



**HAL**  
open science

## Martingales et calcul stochastique

Nils Berglund

► **To cite this version:**

Nils Berglund. Martingales et calcul stochastique. DEA. Université d'Orléans, 2010, pp.91. cel-00443085v2

**HAL Id: cel-00443085**

**<https://cel.hal.science/cel-00443085v2>**

Submitted on 21 Dec 2010 (v2), last revised 7 Jan 2013 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Martingales et calcul stochastique

Master 2 Recherche de Mathématiques

Université d'Orléans

Nils Berglund

Version de Décembre 2010



# Table des matières

<b>I</b>	<b>Processus en temps discret</b>	<b>1</b>
<b>1</b>	<b>Exemples de processus stochastiques</b>	<b>3</b>
1.1	Variables aléatoires i.i.d. . . . . .	3
1.2	Marches aléatoires et chaînes de Markov . . . . .	4
1.3	Urnes de Polya . . . . .	6
1.4	Le processus de Galton–Watson . . . . .	7
1.5	Marches aléatoires auto-évitantes . . . . .	9
1.6	Systèmes dynamiques . . . . .	11
<b>2</b>	<b>Construction générale des processus</b>	<b>13</b>
2.1	Préliminaires et rappels . . . . .	13
2.2	Distributions de dimension finie . . . . .	14
2.3	Noyaux markoviens . . . . .	14
2.4	Le théorème de Ionescu–Tulcea . . . . .	16
<b>3</b>	<b>Filtrations, espérance conditionnelle</b>	<b>19</b>
3.1	Sous-tribus et filtrations . . . . .	19
3.2	Espérance conditionnelle . . . . .	21
3.3	Exercices . . . . .	25
<b>4</b>	<b>Martingales</b>	<b>27</b>
4.1	Définitions et exemples . . . . .	27
4.2	Inégalités . . . . .	28
4.3	Décomposition de Doob, processus croissant . . . . .	29
4.4	Exercices . . . . .	31
<b>5</b>	<b>Temps d’arrêt</b>	<b>33</b>
5.1	Définition et exemples . . . . .	33
5.2	Processus arrêté, inégalité de Doob . . . . .	34
5.3	Exercices . . . . .	36
<b>6</b>	<b>Théorèmes de convergence</b>	<b>39</b>
6.1	Rappel: Notions de convergence . . . . .	39
6.2	Convergence presque sûre . . . . .	40
6.3	Convergence dans $L^p$ , $p > 1$ . . . . .	42
6.4	Convergence dans $L^1$ . . . . .	44
6.5	Loi 0 – 1 de Lévy . . . . .	46
6.6	Exercices . . . . .	47

<b>II</b>	<b>Processus en temps continu</b>	<b>49</b>
<b>7</b>	<b>Le mouvement Brownien</b>	<b>51</b>
7.1	Limite d'échelle d'une marche aléatoire . . . . .	51
7.2	Construction du mouvement Brownien . . . . .	53
7.3	Propriétés de base . . . . .	55
7.4	Temps d'arrêt . . . . .	57
7.5	Mouvement Brownien et martingales . . . . .	59
7.6	Exercices . . . . .	61
<b>8</b>	<b>L'intégrale d'Itô</b>	<b>63</b>
8.1	Définition . . . . .	63
8.2	Propriétés élémentaires . . . . .	66
8.3	Un exemple . . . . .	67
8.4	La formule d'Itô . . . . .	68
8.5	Exercices . . . . .	70
<b>9</b>	<b>Equations différentielles stochastiques</b>	<b>73</b>
9.1	Solutions fortes . . . . .	73
9.2	Existence et unicité de solutions . . . . .	75
9.3	Exercices . . . . .	77
<b>10</b>	<b>Diffusions</b>	<b>79</b>
10.1	La propriété de Markov . . . . .	79
10.2	Semigroupes et générateurs . . . . .	81
10.3	La formule de Dynkin . . . . .	83
10.4	Les équations de Kolmogorov . . . . .	86
10.5	Exercices . . . . .	89

# Partie I

## Processus en temps discret



# Chapitre 1

## Exemples de processus stochastiques

D'une manière générale, un *processus stochastique à temps discret* est simplement une suite  $(X_0, X_1, X_2, \dots)$  de variables aléatoires, définies sur un même espace probabilisé. Avant de donner une définition précise, nous discutons quelques exemples de tels processus.

### 1.1 Variables aléatoires i.i.d.

Supposons que les variables aléatoires  $X_0, X_1, X_2, \dots$  ont toutes la même loi et sont indépendantes. On dit alors qu'elles sont *indépendantes et identiquement distribuées*, abrégé i.i.d. C'est d'une certaine manière la situation la plus aléatoire que l'on puisse imaginer. On peut considérer que les  $X_i$  sont définies chacune dans une copie différente d'un espace probabilisé de base, et que le processus vit dans l'espace produit de ces espaces.

La suite des  $X_i$  en soi n'est pas très intéressante. Par contre la suite des sommes partielles

$$S_n = \sum_{i=0}^{n-1} X_i \quad (1.1.1)$$

admet plusieurs propriétés remarquables. En particulier, rappelons les théorèmes de convergence suivants.

1. La *loi faible des grands nombres*: Si l'espérance  $\mathbb{E}(X_0)$  est finie, alors  $S_n/n$  converge vers  $\mathbb{E}(X_0)$  en probabilité, c'est-à-dire

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{S_n}{n} - \mathbb{E}(X_0) \right| > \varepsilon \right\} = 0 \quad (1.1.2)$$

pour tout  $\varepsilon > 0$ .

2. La *loi forte des grands nombres*: Si l'espérance est finie, alors  $S_n/n$  converge presque sûrement vers  $\mathbb{E}(X_0)$ , c'est-à-dire

$$\mathbb{P} \left\{ \lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbb{E}(X_0) \right\} = 1. \quad (1.1.3)$$

3. Le *théorème de la limite centrale*: Si l'espérance et la variance  $\text{Var}(X_0)$  sont finies, alors  $(S_n - n\mathbb{E}(X_0))/\sqrt{n \text{Var}(X_0)}$  tend en loi vers une variable normale centrée réduite,

c'est-à-dire

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ a \leq \frac{S_n - n\mathbb{E}(X_0)}{\sqrt{n \operatorname{Var}(X_0)}} \leq b \right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (1.1.4)$$

pour tout choix de  $-\infty \leq a < b \leq \infty$ .

Il existe d'autres théorèmes limites, comme par exemple des principes de grandes déviations et des lois du logarithme itéré. Le théorème de Cramér affirme que sous certaines conditions de croissance sur la loi des  $X_i$ , on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left\{ \frac{S_n}{n} > x \right\} = -I(x), \quad (1.1.5)$$

où la *fonction taux*  $I$  est la transformée de Legendre de la fonction génératrice des cumulants de  $X_0$ , à savoir  $\lambda(\theta) = \log \mathbb{E}(e^{\theta X_0})$ . C'est un principe de grandes déviations, qui implique que  $\mathbb{P}\{S_n > nx\}$  décroît exponentiellement avec un taux  $-nI(x)$ .

La loi du logarithme itéré, quant à elle, affirme que

$$\mathbb{P} \left\{ \limsup_{n \rightarrow \infty} \frac{S_n - n\mathbb{E}(X_0)}{\sqrt{n \log(\log n) \operatorname{Var}(X_0)}} = \sqrt{2} \right\} = 1. \quad (1.1.6)$$

Les fluctuations de  $S_n$ , qui sont typiquement d'ordre  $\sqrt{n \operatorname{Var}(X_0)}$  d'après le théorème de la limite centrale, atteignent donc avec probabilité 1 la valeur  $\sqrt{2n \log(\log n) \operatorname{Var}(X_0)}$ , sans jamais la dépasser.

