{\nu_x(E)} = \frac{1}{\nu_x(E)}$. On détermine alors $\nu_x(E)$

$$\nu_x(E) = \sum_{y \in E} \mathbf{E}_x \left[\sum_{k=0}^{H_x-1} \mathbb{1}_{\{X_k=y\}} \right] = \mathbf{E}_x \left[\sum_{k=0}^{H_x-1} \sum_{y \in E} \mathbb{1}_{\{X_k=y\}} \right] = \mathbf{E}_x[H_x].$$

On a donc $\mu(x) = \frac{1}{\mathbf{E}_x[H_x]}$.

- (ii) Dans le cas où toutes les mesures invariantes sont infinies, le calcul précédent de $\nu_x(E)$ reste valable et on a $\mathbf{E}_x[H_x] = \nu_x(E) = +\infty$.

□

Proposition 5.3.8

Soit $X = \{X_n; n \in \mathbb{N}\}$ une chaîne de Markov irréductible.

Si X possède une probabilité invariante, alors X est récurrente.

Preuve Soit μ une mesure invariante finie et soit $y \in E$ tel que $\mu(y) > 0$.

Pour tout $x \in E$, on a

$$U(x, y) = \sum_{n=0}^{\infty} Q^n(x, y) \leq U(y, y),$$

par le principe du maximum.

En multipliant par $\mu(x)$ et en sommant l'inégalité sur tous les $x \in E$, on obtient

$$\sum_{n=0}^{\infty} \mu Q^n(y) \leq \mu(E) U(y, y).$$

Or, $\mu Q^n(y) = \mu(y)$ car μ est invariante pour Q .

Comme $\mu(y) > 0$, on obtient

$$+\infty = \sum_{n=0}^{\infty} \mu(y) \leq \mu(E) U(y, y),$$

i.e. $\mu(E) U(y, y) = +\infty$ soit $U(y, y) = +\infty$. L'état y est donc récurrent et la chaîne est récurrente. \square

Chapitre 6

Comportement asymptotique

Dans cette partie, on s'intéresse au comportement de quantités

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \quad \text{ou} \quad \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)}$$

lorsque n tend vers ∞ .

De tels résultats sont appelés *théorèmes ergodiques*.

6.1 Cas récurrent irréductible

Théorème 6.1.1

Soit X une chaîne de Markov récurrente irréductible et soit μ une mesure invariante. Soient f et g deux fonctions positives dans $L^1(\mu)$ avec $\mu(g) \neq 0$, alors pour tout $x \in E$,

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \rightarrow \frac{\mu(f)}{\mu(g)} \quad \mathbf{P}_x - p.s.$$

lorsque n tend vers ∞ .

Preuve La preuve repose sur la loi forte des grands nombres. On distingue plusieurs étapes :

- Etape 1 : Définition des variables aléatoires $Z_n(f)$.

Pour x fixé, on définit la suite $(T_n)_{n \in \mathbb{N}}$ par $T_0 = 0$, $T_1 = H_x = \inf\{k > 0 : X_k = x\}$ et

$$\forall n \geq 1, \quad T_n = \inf\{k > T_{n-1} : X_k = x\}.$$

Pour tout $n \in \mathbb{N}$, on définit

$$Z_n(f) = \sum_{k=T_n}^{T_{n+1}-1} f(X_k).$$

On fixe $n \geq 1$. Pour tout $k \leq n-1$, on pose $Y_{k+1} = \sum_{l=0}^{T_{k+1}-1} f(X_l)$. On a alors $Z_k(f) = Y_{k+1} - Y_k$.

Montrons que $Z_k(f)$ est \mathcal{F}_{T_n} -mesurable.

Pour tout $A \in \mathcal{B}(\mathbb{R}_+)$ et pour tout $m \in \mathbb{N}$, on a

$$\{Y_p \in A\} \cap \{T_p = m\} = \left\{ \sum_{l=0}^{m-1} f(X_l) \in A \right\} \cap \{T_p = m\}.$$

Or, pour tout $l \leq m-1$, $f(X_l)$ est \mathcal{F}_{m-1} -mesurable. Ceci implique que $\sum_{l=0}^{m-1} f(X_l)$ est \mathcal{F}_{m-1} -mesurable, puis que $\left\{ \sum_{l=0}^{m-1} f(X_l) \in A \right\} \in \mathcal{F}_{m-1}$. En utilisant le fait que T_p est un temps d'arrêt, on en déduit que

$$\{Y_p \in A\} \cap \{T_p = m\} \in \mathcal{F}_m,$$

puis que Y_p est \mathcal{F}_{T_p} -mesurable.

Comme la suite des temps d'arrêt $(T_k)_{k \in \mathbb{N}}$ est croissante, on a $\mathcal{F}_{T_k} \subset \mathcal{F}_{T_{k+1}} \subset \mathcal{F}_{T_n}$ pour tout $k \leq n-1$. Donc $Z_k(f)$ est \mathcal{F}_{T_n} -mesurable.

• Etape 2 : Les variables aléatoires $(Z_n(f))_{n \geq 0}$ sont indépendantes et de même loi.

Pour tout $n \in \mathbb{N}$, on a $Z_0(f) \circ \theta_{T_n} = Z_n(f)$. On montre, par récurrence, que pour tout $n \in \mathbb{N}$,

$$\mathbf{E}_x \left[\prod_{k=0}^n g_k(Z_k(f)) \right] = \prod_{k=0}^n \mathbf{E}_x [g_k(Z_0(f))]. \quad (6.1)$$

- Pour $n = 0$, l'égalité est triviale.

- Supposons l'égalité vérifiée pour $n \in \mathbb{N}$. Montrons qu'elle l'est pour $n+1$.

On utilise la propriété de Markov forte pour le temps d'arrêt T_{n+1} pour passer de n à $n+1$. On écrit

$$\begin{aligned} \mathbf{E}_x \left[\prod_{k=0}^{n+1} g_k(Z_k(f)) \right] &= \mathbf{E}_x \left[\left(\prod_{k=0}^n g_k(Z_k(f)) \right) g_{n+1}(Z_{n+1}(f)) \right] \\ &= \mathbf{E}_x \left[\left(\prod_{k=0}^n g_k(Z_k(f)) \right) g_{n+1}(Z_0(f) \circ \theta_{T_{n+1}}) \right]. \end{aligned}$$

Comme $\prod_{k=0}^n g_k(Z_k(f))$ est $\mathcal{F}_{T_{n+1}}$ -mesurable, la propriété de Markov forte appliquée au temps d'arrêt T_{n+1} fini presque sûrement implique,

$$\begin{aligned} \mathbf{E}_x \left[\prod_{k=0}^{n+1} g_k(Z_k(f)) \right] &= \mathbf{E}_x \left[\left(\prod_{k=0}^n g_k(Z_k(f)) \right) g_{n+1}(Z_0(f) \circ \theta_{T_{n+1}}) \right] \\ &= \mathbf{E}_x \left[\left(\prod_{k=0}^n g_k(Z_k(f)) \right) \mathbf{E}_{X_{T_{n+1}}} [g_{n+1}(Z_0(f))] \right] \\ &= \mathbf{E}_x \left[\prod_{k=0}^n g_k(Z_k(f)) \right] \mathbf{E}_x [g_{n+1}(Z_0(f))], \end{aligned}$$

en utilisant le fait que $X_{T_{n+1}} = x$.