## 1.2 Marches aléatoires et chaînes de Markov

La *marche aléatoire symétrique sur  $\mathbb{Z}^d$*  est construite de la manière suivante. On pose  $X_0 = 0$  (l'origine de  $\mathbb{Z}^d$ ), puis pour  $X_1$  on choisit l'un des  $2d$  plus proches voisins de l'origine avec probabilité  $1/2d$ , et ainsi de suite. Chaque  $X_n$  est choisi de manière équiprobable parmi les  $2d$  plus proches voisins de  $X_{n-1}$ , indépendamment des variables  $X_{n-2}, \dots, X_0$ . On obtient ainsi une distribution de probabilité sur les lignes brisées de  $\mathbb{Z}^d$ .

La marche symétrique sur  $\mathbb{Z}^d$  est très étudiée et relativement bien comprise. On sait par exemple que

1. en dimensions  $d = 1$  et  $2$ , la marche est *récurrente*, c'est-à-dire qu'elle revient presque sûrement en zéro;
2. en dimensions  $d \geq 3$ , la marche est *transiente*, c'est-à-dire qu'il y a une probabilité strictement positive qu'elle ne revienne jamais en zéro.

Il existe de nombreuses variantes des marches symétriques sur  $\mathbb{Z}^d$  : Marches asymétriques, sur un sous-ensemble de  $\mathbb{Z}^d$  avec réflexion ou absorption au bord, marches sur des graphes, sur des groupes...

Tous ces exemples de marches font partie de la classe plus générale des *chaînes de Markov*. Pour définir une telle chaîne, on se donne un ensemble dénombrable  $\mathcal{X}$  et une *matrice stochastique*  $P = (p_{ij})_{i,j \in \mathcal{X}}$ , satisfaisant les deux conditions

$$p_{ij} \geq 0 \quad \forall i, j \in \mathcal{X} \quad \text{et} \quad \sum_{j \in \mathcal{X}} p_{ij} = 1 \quad \forall i \in \mathcal{X}. \quad (1.2.1)$$

On choisit alors une distribution initiale pour  $X_0$ , et on construit les  $X_n$  de telle manière que

$$\mathbb{P}\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} = \mathbb{P}\{X_{n+1} = j \mid X_n = i\} = p_{ij} \quad (1.2.2)$$

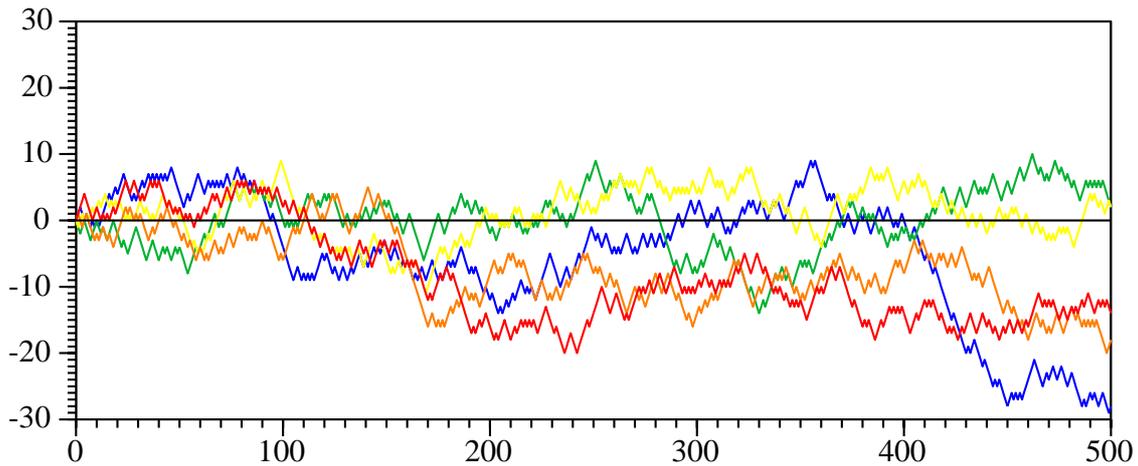


FIGURE 1.1. Cinq réalisations d'une marche aléatoire unidimensionnelle symétrique.

pour tout temps  $n$  et tout choix des  $i_0, \dots, i_{n-1}, i, j \in \mathcal{X}$ . La première égalité s'appelle la *propriété de Markov*. Les chaînes de Markov ont un comportement un peu plus riche que les variables i.i.d. parce que chaque variable dépend de celle qui la précède immédiatement. Toutefois, ces processus ont toujours une mémoire courte, tout ce qui s'est passé plus de deux unités de temps dans le passé étant oublié.

Un exemple intéressant de chaîne de Markov, motivé par la physique, est le *modèle d'Ehrenfest*. On considère deux récipients contenant en tout  $N$  molécules de gaz. Les récipients sont connectés par un petit tube, laissant passer les molécules une à une dans les deux sens. Le but est d'étudier comment évolue la proportion de molécules contenues dans l'un des récipients au cours du temps. Si par exemple ce nombre est nul au début, on s'attend à ce qu'assez rapidement, on atteigne un équilibre dans lequel le nombre de molécules dans les deux récipients est à peu près égal, alors qu'il n'arrivera quasiment jamais qu'on revoie toutes les molécules revenir dans le récipient de droite.

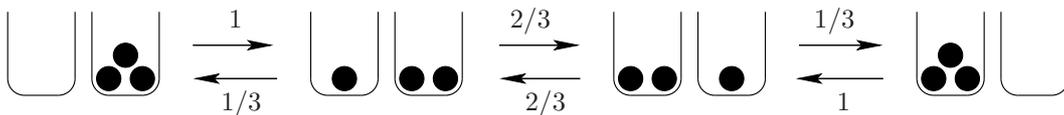


FIGURE 1.2. Le modèle d'urnes d'Ehrenfest, dans le cas de 3 boules.

On modélise la situation en considérant deux urnes contenant en tout  $N$  boules. A chaque unité de temps, l'une des  $N$  boules, tirée uniformément au hasard, passe d'un récipient à l'autre. S'il y a  $i$  boules dans l'urne de gauche, alors on tirera avec probabilité  $i/N$  l'une de ces boules, et le nombre de boules dans l'urne passera de  $i$  à  $i - 1$ . Dans le cas contraire, ce nombre passe de  $i$  à  $i + 1$ . Si  $X_n$  désigne le nombre de boules dans l'urne de gauche au temps  $n$ , l'évolution est donnée par une chaîne de Markov sur  $\{0, 1, \dots, N\}$  de probabilités de transition

$$p_{ij} = \begin{cases} \frac{i}{N} & \text{si } j = i - 1, \\ 1 - \frac{i}{N} & \text{si } j = i + 1, \\ 0 & \text{sinon.} \end{cases} \quad (1.2.3)$$

La Figure 1.3 montre des exemples de suites  $\{X_n/N\}$ . On observe effectivement que

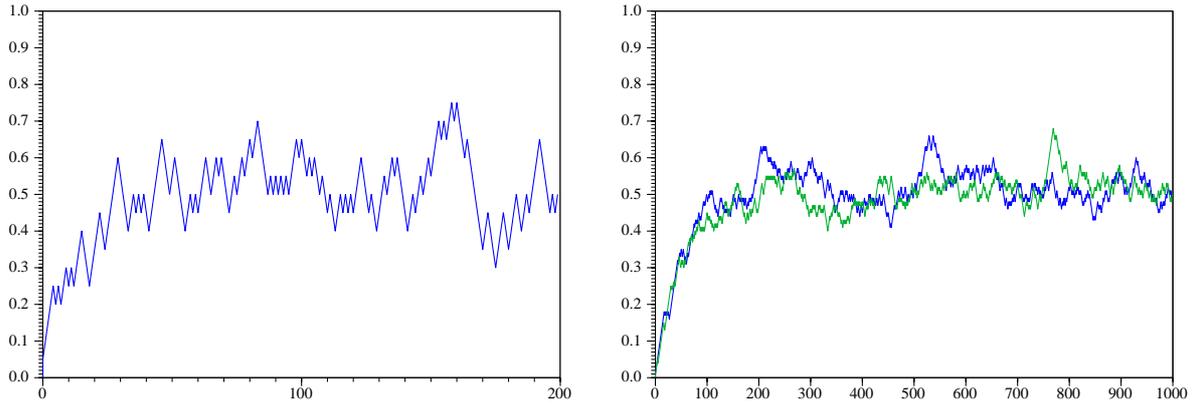


FIGURE 1.3. Réalisations du modèle d'Ehrenfest, montrant la proportion de boules dans l'urne de gauche en fonction du temps. Le nombre initial de boules dans l'urne de gauche est 0. La figure de gauche montre une réalisation pour  $N = 20$  boules, celle de droite montre deux réalisations pour  $N = 100$  boules.

$X_n/N$  approche assez rapidement la valeur  $1/2$ , et tend à fluctuer autour de cette valeur sans trop s'en éloigner. En fait on peut montrer que la loi de  $X_n$  tend vers une loi binomiale  $b(N, 1/2)$ , qui est invariante sous la dynamique. L'espérance de  $X_n/N$  tend donc bien vers  $1/2$  et sa variance tend vers  $1/4N$ . De plus, le temps de récurrence moyen vers un état est égal à l'inverse de la valeur de la loi invariante en cet état. En particulier, le temps de récurrence moyen vers l'état où toutes les boules sont dans la même urne est égal à  $2^N$ . Ceci permet d'expliquer pourquoi dans un système macroscopique, dont le nombre de molécules  $N$  est d'ordre  $10^{23}$ , on n'observe jamais toutes les molécules dans le même récipient – il faudrait attendre un temps d'ordre  $2^{10^{23}} \simeq 10^{3 \cdot 10^{22}}$  pour que cela arrive.

### 1.3 Urnes de Polya

La définition du processus de l'urne de Polya est à priori semblable à celle du modèle d'Ehrenfest, mais elle résulte en un comportement très différent. On considère une urne contenant initialement  $r_0 \geq 1$  boules rouges et  $v_0 \geq 1$  boules vertes. De manière répétée, on tire une boule de l'urne. Si la boule est rouge, on la remet dans l'urne, et on ajoute  $c \geq 1$  boules rouges dans l'urne. Si la boule tirée est verte, on la remet dans l'urne, ainsi que  $c$  boules vertes. Le nombre  $c$  est constant tout au long de l'expérience.

Au temps  $n$ , le nombre total de boules dans l'urne est  $N_n = r_0 + v_0 + nc$ . Soient  $r_n$  et  $v_n$  le nombre de boules rouges et vertes au temps  $n$ , et soit  $X_n = r_n/(r_n + v_n) = r_n/N_n$  la proportion de boules rouges. La probabilité de tirer une boule rouge vaut  $X_n$ , celle de tirer une boule verte vaut  $1 - X_n$ , et on obtient facilement les valeurs correspondantes de  $X_{n+1}$  en fonction de  $X_n$  et  $n$ :

$$X_{n+1} = \begin{cases} \frac{r_n + c}{r_n + v_n + c} = \frac{X_n N_n + c}{N_n + c} & \text{avec probabilité } X_n, \\ \frac{r_n}{r_n + v_n + c} = \frac{X_n N_n}{N_n + c} & \text{avec probabilité } 1 - X_n. \end{cases} \quad (1.3.1)$$

Le processus n'est plus à strictement parler une chaîne de Markov, car l'ensemble des valeurs possibles de  $X_n$  change au cours du temps.

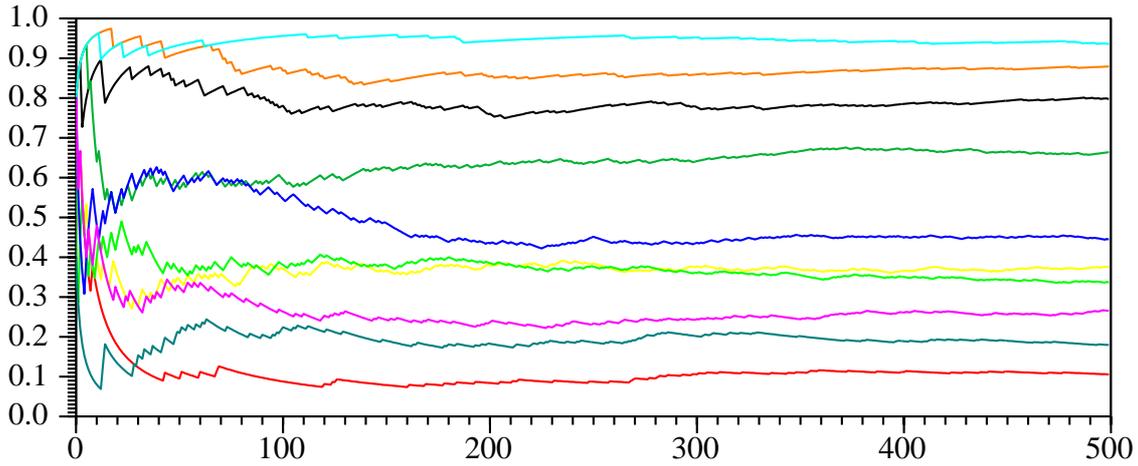


FIGURE 1.4. Dix réalisations du processus d'urne de Polya, avec initialement  $r_0 = 2$  boules rouges et  $v_0 = 1$  boule verte. A chaque pas, on ajoute  $c = 2$  boules.

La Figure 1.4 montre plusieurs réalisations différentes du processus  $\{X_n\}_n$ , pour les mêmes valeurs initiales. Contrairement au modèle d'Ehrenfest,  $X_n$  semble converger vers une constante, qui est différente pour chaque réalisation. Nous montrerons que c'est effectivement le cas, en utilisant le fait que  $X_n$  est une martingale bornée, qui dans ce cas converge presque sûrement. On sait par ailleurs que la loi limite de  $X_n$  est une loi Beta.

## 1.4 Le processus de Galton–Watson

Le processus de Galton–Watson est l'exemple le plus simple de processus stochastique décrivant l'évolution d'une population. Soit  $Z_0$  le nombre d'individus dans la population au temps 0 (souvent on choisit  $Z_0 = 1$ ). On considère alors que chaque individu a un nombre aléatoire de descendants. Les nombres de descendants des différents individus sont indépendants et identiquement distribués.

Cela revient à supposer que

$$Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} \xi_{i,n+1} & \text{si } Z_n \geq 1, \\ 0 & \text{si } Z_n = 0, \end{cases} \quad (1.4.1)$$

où les variables aléatoires  $\{\xi_{i,n}\}_{i,n \geq 1}$ , qui correspondent au nombre de descendants de l'individu  $i$  au temps  $n$ , sont i.i.d. On supposera de plus que chaque  $\xi_{i,n}$  admet une espérance  $\mu$  finie.