On en déduit l'égalité recherchée au rang $n+1$.

En appliquant l'identité (6.1) à toute application g_k mesurable bornée sur \mathbb{R}_+ et $g_l = 1$ pour tout $l \neq k$, on obtient

$$\mathbf{E}_x [g_k(Z_k(f))] = \mathbf{E}_x [g_k(Z_0(f))].$$

On en déduit d'une part que $Z_k(f)$ a même loi que $Z_0(f)$, d'autre part que pour toute suite $(g_n)_{n \in \mathbb{N}}$ de fonctions mesurables bornées sur \mathbb{R}_+ et pour tout $n \in \mathbb{N}$,

$$\mathbf{E}_x \left[\prod_{k=0}^n g_k(Z_k(f)) \right] = \prod_{k=0}^n \mathbf{E}_x [g_k(Z_k(f))],$$

ce qui montre que les variables aléatoires $(Z_k(f))_{k \geq 0}$ sont indépendantes.

Ainsi, les variables aléatoires $(Z_n(f))_{n \geq 0}$ sont indépendantes et de même loi.

• Etape 3 : Application de la loi forte des grands nombres.

Comme la chaîne de Markov X est récurrente irréductible, on sait que toutes les mesures invariantes sont proportionnelles. En notant ν_x la mesure sur E définie par

$$\forall y \in E, \quad \nu_x(y) = \mathbf{E}_x \left[\sum_{k=0}^{H_x-1} \mathbb{1}_{\{X_k=y\}} \right],$$

on a $\nu_x(x) = 1$ et $\mu = \mu(x) \nu_x$.

Par définition de $Z_0(f)$, on a

$$\begin{aligned} \mathbf{E}_x [Z_0(f)] &= \mathbf{E}_x \left[\sum_{k=0}^{H_x-1} \sum_{y \in E} f(y) \mathbb{1}_{\{X_k=y\}} \right] \\ &= \sum_{y \in E} f(y) \nu_x(y) \\ &= \frac{1}{\mu(x)} \sum_{y \in E} f(y) \mu(y). \end{aligned}$$

Comme les variables aléatoires $(Z_n(f))_{n \in \mathbb{N}}$ sont i.i.d. et dans $L^1(\Omega, \mathcal{F}, \mathbf{P})$, la loi forte des grands nombres implique

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_k(f) \longrightarrow \frac{1}{\mu(x)} \int_E f(y) \mu(dy) \quad \text{lorsque } n \rightarrow \infty.$$

En désignant par $N_x(n)$ le nombre de retours en x effectués par la chaîne de Markov X avant le temps n , on a $T_{N_x(n)} \leq n < T_{N_x(n)+1}$.

On écrit alors (puisque f est à valeurs positives)

$$\frac{1}{N_x(n)} \sum_{k=0}^{T_{N_x(n)}-1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{n-1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{T_{N_x(n)+1}-1} f(X_k).$$

On en déduit

$$\frac{1}{N_x(n)} \sum_{k=0}^{N_x(n)-1} Z_k(f) \leq \frac{1}{N_x(n)} \sum_{k=0}^{n-1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{N_x(n)} Z_k(f).$$

Cet encadrement permet de conclure que \mathbf{P}_x -presque sûrement

$$\frac{1}{N_x(n)} \sum_{k=0}^{n-1} f(X_k) \longrightarrow \frac{1}{\mu(x)} \int_E f(y) \mu(dy) \quad \text{lorsque } n \rightarrow \infty,$$

puis que

$$\frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \longrightarrow \frac{1}{\mu(x)} \int_E f(y) \mu(dy) \quad \text{lorsque } n \rightarrow \infty.$$

On obtient le même résultat pour la fonction g . Ainsi, on peut conclure que

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \longrightarrow \frac{\int_E f(y) \mu(dy)}{\int_E g(y) \mu(dy)} \quad \mathbf{P}_x - \text{p.s.}$$

lorsque n tend vers ∞ .

□

Si on choisit $f = \mathbb{1}_{\{y\}}$ et $g = \mathbb{1}_{\{x\}}$ dans le théorème précédent, on obtient

$$\frac{\text{Nombre de visites en } y \text{ avant } n}{\text{Nombre de visites en } x \text{ avant } n} \rightarrow \frac{\mu(y)}{\mu(x)} \quad \text{p.s.}$$

Corollaire 6.1.2

Soit X une chaîne de Markov irréductible et récurrente positive.
Si μ est l'unique mesure de probabilité invariante, alors on a \mathbf{P}_x -p.s.

$$\frac{1}{n} \sum_{k=0}^n f(X_k) \rightarrow \int f.d\mu$$

lorsque n tend vers ∞ .

Preuve Il suffit de prendre $g = 1$ dans le théorème précédent. \square

Théorème 6.1.3

Soit X une chaîne de Markov récurrente irréductible.

(i) Dans le cas récurrent positif, si on note μ la probabilité invariante alors

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \rightarrow \mu(x) \quad p.s.$$

(ii) Dans le cas récurrent nul, on a

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \rightarrow 0 \quad p.s.$$

Preuve D'après le Corollaire 6.1.2, seul le point (ii) reste à démontrer.

On applique le Théorème 6.1.1 avec $f = \mathbb{1}_{\{x\}}$ et $g = \mathbb{1}_F$ pour $F \subset E$ tel que $\mu(F) < \infty$. On a alors

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \leq \frac{\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}}{\sum_{k=0}^{n-1} \mathbb{1}_F(X_k)} \rightarrow \frac{\mu(x)}{\mu(F)} \quad p.s.$$

On en déduit que pour tout $F \subset E$ tel que $\mu(F) < \infty$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \leq \frac{\mu(x)}{\mu(F)} \quad p.s.$$

Par croissance monotone (prendre une suite de parties $(F_n)_{n \in \mathbb{N}}$ qui croît vers F), on obtient

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} = 0 \quad p.s.$$

On a donc

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} = 0 \quad \text{p.s.}$$

□

6.2 Cas apériodique

Définition 6.2.1

Soit X une chaîne de Markov de matrice de transition Q , et soit $x \in E$ un point récurrent. On appelle période de x , noté $d(x)$, le PGCD de

$$L_x = \{n \geq 0 : Q^n(x, x) > 0\}.$$

Si tous les états ont la même période égale à 1, la chaîne est dite apériodique.

Proposition 6.2.2

Si X est une chaîne de Markov récurrente irréductible, alors tous les points ont la même période.

Preuve Montrons tout d'abord que pour tout $x \in E$, le sous-groupe engendré par L_x est $d(x)\mathbb{Z}$.

Pour tous $n, m \in \mathbb{N}$, on a $Q^{n+m}(x, x) \geq Q^n(x, x) Q^m(x, x)$, ce qui montre que L_x est stable par addition.