Galton et Watson ont introduit leur modèle dans le but d'étudier la disparition de noms de famille (à leur époque, les noms de familles étant transmis de père en fils uniquement, on ne considère que les descendants males). On observe à ce sujet que dès que  $Z_n = 0$  pour un certain  $n = n_0$ , on aura  $Z_n = 0$  pour tous les  $n \geq n_0$ . Cette situation correspond à l'extinction de l'espèce (ou du nom de famille).

Nous montrerons dans ce cours que

- Si  $\mu < 1$ , alors la population s'éteint presque sûrement;
- Si  $\mu > 1$ , alors la population s'éteint avec une probabilité  $\rho = \rho(\mu)$  comprise strictement entre 0 et 1.

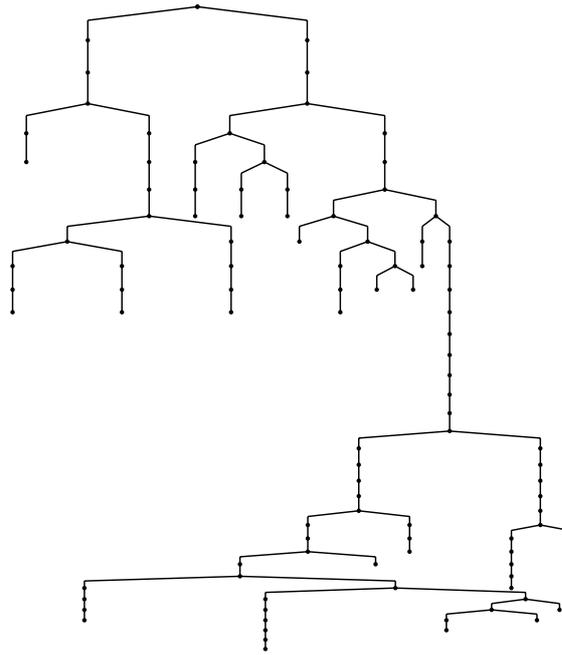


FIGURE 1.5. Une réalisation du processus de Galton–Watson. La distribution des descendants est binomiale, de paramètres  $n = 2$  et  $p = 0.54$ .

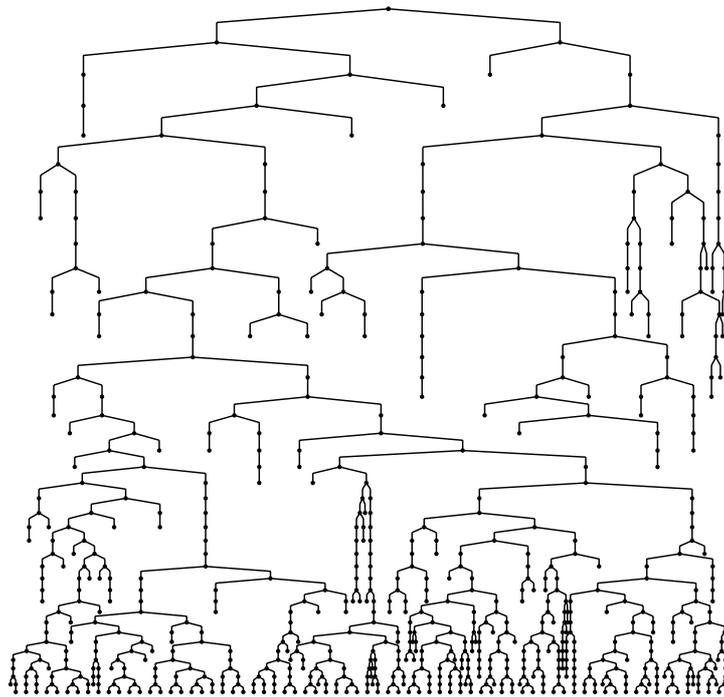


FIGURE 1.6. Une réalisation du processus de Galton–Watson. La distribution des descendants est binomiale, de paramètres  $n = 2$  et  $p = 0.56$ .

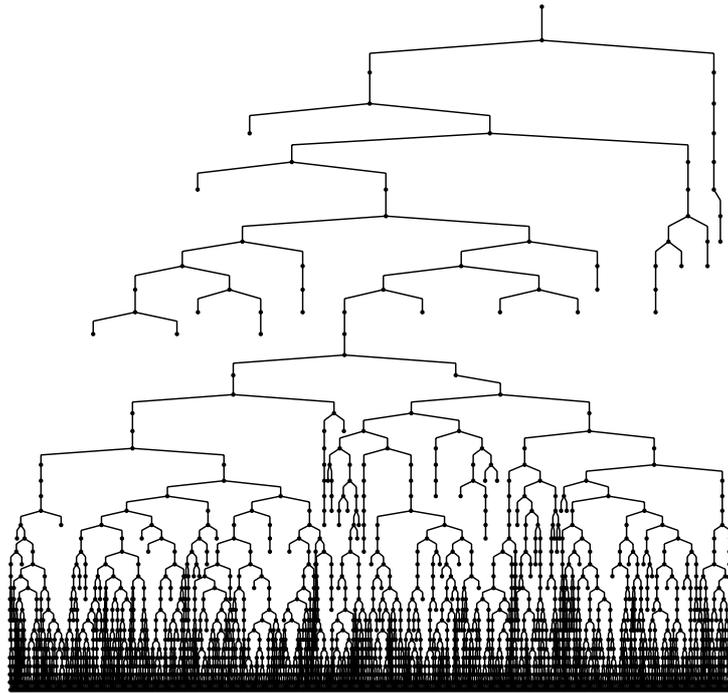


FIGURE 1.7. Une réalisation du processus de Galton–Watson. La distribution des descendants est binomiale, de paramètres  $n = 2$  et  $p = 0.6$ .

La preuve s’appuie sur le fait que  $X_n = Z_n/\mu^n$  est une martingale.

Les Figures 1.5 à 1.7 montrent des réalisations du processus de Galton–Watson sous forme d’arbre. Le temps s’écoule de haut en bas. La distribution des descendants est binomiale  $b(2, p)$  pour des valeurs croissantes de  $p$  (dans ce cas on a  $\mu = 2p$ ). Pour  $p = 0.54$ , on observe l’extinction de l’espèce. Dans les deux autres cas, on a pris des exemples dans lesquels l’espèce survit, et en fait le nombre d’individus croît exponentiellement vite.

Il existe de nombreux modèles d’évolution plus détaillés que le processus de Galton–Watson, tenant compte de la distribution géographique, de la migration, des mutations.

## 1.5 Marches aléatoires auto-évitantes

Les exemples précédents jouissent tous de la propriété de Markov, c’est-à-dire que  $X_{n+1}$  dépend uniquement de  $X_n$ , de l’aléa, et éventuellement du temps  $n$ . Il est facile de modifier ces exemples afin de les rendre non markoviens. Il suffit pour cela de faire dépendre chaque  $X_{n+1}$  de tous ses prédécesseurs.

Un exemple intéressant de processus non markovien est la *marche aléatoire auto-évitante*, qui modélise par exemple certains polymères. Il existe en fait deux variantes de ce processus. Dans la première, on définit  $X_0$  et  $X_1$  comme dans le cas de la marche symétrique simple, mais ensuite chaque  $X_n$  est choisi uniformément parmi tous les plus proches voisins jamais visités auparavant. La seconde variante, un peu plus simple à étudier, est obtenue en considérant comme équiprobables toutes les lignes brisées de longueur donnée ne passant jamais plus d’une fois au même endroit.

Une question importante pour les marches aléatoires auto-évitantes est le comportement asymptotique du déplacement quadratique moyen  $\mathbb{E}(\|X_n\|^2)^{1/2}$ . On sait par exemple

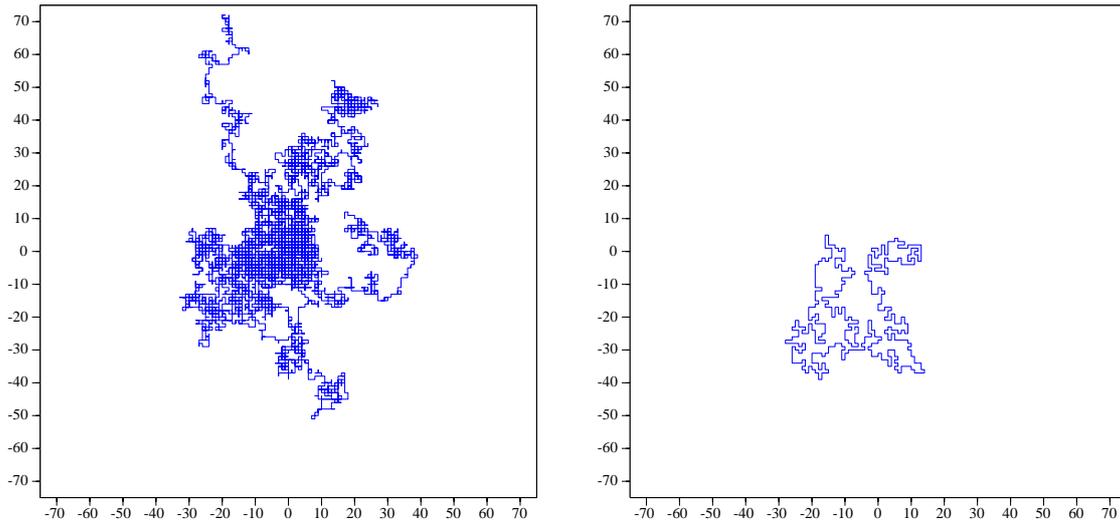


FIGURE 1.8. A gauche, une réalisation d'une marche aléatoire simple dans  $\mathbb{Z}^2$ . A droite, une réalisation d'une marche aléatoire auto-évitante dans  $\mathbb{Z}^2$ , se piégeant après 595 pas.

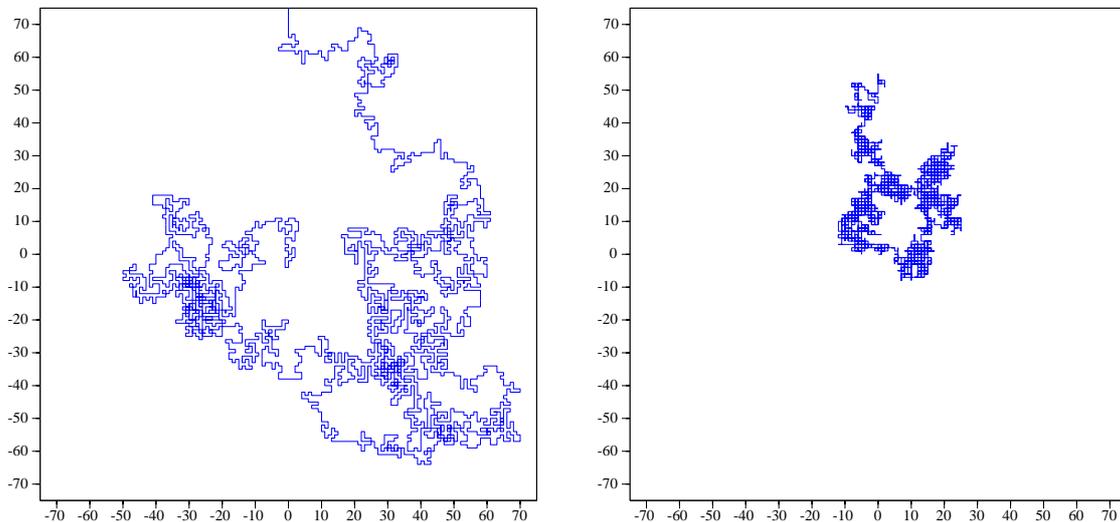


FIGURE 1.9. A gauche, une réalisation d'une marche aléatoire choisissant un site jamais visité avec une probabilité  $10^5$  fois supérieure à un site déjà visité. A droite, la marche choisit un site déjà visité deux fois plus souvent qu'un site jamais visité.

démontrer rigoureusement que pour  $d \geq 5$ , ce déplacement croît comme  $n^{1/2}$ , comme c'est le cas pour les marches aléatoires simples. Pour les dimensions inférieures, la question de la vitesse de croissance est encore ouverte (le déplacement devrait croître plus rapidement que  $n^{1/2}$  à cause des contraintes géométriques). La conjecture (pour la seconde variante) est que la croissance est en  $n^{3/4}$  en dimension 2, en  $n^\nu$  avec  $\nu \simeq 0.59$  en dimension 3, et en  $n^{1/2}(\log n)^{1/8}$  en dimension 4.

Parmi les variantes de ce modèle, mentionnons les marches aléatoires attirées ou repoussées par les sites déjà visités, c'est-à-dire que quand elles passent à côté d'un tel site, la probabilité d'y revenir est soit plus grande, soit plus petite que la probabilité de choisir un site jamais encore visité (Figure 1.9). On peut également faire dépendre les probabilités du nombre de fois qu'un site a déjà été visité dans le passé.

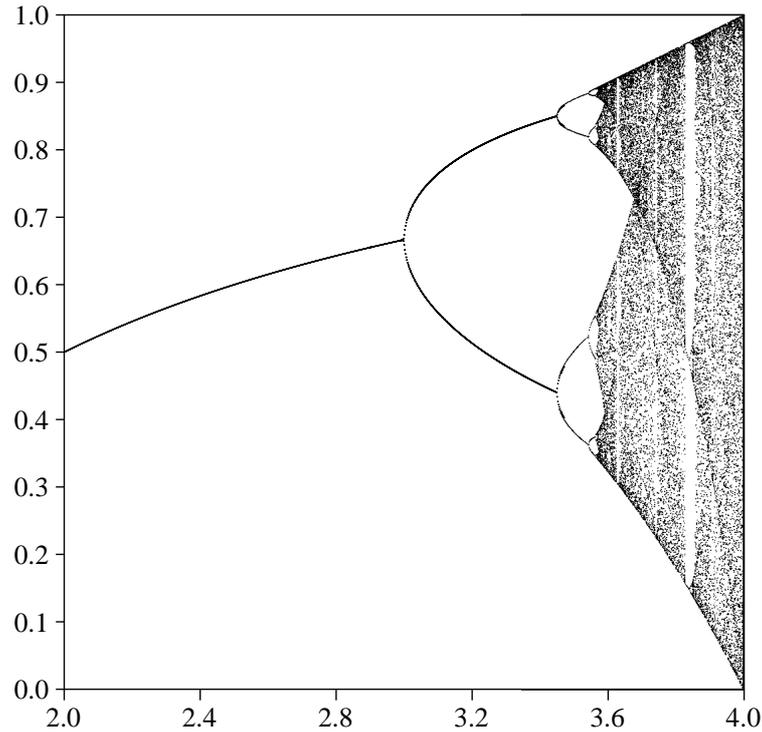


FIGURE 1.10. Diagramme de bifurcation de l'application logistique. Pour chaque valeur du paramètre  $\lambda$  en abscisse, on a représenté les valeurs  $X_{1001}$  à  $X_{1100}$ . La condition initiale est toujours  $X_0 = 1/2$ .