$d(x)\mathbb{Z}$ est visiblement un sous-groupe de $(\mathbb{Z}, +)$ qui contient L_x . De plus, par définition de $d(x)$, tout sous-groupe qui contient L_x contient $d(x)\mathbb{Z}$.

On en déduit que $d(x)\mathbb{Z}$ est le sous-groupe engendré par L_x .

Soient $x, y \in E$. Comme la chaîne de Markov est irréductible, il existe deux entiers n_1 et n_2 tels que $Q^{n_1}(x, y) > 0$ et $Q^{n_2}(y, x) > 0$ ($x \rightsquigarrow y$ et $y \rightsquigarrow x$).

Pour tout $n \in L_x$, on a alors $Q^{n_1+n+n_2}(y, y) \geq Q^{n_2}(y, x) Q^n(x, x) Q^{n_1}(x, y) > 0$, ce qui entraîne $n_1 + n + n_2 \in L_y$.

On en déduit que le sous-groupe engendré par L_x est inclus dans celui engendré par L_y , c'est-à-dire $d(x)\mathbb{Z} \subset d(y)\mathbb{Z}$. Ainsi $d(y)$ divise $d(x)$.

Par symétrie entre x et y , on conclut que $d(x) = d(y)$.

□

Proposition 6.2.3

Soit X une chaîne de Markov récurrente irréductible et apériodique. Alors pour tous $x, y \in E$, il existe un entier n_0 tel que pour tout $n \geq n_0$, $Q^n(x, y) > 0$.

Preuve Supposons d'abord que $x = y$.

Comme $d(x) = 1$, il existe $n_1 \in \mathbb{N}$ tel que $Q^{n_1}(x, x) > 0$ et $Q^{n_1+1}(x, x) > 0$.

- Si $n_1 = 0$, on montre facilement par récurrence que pour tout $n \geq 0$, $Q^n(x, x) > 0$.
- Si $n_1 \geq 1$, alors pour tout $j \in \{0, \dots, n_1 - 1\}$, on a

$$Q^{n_1^2+j}(x, x) = Q^{(n_1-j)n_1+j(n_1+1)}(x, x) \geq (Q^{n_1}(x, x))^{n_1-j} (Q^{n_1+1}(x, x))^j > 0.$$

En posant $n_0 = n_1^2$, on en déduit que pour tout $j \geq 0$ (distinguer les cas $k.n_1 \leq j \leq (k+1).n_1 - 1$ pour tout $k \in \mathbb{N}$),

$$Q^{n_0+j}(x, x) > 0.$$

Dans le cas où $x \neq y$, comme la chaîne est irréductible, il existe $n_2 \in \mathbb{N}$ tel que $Q^{n_2}(x, y) > 0$. D'après le cas précédent, il existe deux entiers n_0 et n'_0 tels que

$$\forall n \geq n_0, \quad Q^n(x, x) > 0$$

et

$$\forall n \geq n'_0, \quad Q^n(y, y) > 0.$$

Pour tout $n \geq n_0 + n'_0 + n_2$, en écrivant $n = i + j + n_2$ avec $i \geq n_0$ et $j \geq n'_0$, on a

$$Q^n(x, y) \geq Q^i(x, x) Q^{n_2}(x, y) Q^j(y, y) > 0.$$

□

Théorème 6.2.4

Soit X une chaîne de Markov irréductible, récurrente positive et apériodique. Alors, si μ désigne l'unique probabilité invariante, on a pour tout $x \in E$,

$$\lim_{n \rightarrow \infty} \mathbf{P}_x(X_n = y) = \mu(y).$$

Preuve Cf. [15] Théorème 2.6.4 p. 53. □

Troisième partie

Quelques processus à temps continu (2 séances)

Chapitre 7

Processus gaussiens

Pour définir un processus à temps continu (c'est-à-dire sans se restreindre au cas où l'ensemble \mathcal{T} des indices est au plus dénombrable), deux axes peuvent être considérés :

- Approximation par des processus indexés par \mathbb{N} ,
- Définition directe d'un processus à temps continu par sa loi.

7.1 Définition et existence

Un prolongement de la notion de vecteur aléatoire gaussien est celle de processus gaussien.

Définition 7.1.1

Un processus stochastique $\{X_t; t \in \mathcal{T}\}$ est dit processus gaussien si pour tout $n \in \mathbb{N}^*$, et tous $t_1, \dots, t_n \in \mathcal{T}$, le vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ est gaussien.

A tout processus gaussien $\{X_t; t \in \mathcal{T}\}$, on peut associer

- une fonction moyenne $\mu : t \mapsto \mu(t) = \mathbf{E}[X_t]$,
- et une fonction de covariances $\Sigma : (s, t) \mapsto \Sigma(s, t) = \text{Cov}(X_s, X_t)$.

Rappelons qu'une fonction $f : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ est dite *symétrique* si $f(s, t) = f(t, s)$ pour tous $s, t \in \mathcal{T}$ et qu'elle est dite *semi-définie positive* si pour tous $s_1, \dots, s_n \in \mathcal{T}$ et tous $\xi_1, \dots, \xi_n \in \mathbb{R}$,

$$\sum_{i=1}^n \sum_{j=1}^n \xi_i f(s_i, s_j) \xi_j \geq 0.$$

Lemme 7.1.2

La fonction de covariances d'un processus gaussien est symétrique et semi-définie positive.

Preuve Soit $\Sigma : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ la fonction de covariances d'un processus gaussien $\{X_t; t \in \mathcal{T}\}$.

Pour tous $s, t \in \mathcal{T}$, on a $\Sigma(s, t) = \text{Cov}(X_s, X_t) = \text{Cov}(X_t, X_s) = \Sigma(t, s)$. L'application Σ est donc symétrique.

Pour tous $t_1, \dots, t_n \in \mathcal{T}$ et tous $\xi_1, \dots, \xi_n \in \mathbb{R}$, on a

$$\sum_{i,j} \xi_i \Sigma(t_i, t_j) \xi_j = \sum_{i,j} \xi_i \text{Cov}(X_{t_i}, X_{t_j}) \xi_j = \text{Var} \left(\sum_{i=1}^n \xi_i X_{t_i} \right) \geq 0.$$

L'application Σ est donc semi-définie positive. \square

Le résultat suivant montre que ces deux propriétés caractérisent la loi d'un processus gaussien.

Théorème 7.1.3

Etant donné un ensemble abstrait \mathcal{T} , une fonction quelconque $\mu : \mathcal{T} \rightarrow \mathbb{R}$ et une fonction $\Sigma : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ symétrique et définie positive, il existe un processus gaussien $X = \{X_t; t \in \mathcal{T}\}$ de fonction moyenne μ et de fonction de covariances Σ .

Preuve La preuve de ce résultat repose sur le résultat analogue pour les vecteurs gaussiens.

On montre le résultat dans le cas où $\mu = 0$. On montre ainsi l'existence d'un processus gaussien $\{\tilde{X}_t; t \in \mathcal{T}\}$ de moyenne nulle et de fonction de covariance Σ . Le processus final est défini par $X_t = \tilde{X}_t + \mu(t)$ pour tout $t \in \mathcal{T}$.

Sans perte de généralité, on suppose dans la suite $\mu = 0$.

Pour tout sous-ensemble fini $I = \{t_1, \dots, t_n\}$ de \mathcal{T} , on considère la matrice Σ_I de taille $n \times n$ définie par

$$\forall i, j \in \{1, \dots, n\}, \quad \Sigma_I(i, j) = \Sigma(t_i, t_j).$$

La matrice Σ_I est symétrique et définie positive, donc on peut construire un vecteur gaussien Z_I de taille n suivant la loi $\mathcal{N}(0, \Sigma_I)$. On considère alors la mesure de probabilité μ_I sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \quad \mu_I(A) = P_{Z_I}(A) = \mathbf{P}(Z_I \in A).$$

μ_I est la loi de Z_I .

Supposons maintenant $n > 1$ et posons $I_1 = \{t_1, t_2, \dots, t_{n-1}\}$. On a $I_1 \subset I$ et Z_{I_1} a la même loi que les $n - 1$ premières coordonnées de Z_I . Il s'ensuit que

$$\forall A_1, \dots, A_{n-1} \in \mathcal{B}(\mathbb{R}), \quad \mu_{I_1}(A_1 \times \dots \times A_{n-1}) = \mu_I(A_1 \times \dots \times A_{n-1} \times \mathbb{R}).$$

La famille de mesures de probabilité $(\mu_I; I \subset \mathcal{T} \text{ fini})$ vérifie donc les conditions de compatibilité du théorème d'extension de Kolmogorov. Il existe donc un processus $\{X_t; t \in \mathcal{T}\}$ dont les distributions fini-dimensionnelles sont les $(\mu_I; I \subset \mathcal{T} \text{ fini})$. X est un processus gaussien de moyenne nulle et de fonction de covariance Σ . \square

7.2 Bruit blanc de \mathbb{R}

Dans cette section, l'ensemble d'indices \mathcal{T} considéré est la tribu de Borel de \mathbb{R} .

Lemme 7.2.1

L'application $\Sigma : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}$ définie par

$$\forall A, B \in \mathcal{T}, \quad \Sigma(A, B) = \lambda(A \cap B) \mathbb{1}_{\lambda(A) < \infty} \mathbb{1}_{\lambda(B) < \infty},$$

où λ désigne la mesure de Lebesgue sur \mathbb{R} , est symétrique et définie positive.

Preuve La symétrie est évidente.

Pour montrer que Σ est définie positive, on écrit, pour $\lambda(A) < \infty$ et $\lambda(B) < \infty$,

$$\Sigma(A, B) = \int_{\mathbb{R}} \mathbb{1}_A(u) \mathbb{1}_B(u) \lambda(du).$$

Le théorème de Fubini montre que pour tous $A_i \in \mathcal{T}$ et tous $\xi_i \in \mathbb{R}$ ($1 \leq i \leq n$)

$$\sum_{i,j} \xi_i \Sigma(A_i, A_j) \xi_j = \int_{\mathbb{R}} \left(\sum_{i=1}^n \mathbb{1}_{A_i}(u) \xi_i \mathbb{1}_{\lambda(A_i) < \infty} \right)^2 \lambda(du) \geq 0.$$

\square

Le lemme précédent permet de définir le bruit blanc de \mathbb{R} par :

Définition 7.2.2

On appelle bruit blanc sur \mathbb{R} , un processus gaussien $\mathbb{W} = \{\mathbb{W}(A); A \in \mathcal{T}\}$ de moyenne nulle et de fonction de covariances définie par $\Sigma(A, B) = \lambda(A \cap B)$ pour tous $A, B \in \mathcal{T}$.

Théorème 7.2.3

Soit $\mathbb{W} = \{\mathbb{W}(A); A \in \mathcal{T}\}$ un bruit blanc sur \mathbb{R} .

- (i.) Pour tous $A, B \in \mathcal{T}$ disjoints, $\mathbb{W}(A)$ et $\mathbb{W}(B)$ sont indépendants.
- (ii.) Pour tous $A, B \in \mathcal{T}$, $\mathbb{W}(A \cup B) = \mathbb{W}(A) + \mathbb{W}(B) - \mathbb{W}(A \cap B)$ p.s.
- (iii.) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'ensembles de \mathcal{T} disjoints deux à deux et telle que $\sum_{n=1}^{\infty} \lambda(A_n) < \infty$, alors on a

$$\mathbb{W} \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mathbb{W}(A_n) \quad \text{p.s.}$$

Preuve

- (i) Si $A \cap B = \emptyset$, on a $\mathbf{E}[\mathbb{W}(A)\mathbb{W}(B)] = 0$. Comme \mathbb{W} est un processus gaussien, la non-corrélation entre $\mathbb{W}(A)$ et $\mathbb{W}(B)$ implique l'indépendance.
- (ii) Supposons $A \cap B = \emptyset$ dans un premier temps. On pose $D = \mathbb{W}(A \cup B) - \mathbb{W}(A) - \mathbb{W}(B)$. On calcule

$$\begin{aligned} \mathbf{E}[D^2] &= \mathbf{E}[\mathbb{W}(A \cup B)^2] - 2\mathbf{E}[\mathbb{W}(B)\mathbb{W}(A \cup B)] - 2\mathbf{E}[\mathbb{W}(A)\mathbb{W}(A \cup B)] \\ &\quad + \mathbf{E}[\mathbb{W}(A)^2] + 2\mathbf{E}[\mathbb{W}(A)\mathbb{W}(B)] + \mathbf{E}[\mathbb{W}(B)^2] \\ &= (\lambda(A \cup B) - 2\lambda(B) - 2\lambda(A) + \lambda(A) + 2\lambda(A \cap B) + \lambda(B)) \mathbb{1}_{\lambda(A) < \infty} \mathbb{1}_{\lambda(B) < \infty} = 0. \end{aligned}$$

Grâce à l'inégalité de Tchebychev, on en déduit $\mathbf{P}(D = 0) = 1$ ce qui montre le résultat dans le cas $A \cap B = \emptyset$.

Par récurrence, on peut montrer que si $(A_i)_{1 \leq i \leq n}$ est une famille finie de parties 2 à 2 disjointes, on a

$$\mathbb{W} \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n \mathbb{W}(A_i) \quad \text{p.s.}$$

Dans le cas où $A \cap B \neq \emptyset$, on décompose $A \cup B$ suivant la partition

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B).$$

On a alors $\mathbb{W}(A) = \mathbb{W}(A \setminus B) + \mathbb{W}(A \cap B)$, $\mathbb{W}(B) = \mathbb{W}(B \setminus A) + \mathbb{W}(A \cap B)$ et $\mathbb{W}(A \cup B) = \mathbb{W}(A \setminus B) + \mathbb{W}(B \setminus A) + \mathbb{W}(A \cap B)$ presque sûrement.

Le résultat en découle.