## 1.6 Systèmes dynamiques

Les suites de variables i.i.d. forment les processus les plus aléatoires. À l'autre extrême, on trouve les systèmes dynamiques, définis par

$$X_{n+1} = f(X_n), \quad (1.6.1)$$

où  $f$  est une fonction fixée. Dans ce cas, comme

$$\mathbb{P}\{X_{n+1} \in A\} = \mathbb{P}\{X_n \in f^{-1}(A)\}, \quad (1.6.2)$$

les lois  $\nu_n$  des variables aléatoires  $X_n$  évoluent selon la règle simple

$$\nu_{n+1} = \nu_n \circ f^{-1}. \quad (1.6.3)$$

On pourrait penser que ces processus sont plus simples à analyser, la seule source de hasard étant la distribution de la condition initiale  $X_0$ . Dans certaines situations, par exemple si  $f$  est contractante, c'est effectivement le cas, mais de manière générale il n'en est rien. Par exemple dans le cas de l'application logistique

$$f(x) = \lambda x(1 - x), \quad (1.6.4)$$

le comportement de la suite des  $X_n$  est chaotique pour certaines valeurs de  $\lambda$ , en particulier pour  $\lambda = 4$ . En fait l'existence d'une composante probabiliste tend plutôt à simplifier l'étude de la suite des  $X_n$ . Par exemple, pour  $\lambda = 4$  on connaît la mesure de probabilité invariante du processus, c'est-à-dire la mesure  $\nu$  telle que  $\nu \circ f^{-1} = \nu$ .



## Chapitre 2

# Construction générale de processus stochastiques à temps discret

### 2.1 Préliminaires et rappels

Rappelons quelques notions de base de théorie de la mesure et de probabilités.

Soit  $\Omega$  un ensemble non vide, et soit  $\mathcal{F}$  une *tribu* sur  $\Omega$ , c'est-à-dire une collection de sous-ensembles de  $\Omega$ , contenant  $\Omega$  et stable par réunion dénombrable et complémentaire. On dit que  $(\Omega, \mathcal{F})$  forme un *espace mesurable*.

Une *mesure* sur  $(\Omega, \mathcal{F})$  est une application  $\mu : \mathcal{F} \rightarrow [0, \infty]$  telle que  $\mu(\emptyset) = 0$  et satisfaisant

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \quad (2.1.1)$$

pour toute famille  $\{A_n\}$  d'éléments deux à deux disjoints de  $\mathcal{F}$  ( $\sigma$ -*additivité*). Si  $\mu(\Omega) = 1$  on dit que c'est une *mesure de probabilité*, qu'on notera souvent  $\mathbb{P}$ . Le triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  est appelé un *espace probabilisé*.

Si  $(E, \mathcal{E})$  est un espace mesurable, une application  $f : \Omega \rightarrow E$  est dite  $\mathcal{F}$ - $\mathcal{E}$ -*mesurable* si elle satisfait

$$f^{-1}(A) \in \mathcal{F} \quad \forall A \in \mathcal{E} . \quad (2.1.2)$$

Une *variable aléatoire à valeurs dans*  $(E, \mathcal{E})$  sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  est une application  $\mathcal{F}$ - $\mathcal{E}$ -mesurable  $X : \Omega \rightarrow E$ . Nous aurons souvent affaire au cas  $E = \mathbb{R}$ , avec  $\mathcal{E}$  la tribu des boréliens  $\mathcal{B}$ . Dans ce cas nous dirons simplement que  $X$  est  $\mathcal{F}$ -mesurable, et écrirons  $X \subseteq \mathcal{F}$ . La *loi* d'une variable aléatoire  $X$  à valeurs dans  $(E, \mathcal{E})$  est l'application

$$\begin{aligned} \mathbb{P}X^{-1} : \mathcal{E} &\rightarrow [0, 1] \\ A &\mapsto \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{X \in A\} . \end{aligned} \quad (2.1.3)$$

L'*espérance* d'une variable aléatoire  $X$  est définie comme l'intégrale de Lebesgue

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} . \quad (2.1.4)$$

Rappelons que cette intégrale est définie en approchant  $X$  par une suite de fonctions étagées  $X_n = \sum_i a_i 1_{A_i}$ , pour lesquelles  $\int_{\Omega} X_n \, d\mathbb{P} = \sum_i a_i \mathbb{P}(A_i)$ .

## 2.2 Distributions de dimension finie

Soit  $(E, \mathcal{E})$  un espace mesurable. L'ensemble  $E^n = E \times E \times \cdots \times E$  peut être muni d'une tribu  $\mathcal{E}^{\otimes n}$ , définie comme la tribu engendrée par tous les événements du type

$$A^{(i)} = \{\omega \in E^n : \omega_i \in A\}, \quad A \in \mathcal{E}, \quad (2.2.1)$$

appelés *cylindres*. On dénote par  $E^{\mathbb{N}}$  l'ensemble des applications  $x : \mathbb{N} \rightarrow E$ , c'est-à-dire l'ensemble des suites  $(x_0, x_1, x_2, \dots)$  à valeurs dans  $E$ . Cet ensemble peut à nouveau être muni d'une tribu construite à partir de tous les cylindres, notée  $\mathcal{E}^{\otimes \mathbb{N}}$ .

**Définition 2.2.1** (Processus stochastique, distributions de dimension finie).

- Un processus stochastique à valeurs dans  $(E, \mathcal{E})$  est une suite  $\{X_n\}_{n \in \mathbb{N}}$  de variables aléatoires à valeurs dans  $(E, \mathcal{E})$ , définies sur un même espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  (autrement dit, chaque  $X_n$  est une application  $\mathcal{F}$ - $\mathcal{E}$ -mesurable de  $\Omega$  dans  $E$ ). C'est donc également une variable aléatoire à valeurs dans  $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ .
- Soit  $\mathbb{Q}$  une mesure de probabilité sur  $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ . Les distributions de dimension finie de  $\mathbb{Q}$  sont les mesures sur  $(E^{n+1}, \mathcal{E}^{\otimes n+1})$  définies par

$$\mathbb{Q}^{(n)} = \mathbb{Q} \circ (\pi^{(n)})^{-1}, \quad (2.2.2)$$

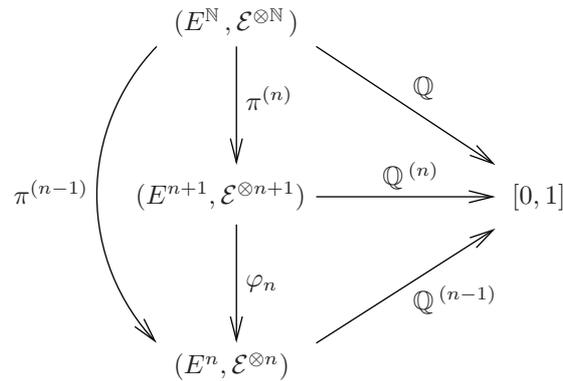
où  $\pi^{(n)}$  est la projection  $\pi^{(n)} : E^{\mathbb{N}} \rightarrow E^{n+1}, (x_0, x_1, x_2, \dots) \mapsto (x_0, \dots, x_n)$ .