- (iii) Pour tout $n \in \mathbb{N}^*$, on définit $M_n = \sum_{i=1}^n \mathbb{W}(A_i)$ et $\epsilon_n = \mathbb{W}(\bigcup_{i=1}^{\infty} A_i) - M_n$. Grâce au point précédent, on a $M_n = \mathbb{W}(\bigcup_{i=1}^n A_i)$ et $\epsilon_n = \mathbb{W}(\bigcup_{i=n+1}^{\infty} A_i)$ presque sûrement. D'autre part,

$$\mathbf{E}[\epsilon_n^2] = \lambda \left(\bigcup_{i=n+1}^{\infty} A_i \right) = \sum_{i=n+1}^{\infty} \lambda(A_i)$$

car les A_i sont disjoints. Par hypothèse, le dernier terme tend vers 0 lorsque $n \rightarrow \infty$. On en déduit que ϵ_n tend vers 0 dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$, c'est-à-dire M_n tend vers $\mathbb{W}(\bigcup_{n=1}^{\infty} A_n)$ dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$.

Pour montrer la convergence presque sûre, on remarque que $\{M_n; n \in \mathbb{N}\}$ est une martingale, telle que

$$\mathbf{E}[M_n^2] = \sum_{i=1}^n \lambda(A_i)$$

qui est bornée. La martingale $\{M_n; n \in \mathbb{N}\}$ bornée dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$, converge donc dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$ et presque sûrement. Comme on a déjà déterminé la limite dans $L^2(\Omega, \mathcal{F}, \mathbf{P})$, le résultat en découle.

□

7.3 Processus isonormal de \mathbb{R}

Les propriétés du bruit blanc de \mathbb{R} permettent de définir des quantités $W(h) = \int_{\mathbb{R}} h(t) \mathbb{W}(dt)$ pour toute fonction $f \in L^2(\lambda)$.

Pour cela, la première idée consiste en la construction de ces intégrales pour des fonctions étagées et d'étendre à $L^2(\lambda)$ par densité (mais attention à l'accumulation non dénombrable des 'p.s.' qui peut devenir un ensemble de mesure nulle).

Heureusement, on peut définir cette intégrale directement.

Définition 7.3.1

On appelle processus isonormal de \mathbb{R} tout processus gaussien $W = \{W(f); f \in L^2(\lambda)\}$ de moyenne nulle et de fonction de covariances vérifiant

$$\forall h_1, h_2 \in L^2(\lambda), \quad \mathbf{E}[W(h_1) W(h_2)] = \int_{\mathbb{R}} h_1(u) h_2(u) \lambda(du).$$

De plus, pour tous $\alpha, \beta \in \mathbb{R}$ et toutes $f, g \in L^2(\lambda)$,

$$W(\alpha f + \beta g) = \alpha W(f) + \beta W(g) \quad p.s.$$

Quelle que soit la manière de construire le processus isonormal ou intégrale par rapport au bruit blanc, on a la caractérisation

$$\forall A \in \mathcal{T}, \quad \mathbb{W}(A) = W(\mathbb{1}_A).$$

Chapitre 8

Introduction au mouvement brownien

8.1 Définition comme processus gaussien

Définition 8.1.1

Le mouvement brownien est un processus gaussien $B = \{B_t; t \in \mathbb{R}_+\}$ de moyenne nulle et de fonction de covariances

$$\forall s, t \in \mathbb{R}_+; \quad \Sigma(s, t) = s \wedge t.$$

Cette définition du mouvement brownien est valide car la fonction Σ est définie positive. Pour voir cela, on remarque que

$$\Sigma(s, t) = \lambda([0, s] \cap [0, t])$$

où λ est la mesure de Lebesgue sur \mathbb{R} .

Théorème 8.1.2 (*Représentation de Centsov*)

Etant donné un bruit blanc \mathbb{W} sur \mathbb{R} , le processus $B = \{B_t; t \in \mathbb{R}_+\}$ défini par $B_t = \mathbb{W}([0, t])$ pour tout $t \in \mathbb{R}_+$ est un mouvement brownien.

Preuve Il suffit de calculer la fonction de covariance de B . \square

En conséquence de cette représentation du mouvement brownien et en utilisant la caractérisation entre bruit blanc et processus isonormal, on a $B_t = W(\mathbb{1}_{[0,t]})$ pour tout $t \in \mathbb{R}_+$. Sous forme intégrale, on peut alors écrire

$$\forall t \in \mathbb{R}_+, \quad B_t = \int_{[0,t]} \mathbb{W}(ds)$$

ce qui amène à considérer le mouvement brownien comme intégrale du bruit blanc.

Grâce à cette observation, pour toute fonction $f \in L^2(\mathbb{R})$, on peut définir

$$\int f(s) dB_s = \int f(s) \mathbb{W}(ds) = W(f).$$

De même, pour toute fonction $f \in L^2_{loc}(\mathbb{R})$ et tout $t \in \mathbb{R}_+$,

$$\int_{[0,t]} f(s) dB_s = \int_{[0,t]} f(s) \mathbb{W}(ds) = W(\mathbb{1}_{[0,t]}f).$$

Dans une certaine mesure, le processus isonormal peut être considéré comme la première pierre de la construction de l'intégrale par rapport au mouvement brownien. On l'appelle *intégrale de Wiener*.

8.2 Quelques problématiques de changement d'échelles

8.2.1 Convergence faible

Définition 8.2.1

Soit (S, ρ) un espace métrique muni de sa tribu de Borel $\mathcal{B}(S)$. Soit $(\mathbf{P}_n)_{n \in \mathbb{N}}$ et \mathbf{P} des mesures de probabilité sur $(S, \mathcal{B}(S))$. On dit que $(\mathbf{P}_n)_{n \in \mathbb{N}}$ converge faiblement vers \mathbf{P} si

$$\lim_{n \rightarrow \infty} \int_S f(s) \mathbf{P}_n(ds) = \int_S f(s) \mathbf{P}(ds)$$

pour toute fonction $f : S \rightarrow \mathbb{R}$ continue bornée.

Définition 8.2.2

Soient $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)_{n \in \mathbb{N}}$ et $(\Omega, \mathcal{F}, \mathbf{P})$ des espaces de probabilité et $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires sur ces espaces et à valeurs dans l'espace métrique (S, ρ) .

On dit que $(X_n)_{n \in \mathbb{N}}$ converge en distribution vers X , si la suite des lois $(\mathbf{P}_n \circ X_n^{-1})_{n \in \mathbb{N}}$ converge faiblement vers la loi $\mathbf{P} \circ X^{-1}$, c'est-à-dire si

$$\lim_{n \rightarrow \infty} \mathbf{E}_n[f(X_n)] = \mathbf{E}[f(X)]$$

pour toute fonction $f : S \rightarrow \mathbb{R}$ continue bornée.

Les notions suivantes de relative compacité et de tension sont essentielles pour caractériser la convergence faible.

Définition 8.2.3

Soit (S, ρ) un espace métrique et soit Π une famille de mesures de probabilité sur $(S, \mathcal{B}(S))$.

- Π est dite relativement compacte si toute suite de Π admet une sous-suite qui converge faiblement.
- Π est dite tendue si pour tout $\epsilon > 0$, il existe un ensemble compact $K \subset S$ tel que

$$\forall \mathbf{P} \in \Pi; \quad \mathbf{P}(K) \geq 1 - \epsilon.$$

Une famille de v.a. $(X_\alpha)_\alpha$ est dite relativement compacte (resp. tendue) si la famille de leur loi est relativement compacte (resp. tendue).

Théorème 8.2.4 (Prohorov)

Soit Π une famille de mesures de probabilité sur un espace métrique complet et séparable S . Alors, Π est relativement compacte si et seulement si elle est tendue.

Rappelons que la loi d'un processus stochastique est engendrée par ses distributions fini-dimensionnelles (Théorème de Kolmogorov). Tout processus stochastique $X = \{X_t; t \in \mathbb{R}_+\}$ à valeurs dans E , peut être considéré comme une variable aléatoire à valeurs dans l'espace des trajectoires

$$X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (E^{\mathbb{R}_+}, \mathcal{G}, P_X),$$

où \mathcal{G} est la tribu de Kolmogorov et P_X la mesure image de \mathbf{P} par X .

Les distributions fini-dimensionnelles de X sont les mesures de probabilité

$$\begin{aligned} P_{t_1, \dots, t_d} : \quad \mathcal{E}^{\otimes d} &\longrightarrow [0, 1] \\ B_1 \times \dots \times B_d &\longmapsto \mathbf{P}(X_{t_1} \in B_1, \dots, X_{t_d} \in B_d), \end{aligned}$$

où $d \geq 1$ est un entier quelconque et t_1, \dots, t_d sont des réels positifs quelconques.

Les mesures P_{t_1, \dots, t_d} sont déterminées par les quantités $\mathbf{E}[f(X_{t_1}, \dots, X_{t_d})]$, pour toute $f : E^d \rightarrow \mathbb{R}$ mesurable bornée.

Pour tout sous-ensemble fini $\{t_1, \dots, t_d\}$ de \mathbb{R}_+ , on définit la fonction projection $\pi_{t_1, \dots, t_d} : C(\mathbb{R}_+, E) \rightarrow E^d$ par

$$\pi_{t_1, \dots, t_d}(\omega) = (\omega(t_1), \dots, \omega(t_d)).$$

Si une fonction $f : E^d \rightarrow \mathbb{R}$ est continue et bornée, alors l'application composée $f \circ \pi_{t_1, \dots, t_d}$ l'est également. Ainsi, si une suite de processus continus $(X^{(n)})_{n \in \mathbb{N}}$ converge faiblement vers un processus X alors

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{E}_n[f(X_{t_1}^{(n)}, \dots, X_{t_d}^{(n)})] &= \lim_{n \rightarrow \infty} \mathbf{E}_n[f \circ \pi_{t_1, \dots, t_d}(X^{(n)})] \\ &= \mathbf{E}[f \circ \pi_{t_1, \dots, t_d}(X)] \\ &= \mathbf{E}[f(X_{t_1}, \dots, X_{t_d})]. \end{aligned}$$

Ceci montre que les distributions fini-dimensionnelles des $X^{(n)}$ convergent vers celles de X .

La réciproque de ce résultat est fausse en général. Cependant elle devient vraie si on ajoute l'hypothèse que les lois des $(X^{(n)})_{n \in \mathbb{N}}$ sont tendues.

Théorème 8.2.5

Soit $(X^{(n)})_{n \in \mathbb{N}}$ une suite tendue de processus continus dont les distributions fini-dimensionnelles convergent.

Alors les mesures de probabilités sur $C(\mathbb{R}_+)$ induites par les $X^{(n)}$ convergent faiblement vers une mesure P , sous laquelle le processus coordonnées $W = \{W_t = \omega(t); t \in \mathbb{R}_+\}$ vérifie

$$(X_{t_1}^{(n)}, \dots, X_{t_d}^{(n)}) \xrightarrow{\mathcal{D}} (W_{t_1}, \dots, W_{t_d})$$

pour tous $t_1, \dots, t_d \in \mathbb{R}_+$.

8.2.2 Principe d'invariance de Donsker

Soit $(\xi_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires i.i.d. de moyenne nulle et de variance $\sigma^2 > 0$. On considère la suite des sommes partielles

$$S_0 = 0, \quad S_n = \sum_{k=1}^n \xi_k \quad \text{pour } n \geq 1.$$

On définit alors le processus continu $Y = \{Y_t; t \in \mathbb{R}_+\}$ par

$$\forall t \in \mathbb{R}_+; \quad Y_t = S_{[t]} + (t - [t])\xi_{[t]+1},$$

puis on renormalise Y pour obtenir un processus $\{X_t^{(n)}; t \in \mathbb{R}_+\}$ défini par

$$\forall t \in \mathbb{R}_+; \quad X_t^{(n)} = \frac{1}{\sigma\sqrt{n}} Y_{nt}.$$

Théorème 8.2.6 (Principe d'invariance de Donsker)

La suite de processus $(X^{(n)})_{n \in \mathbb{N}}$ converge en distribution vers le mouvement brownien.

8.3 Trajectoire du mouvement brownien

Définition 8.3.1

On appelle mouvement brownien standard un processus gaussien centré continu $\{B_t; t \in \mathbb{R}_+\}$, de fonction de covariance $\text{Cov}(B_s, B_t) = \mathbf{E}[B_s B_t] = s \wedge t$, pour tous $s, t \in \mathbb{R}_+$.

8.3.1 Existence d'une modification continue

Théorème 8.3.2 (Kolmogorov-Centsov)

Soit $X = \{X_t; 0 \leq t \leq T\}$ un processus stochastique sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ tel que

$$\forall s, t \in [0, T]; \quad \mathbf{E}[|X_t - X_s|^\alpha] \leq C |t - s|^{1+\beta}$$

pour des constantes positives α, β et C .

Alors, il existe une modification continue $\tilde{X} = \{\tilde{X}_t; 0 \leq t \leq T\}$, qui est localement Hölder-continue d'exposant γ pour tout $\gamma \in]0, \frac{\beta}{\alpha}[$, i.e.

$$\mathbf{P} \left(\left\{ \omega : \sup_{0 < t-s < h(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} \leq \delta \right\} \right) = 1$$

où $h(\omega)$ est une variable aléatoire positive presque sûrement et $\delta > 0$ une constante.

Preuve Sans perte de généralité, on suppose $T = 1$.

Pour tout $\epsilon > 0$, on a

$$\mathbf{P}(|X_t - X_s| \geq \epsilon) \leq \frac{\mathbf{E}[|X_t - X_s|^\alpha]}{\epsilon^\alpha} \leq C \epsilon^{-\alpha} |t - s|^{1+\beta}.$$

On en déduit que $X_s \rightarrow X_t$ en probabilité lorsque $s \rightarrow t$.