On se convainc facilement que la suite des  $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$  détermine  $\mathbb{Q}$  univoquement. Inversément, pour qu'une suite donnée  $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$  corresponde effectivement à une mesure  $\mathbb{Q}$ , les  $\mathbb{Q}^{(n)}$  doivent satisfaire une *condition de compatibilité* :

Soit  $\varphi_n$  la projection  $\varphi_n : E^{n+1} \rightarrow E^n, (x_0, \dots, x_n) \mapsto (x_0, \dots, x_{n-1})$ . Alors on a  $\pi^{(n-1)} = \varphi_n \circ \pi^{(n)}$ , donc pour tout  $A \in \mathcal{E}^{\otimes n}$ ,  $(\pi^{(n-1)})^{-1}(A) = (\pi^{(n)})^{-1}(\varphi_n^{-1}(A))$ . La condition de compatibilité s'écrit donc

$$\mathbb{Q}^{(n-1)} = \mathbb{Q}^{(n)} \circ \varphi_n^{-1}. \quad (2.2.3)$$

Le diagramme suivant illustre la situation (toutes les projections étant mesurables, on peut les considérer à la fois comme applications entre ensembles et entre tribus) :



## 2.3 Noyaux markoviens

Nous allons voir comment construire une suite de  $\mathbb{Q}^{(n)}$  satisfaisant la condition (2.2.3).

**Définition 2.3.1** (Noyau markovien). Soient  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  deux espaces mesurables. Un noyau markovien de  $(E_1, \mathcal{E}_1)$  vers  $(E_2, \mathcal{E}_2)$  est une application  $K : E_1 \times \mathcal{E}_2 \rightarrow [0, 1]$  satisfaisant les deux conditions

1. Pour tout  $x \in E_1$ ,  $K(x, \cdot)$  est une mesure de probabilité sur  $(E_2, \mathcal{E}_2)$ .
2. Pour tout  $A \in \mathcal{E}_2$ ,  $K(\cdot, A)$  est une application  $\mathcal{E}_1$ -mesurable.

**Exemple 2.3.2.**

1. Soit  $\mu$  une mesure de probabilité sur  $(E_2, \mathcal{E}_2)$ . Alors  $K$  défini par  $K(x, A) = \mu(A)$  pour tout  $x \in E_1$  est un noyau markovien.
2. Soit  $f : E_1 \rightarrow E_2$  une application mesurable. Alors  $K$  défini par  $K(x, A) = 1_A(f(x))$  est un noyau markovien.
3. Soit  $\mathcal{X} = \{1, \dots, N\}$  un ensemble fini, et posons  $E_1 = E_2 = \mathcal{X}$  et  $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{P}(\mathcal{X})$ . Alors  $K$  défini par

$$K(i, A) = \sum_{j \in A} p_{ij}, \quad (2.3.1)$$

où  $P = (p_{ij})_{i,j \in \mathcal{X}}$  est une matrice stochastique, est un noyau markovien.

Si  $\mu$  est une mesure de probabilité sur  $(E_1, \mathcal{E}_1)$  et  $K$  est un noyau markovien de  $(E_1, \mathcal{E}_1)$  vers  $(E_2, \mathcal{E}_2)$ , on définit une mesure de probabilité  $\mu \otimes K$  sur  $\mathcal{E}_1 \otimes \mathcal{E}_2$  par

$$(\mu \otimes K)(A) := \int_{E_1} K(x_1, A_{x_1}) \mu(dx_1), \quad (2.3.2)$$

où  $A_{x_1} = \{x_2 \in E_2 : (x_1, x_2) \in A\} \in \mathcal{E}_2$  est la *section* de  $A$  en  $x_1$ . On vérifie que c'est bien une mesure de probabilité. Afin de comprendre sa signification, calculons ses marginales. Soient  $\pi_1$  et  $\pi_2$  les projections définies par  $\pi_i(x_1, x_2) = x_i$ ,  $i = 1, 2$ .

1. Pour tout ensemble mesurable  $A_1 \in \mathcal{E}_1$ , on a

$$\begin{aligned} ((\mu \otimes K) \circ \pi_1^{-1})(A_1) &= (\mu \otimes K)(A_1 \times E_2) \\ &= \int_{E_1} 1_{A_1}(x_1) K(x_1, E_2) \mu(dx_1) \\ &= \int_{A_1} K(x_1, E_2) \mu(dx_1) = \mu(A_1), \end{aligned} \quad (2.3.3)$$

où on a utilisé le fait que la section  $(A_1 \times E_2)_{x_1}$  est donnée par  $E_2$  si  $x_1 \in A_1$ , et  $\emptyset$  sinon. Ceci implique

$$(\mu \otimes K) \circ \pi_1^{-1} = \mu. \quad (2.3.4)$$

La première marginale de  $\mu \otimes K$  est donc simplement  $\mu$ .

2. Pour tout ensemble mesurable  $A_2 \in \mathcal{E}_2$ , on a

$$\begin{aligned} ((\mu \otimes K) \circ \pi_2^{-1})(A_2) &= (\mu \otimes K)(E_1 \times A_2) \\ &= \int_{E_1} K(x_1, A_2) \mu(dx_1). \end{aligned} \quad (2.3.5)$$

La seconde marginale de  $\mu \otimes K$  s'interprète comme suit: c'est la mesure sur  $E_2$  obtenue en partant avec la mesure  $\mu$  sur  $E_1$ , et en "allant de tout  $x \in E_1$  vers  $A_2 \in \mathcal{E}_2$  avec probabilité  $K(x_1, A_2)$ ".

Enfin, par une variante du théorème de Fubini–Tonelli, on vérifie que pour toute fonction  $(\mu \otimes K)$ -intégrable  $f : E_1 \times E_2 \rightarrow \mathbb{R}$ , on a

$$\int_{E_1 \times E_2} f d(\mu \otimes K) = \int_{E_1} \left( \int_{E_2} f(x_1, x_2) K(x_1, dx_2) \right) \mu(dx_1). \quad (2.3.6)$$

Nous pouvons maintenant procéder à la construction de la suite  $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$  de distributions de dimension finie, satisfaisant la condition de compatibilité (2.2.3). Sur l'espace mesurable  $(E, \mathcal{E})$ , on se donne une mesure de probabilité  $\nu$ , appelée *mesure initiale*. On se donne pour tout  $n \in \mathbb{N}$  un noyau markovien  $K_n$  de  $(E^{n+1}, \mathcal{E}^{\otimes n+1})$  vers  $(E, \mathcal{E})$ . On définit alors la suite  $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$  de mesures de probabilité sur  $(E^{n+1}, \mathcal{E}^{\otimes n+1})$  récursivement par

$$\begin{aligned} \mathbb{Q}^{(0)} &= \nu, \\ \mathbb{Q}^{(n)} &= \mathbb{Q}^{(n-1)} \otimes K_{n-1}, \quad n \geq 1. \end{aligned} \quad (2.3.7)$$

Par (2.3.4), on a  $\mathbb{Q}^{(n)} \circ \varphi_n^{-1} = (\mathbb{Q}^{(n-1)} \otimes K_{n-1}) \circ \varphi_n^{-1} = \mathbb{Q}^{(n-1)}$ , donc la condition de compatibilité est bien satisfaite.

L'interprétation de (2.3.7) est simplement que chaque noyau markovien  $K_n$  décrit les probabilités de transition entre les temps  $n$  et  $n+1$ , et permet ainsi de définir une mesure sur les segments de trajectoire plus longs d'une unité. Remarquons enfin qu'on peut également construire pour tout  $m, n$  un noyau  $K_{n,m}$  de  $(E^n, \mathcal{E}^{\otimes n})$  vers  $(E^m, \mathcal{E}^{\otimes m})$  tel que  $\mathbb{Q}^{(n+m)} = \mathbb{Q}^{(n)} \otimes K_{n,m}$ .