On prend ensuite $t = k \cdot 2^{-n}$, $s = (k-1) \cdot 2^{-n}$ et $\epsilon = 2^{-\gamma n}$ (où $0 < \gamma < \beta/\alpha$). On obtient

$$\mathbf{P}(|X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \geq 2^{-\gamma n}) \leq C 2^{-n(1+\beta-\alpha\gamma)}.$$

Ainsi, on peut majorer

$$\begin{aligned} \mathbf{P} \left(\max_{1 \leq k \leq 2^n} |X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \geq 2^{-\gamma n} \right) &= \mathbf{P} \left(\bigcup_{1 \leq k \leq 2^n} \{|X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \geq 2^{-\gamma n}\} \right) \\ &= \sum_{k=1}^{2^n} \mathbf{P}(|X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \geq 2^{-\gamma n}) \\ &\leq C 2^{-n(\beta-\alpha\gamma)}. \end{aligned}$$

Cette inégalité montre que la série $\sum_{n \in \mathbb{N}} \mathbf{P}(\max_{1 \leq k \leq 2^n} |X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \geq 2^{-\gamma n})$ converge, donc on peut appliquer le lemme de Borel-Cantelli : Il existe un ensemble $\Omega^* \in \mathcal{F}$ tel que $\mathbf{P}(\Omega^*) = 1$ et une variable aléatoire n^* à valeurs dans \mathbb{N} tels que pour tout $\omega \in \Omega^*$,

$$\forall n \geq n^*(\omega); \quad \max_{1 \leq k \leq 2^n} |X_{k \cdot 2^{-n}} - X_{(k-1) \cdot 2^{-n}}| \leq 2^{-\gamma n}. \quad (8.1)$$

Pour tout entier $n \geq 1$, on considère $D_n = \{(k \cdot 2^{-n}); 1 \leq k \leq 2^n\} \subset [0, 1]$. L'ensemble $D = \bigcup_{n \geq 1} D_n$ est alors l'ensemble des rationnels dyadiques de $[0, 1]$.

Fixons $\omega \in \Omega^*$ et $n \geq n^*(\omega)$. Montrons que pour tout $m > n$, on a

$$\forall s, t \in D_m, \quad 0 < t - s < 2^{-n}; \quad |X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^m 2^{-\gamma j}. \quad (8.2)$$

Pour $m = n + 1$, on doit juste montrer le résultat pour $t = k \cdot 2^{-n}$ et $s = (k - 1) \cdot 2^{-n}$, qui résulte immédiatement de (8.1). Supposons (8.2) vérifiée pour $m = n + 1, n + 2, \dots, M - 1$. On considère $s, t \in D_M$ tels que $s < t$ et on pose

$$t^1 = \max(u \in D_{M-1}; u \leq t) \quad \text{et} \quad s^1 = \max(u \in D_{M-1}; u \geq s).$$

On a alors

$$\begin{cases} s \leq s^1 \leq t^1 \leq t; \\ s^1 - s \leq 2^{-M} \quad \text{et} \quad t - t^1 \leq 2^{-M}. \end{cases}$$

D'après (8.1), on a

$$|X_{s^1}(\omega) - X_s(\omega)| \leq 2^{-\gamma M} \quad \text{et} \quad |X_t(\omega) - X_{t^1}(\omega)| \leq 2^{-\gamma M}.$$

En appliquant (8.2) vérifiée pour $m = M - 1$ par hypothèse de récurrence, on a

$$|X_{t^1}(\omega) - X_{s^1}(\omega)| \leq 2 \sum_{j=n+1}^{M-1} 2^{-\gamma j}.$$

Par inégalité triangulaire, on déduit que (8.2) est vérifiée pour $m = M$.

On montre alors que pour tout $\omega \in \Omega^*$, $\{X_t(\omega); t \in D\}$ est uniformément continu en t .

Pour tous $s, t \in D$ tels que $0 < t - s < 2^{-n^*(\omega)}$, on choisit $n \geq n^*(\omega)$ tel que $2^{-(n+1)} \leq t - s < 2^{-n}$. On a alors

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{\infty} 2^{-\gamma j} \leq \frac{2}{1 - 2^{-\gamma}} |t - s|^{\gamma}.$$

On définit alors \tilde{X} de la manière suivante :

- Pour tout $\omega \notin \Omega^*$, on pose $\tilde{X}_t(\omega) = 0$ pour tout $t \in [0, 1]$;
- pour tout $\omega \in \Omega^*$, on pose $\tilde{X}_t(\omega) = X_t(\omega)$ pour tout $t \in D$;
- et pour tout $\omega \in \Omega^*$ et $t \in [0, 1] \setminus D$, on pose $\tilde{X}_t(\omega) = \lim_n X_{s_n}(\omega)$ où $(s_n)_{n \in \mathbb{N}}$ est une suite dans D qui converge vers t .

On vérifie que \tilde{X} est une modification de X . \square

Comme $B_t - B_s$ est une variable aléatoire gaussienne centrée, pour tout $n \in \mathbb{N}^*$, il existe une constante $\lambda_n > 0$ telle que

$$\mathbf{E}[(B_t - B_s)^{2n}] = \lambda_n \cdot (\mathbf{E}[(B_t - B_s)^2])^n.$$

Par ailleurs, on a

$$\mathbf{E}[(B_t - B_s)^2] = |t - s|$$

pour tous $s, t \in \mathbb{R}_+$, donc

$$\mathbf{E}[(B_t - B_s)^{2n}] = \lambda_n |t - s|^n.$$

En choisissant $n = 2$, on peut appliquer le théorème précédent pour montrer l'existence d'une modification continue du mouvement brownien.

8.4 Premières propriétés du mouvement brownien

Proposition 8.4.1

Soit $B = \{B_t; t \in \mathbb{R}_+\}$ un mouvement brownien.

- (i) Les accroissements de B sont indépendants :
Pour toute suite finie $0 = t_0 < t_1 < \dots < t_n$, les variables aléatoires $B_{t_k} - B_{t_{k-1}}$ sont indépendantes ;
- (ii) Les accroissements de B sont stationnaires : Plus précisément, pour tous $s, t \in \mathbb{R}_+$ tels que $s < t$, la loi de la v.a. $B_t - B_s$ est gaussienne centrée de variance $t - s$.

Preuve Soit $\{\mathbb{W}(A); A \in \mathcal{B}(\mathbb{R})\}$ un bruit blanc de \mathbb{R} . Le processus $X = \{X_t; t \in \mathbb{R}_+\}$ défini par $X_t = \mathbb{W}([0, t])$ pour tout $t \in \mathbb{R}_+$, a même loi que le mouvement brownien B .

Il suffit donc de montrer les assertions (i) et (ii) pour le processus X .

- (i) Pour $0 = t_0 < t_1 < \dots < t_n$, on a $X_{t_k} - X_{t_{k-1}} = \mathbb{W}([t_{k-1}, t_k])$. Comme les intervalles $[t_{k-1}, t_k]$ sont 2 à 2 disjoints, les variables aléatoires $\mathbb{W}([t_{k-1}, t_k])$ (où $1 \leq k \leq n$) sont indépendantes.
- (ii) Pour $s < t$, $X_t - X_s = \mathbb{W}([s, t])$ est une variable aléatoire gaussienne de moyenne nulle et de variance $\lambda([s, t]) = t - s$.

\square

Proposition 8.4.2

Soit $B = \{B_t; t \in \mathbb{R}_+\}$ un mouvement brownien.

- (i) (Homogénéité en temps). Pour tout $s > 0$, le processus $\{B_{t+s} - B_s; t \in \mathbb{R}_+\}$ est un mouvement brownien indépendant de $\sigma(B_u; u \leq s)$.
- (ii) (Symétrie). Le processus $-B$ est un mouvement brownien.
- (iii) (Autosimilarité). Pour tout $c > 0$, les processus $\{cB_{t/c^2}; t \in \mathbb{R}_+\}$ est un mouvement brownien.
- (iv) (Inversion du temps). Le processus X défini par $X_0 = 0$ et $X_t = tB_{1/t}$ pour tout $t > 0$ est un mouvement brownien.

Preuve

- (i) Soit $\{\mathbb{W}(A); A \in \mathcal{B}(\mathbb{R})\}$ un bruit blanc de \mathbb{R} . Le processus $X = \{X_t; t \in \mathbb{R}_+\}$ défini par $X_t = \mathbb{W}([0, t])$ pour tout $t \in \mathbb{R}_+$, a même loi que le mouvement brownien B .

Il suffit de montrer l'assertion pour le processus X .

On a $X_{t+s} - X_s = \mathbb{W}([s, s+t])$. Comme \mathbb{W} est un processus gaussien centré, on en déduit que $\{X_{t+s} - X_s; t \in \mathbb{R}_+\}$ est un processus gaussien centré. Sa fonction de covariance est

$$\begin{aligned} \mathbf{E}[(X_{t+s} - X_s)(X_{t'+s} - X_s)] &= \mathbf{E}[\mathbb{W}([s, s+t])\mathbb{W}([s, s+t'])]] \\ &= \lambda([s, s+t] \cap [s, s+t']) \\ &= \lambda([0, t] \cap [0, t']) = t \wedge t'. \end{aligned}$$

Donc $\{X_{t+s} - X_s; t \in \mathbb{R}_+\}$ est un mouvement brownien.

Comme $X_{t+s} - X_s$ est indépendante des variables X_u pour tout $u \leq s$, on en déduit que $X_{t+s} - X_s$ est indépendante de $\sigma(X_u; u \leq s)$.

- (ii) On vérifie sans peine que le processus $\{X_t; t \in \mathbb{R}_+\}$ défini par $X_t = -B_t$ pour tout $t \in \mathbb{R}_+$ est un mouvement brownien.
- (iii) On considère le processus $\{X_t; t \in \mathbb{R}_+\}$ défini par $X_t = c B_{t/c^2}$ pour tout $t \in \mathbb{R}_+$. Toute combinaison linéaire finie de variables aléatoires X_t ($t \in \mathbb{R}_+$) est une combinaison linéaire finie de variables B_s ($s \in \mathbb{R}_+$), donc suit une loi normale. Le processus X est donc gaussien.

Pour tout $t \in \mathbb{R}_+$, on a $\mathbf{E}[X_t] = c \mathbf{E}[B_{t/c^2}] = 0$.

Pour tous $s, t \in \mathbb{R}_+$, on a $\mathbf{E}[X_s X_t] = c^2 \mathbf{E}[B_{s/c^2} B_{t/c^2}] = s \wedge t$.

On conclut donc que X est un mouvement brownien.

- (iv) La preuve, qui est analogue au point précédent, est laissée en exercice.

□

Les résultats suivants seront prouvés dans le cadre d'un cours de *mouvement brownien et calcul stochastique* de niveau Master 2...

Le premier d'entre eux montre que les trajectoires du mouvement brownien standard sont continues mais irrégulières. Elles ne sont pas dérivables et ne sont pas à variations bornées. Par conséquent, il est impossible de définir une intégrale par rapport au mouvement brownien ω par ω (intégrale de Stieljes).

Proposition 8.4.3

Le mouvement brownien est presque sûrement à variations infinies sur tout intervalle.

L'intégrale stochastique par rapport au mouvement brownien s'inscrit dans le cadre général de l'intégrale stochastique par rapport à des *semi-martingales*.

Définition 8.4.4

Un processus $\{X_t; t \in \mathbb{R}_+\}$ est une martingale par rapport à la filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ si les trois conditions suivantes sont vérifiées :

- (i) Pour tout $t \in \mathbb{R}_+$, X_t est \mathcal{F}_t -mesurable ;
- (ii) Pour tout $t \in \mathbb{R}_+$, $X_t \in L^1(\Omega, \mathcal{F}, \mathbf{P})$;
- (iii) Pour tous $s, t \in \mathbb{R}_+$ tels que $s < t$, on a

$$X_s = \mathbf{E}[X_t \mid \mathcal{F}_s] \quad p.s.$$

Ainsi, on termine ce chapitre (et donc ce cours) par le résultat suivant :

Proposition 8.4.5

Le mouvement brownien est une martingale par rapport à sa filtration naturelle.

Comme dans le cas discret, les martingales à temps continu vérifient un théorème d'arrêt. On en déduit

Théorème 8.4.6

Soient $B = \{B_t; t \in \mathbb{R}_+\}$ un mouvement brownien standard et T un temps d'arrêt borné. On a

$$\mathbf{E}[B_T] = \mathbf{E}[B_0].$$

Application au temps de sortie d'une bande $[a, b]$.

Bibliographie

- [1] M. Benaïm et N. El Karoui, *Promenade aléatoire*, Editions de l'Ecole Polytechnique, Ellipse, 2004.
- [2] Delachellerie et P. A. Meyer, *Probabilités et Potentiels*, Hermann.
- [3] R. J. Durrett, *Probability : Theory and Examples*, 2nd edition, Duxbury Press, 1995.
- [4] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. I-II, Wiley, 2nd edition, 1971.
- [5] E. Herbin and P. Lafitte, *Lecture Handout of “Convergence, Integration, Probability” and “Partial Differential Equations”*, Notes des cours de 1ère année de CentraleSupélec, 2019.
- [6] J. Jacod and P. Protter, *Probability Essentials*, Springer, 2000.
- [7] O. Kalenberg, *Foundation of modern probability*, Springer.
- [8] I. Karatsas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, 2nd edition 1997.
- [9] D. Khoshnevisan, *Multiparameter Processes, An Introduction to Random Fields*, Springer, 2002.
- [10] J. Lacroix, P. Priouret et L. Zambotti, *Probabilités approfondies*, Polycopié du Master de mathématiques et applications de l'Université Paris VI, 2007.
- [11] J.-F. Delmas, B. Jourdain et B. Lapeyre, *Cours de processus aléatoires*, Polycopié de l'ENPC, 2001.
- [12] J.-F. Legall, *Cours de processus aléatoires*, Polycopié de l'Ecole Normale Supérieure de Paris, 2005.
- [13] J.-F. Legall, *Intégration, Probabilités et Processus Aléatoires*, Polycopié de l'Ecole Normale Supérieure de Paris, 2006.
- [14] J. Neveu, *Martingales à temps discret*, Masson.
- [15] E. Pardoux, *Processus de Markov et applications*, 2006.
- [16] D. Williams, *Probability with Martingales*, Cambridge, 1991.