On peut noter par ailleurs que si  $\psi_n : E^{n+1} \rightarrow E$  désigne la projection sur la dernière composante  $(x_0, \dots, x_n) \mapsto x_n$ , alors la formule (2.3.5) montre que la loi de  $X_n$ , qui est donnée par la marginale  $\nu_n = \mathbb{Q}^{(n)} \circ \psi_n^{-1}$ , s'exprime comme

$$\mathbb{P}\{X_n \in A\} = \nu_n(A) = \int_{E^n} K_{n-1}(x, A) \mathbb{Q}^{(n-1)}(dx). \quad (2.3.8)$$

La situation est illustrée par le diagramme suivant :

$$\begin{array}{ccc} & (E, \mathcal{E}) & \\ & \uparrow \psi_n & \searrow \nu_n \\ (E^{n+1}, \mathcal{E}^{\otimes n+1}) & \xrightarrow{\mathbb{Q}^{(n)}} & [0, 1] \\ & \downarrow \varphi_n & \nearrow \mathbb{Q}^{(n-1)} \\ & (E^n, \mathcal{E}^{\otimes n}) & \end{array}$$

## 2.4 Le théorème de Ionescu–Tulcea

Nous donnons maintenant, sans démonstration, le résultat général assurant la légitimité de toute la procédure.

**Théorème 2.4.1** (Ionescu–Tulcea). *Pour la suite de mesures  $\{\mathbb{Q}^{(n)}\}_{n \in \mathbb{N}}$  construites selon (2.3.7), il existe une unique mesure de probabilité  $\mathbb{Q}$  sur  $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$  telle que  $\mathbb{Q}^{(n)} = \mathbb{Q} \circ (\pi^{(n)})^{-1}$  pour tout  $n$ , c'est-à-dire que les  $\mathbb{Q}^{(n)}$  sont les distributions de dimension finie de  $\mathbb{Q}$ .*

**Exemple 2.4.2.**

1. **Mesures produit:** On se donne une suite  $\{\mu_n\}_{n \in \mathbb{N}}$  de mesures de probabilité sur l'espace mesurable  $(E, \mathcal{E})$ . Soit, pour tout  $n$ ,  $\mathbb{Q}^{(n)} = \mu_0 \otimes \mu_1 \otimes \cdots \otimes \mu_n$  la mesure produit. C'est une mesure de la forme ci-dessus, avec noyau markovien

$$K_n(x, A) = \mu_{n+1}(A) \quad \forall x \in E^n, \forall A \in \mathcal{E}. \quad (2.4.1)$$

La relation (2.3.8) montre que la loi  $\nu_n$  de  $X_n$  est donnée par  $\mu_n$ . On dit que les variables aléatoires  $X_n$  sont *indépendantes*. Si tous les  $\mu_n$  sont les mêmes, on dit qu'elles sont *indépendantes et identiquement distribuées (i.i.d.)*.

2. **Système dynamique:** On se donne une application mesurable  $f : E \rightarrow E$ , une mesure de probabilité initiale  $\nu$  sur  $E$ . Soit pour tout  $n$  le noyau markovien

$$K_n(x, A) = 1_A f(x_n), \quad (2.4.2)$$

et construisons les  $\mathbb{Q}^{(n)}$  comme ci-dessus. Alors la formule (2.3.8) montre qu'on a pour tout  $A \in E$

$$\begin{aligned} \nu_{n+1}(A) &= \int_{E^{n+1}} 1_A f(x_n) \mathbb{Q}^{(n)}(dx) \\ &= \int_E 1_A f(x_n) \nu_n(dx_n) \\ &= \int_{f^{-1}(A)} \nu_n(dx_n) = \nu_n(f^{-1}(A)). \end{aligned} \quad (2.4.3)$$

Il suit que

$$\nu_n = \nu \circ f^{-n} \quad \forall n \in \mathbb{N}. \quad (2.4.4)$$

Cette situation correspond à un système dynamique déterministe. Par exemple, si  $f$  est bijective et  $\nu = \delta_{x_0}$  est concentrée en un point, on a  $\nu_n = \delta_{f^n(x_0)}$ .

3. **Chaînes de Markov:** Soit  $\mathcal{X}$  un ensemble fini ou dénombrable, muni de la tribu  $\mathcal{P}(\mathcal{X})$ , et  $P = (p_{ij})_{i,j \in \mathcal{X}}$  une matrice stochastique sur  $\mathcal{X}$ , c'est-à-dire que  $0 \leq p_{ij} \leq 1$   $\forall i, j \in \mathcal{X}$  et  $\sum_{j \in \mathcal{X}} p_{ij} = 1$   $\forall i \in \mathcal{X}$ . On se donne une mesure de probabilité  $\nu$  sur  $\mathcal{X}$ , et une suite  $\mathbb{Q}^{(n)}$  construite à partir de (2.3.7) avec pour noyaux markoviens

$$K_n(i_{[0,n-1]}, i_n) = p_{i_{n-1}i_n}. \quad (2.4.5)$$

Le processus stochastique de mesure  $\mathbb{Q}$ , dont les distributions de dimension finie sont les  $\mathbb{Q}^{(n)}$ , est la chaîne de Markov sur  $\mathcal{X}$  de distribution initiale  $\nu$  et de matrice de transition  $P$ . Dans ce cas la relation (2.3.8) se traduit en

$$\mathbb{P}\{X_n \in A\} = \nu_n(A) = \sum_{i \in \mathcal{X}} \sum_{j \in A} p_{ij} \nu_{n-1}(\{i\}) = \sum_{j \in A} \sum_{i \in \mathcal{X}} \mathbb{P}\{X_{n-1} = i\} p_{ij}. \quad (2.4.6)$$



## Chapitre 3

# Filtrations, espérance conditionnelle

### 3.1 Sous-tribus et filtrations

Soit  $(\Omega, \mathcal{F})$  un espace mesurable. Une *sous-tribu* de  $\mathcal{F}$  est une sous-famille  $\mathcal{F}_1 \subset \mathcal{F}$  qui est également une tribu. On dit parfois que la tribu  $\mathcal{F}_1$  est plus grossière que la tribu  $\mathcal{F}$ , et que  $\mathcal{F}$  est plus fine que  $\mathcal{F}_1$ . On notera que si  $X$  est une variable aléatoire réelle, on a l'implication

$$X \subseteq \mathcal{F}_1 \Rightarrow X \subseteq \mathcal{F} . \quad (3.1.1)$$

Une fonction non mesurable peut donc être rendue mesurable en choisissant une tribu plus fine.

#### Exemple 3.1.1.

1. La plus petite sous-tribu de  $\mathcal{F}$  est la *tribu triviale*  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ . On remarquera que si  $X$  est mesurable par rapport à  $\mathcal{F}_0$ , alors la préimage  $X^{-1}(y)$  d'un point  $y$  doit être égal soit à  $\Omega$  tout entier, soit à l'ensemble vide, ce qui implique que  $X$  doit être constante. En d'autres termes, les variables aléatoires mesurables par rapport à la tribu triviale sont les fonctions constantes.
2. Soit  $X$  une variable aléatoire à valeurs dans un espace mesurable  $(E, \mathcal{E})$ . Alors on vérifie aisément que

$$\sigma(X) := \{X^{-1}(A) : A \in \mathcal{E}\} , \quad (3.1.2)$$

où  $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}$ , est une sous-tribu de  $\mathcal{F}$ . C'est la plus petite sous-tribu de  $\mathcal{F}$  par rapport à laquelle  $X$  soit mesurable.

3. Si  $X_1, \dots, X_n$  sont des variables aléatoires à valeurs dans  $(E, \mathcal{E})$ , alors

$$\sigma(X_1, \dots, X_n) := \{(X_1, \dots, X_n)^{-1}(A) : A \in \mathcal{E}^{\otimes n}\} , \quad (3.1.3)$$

où  $(X_1, \dots, X_n)^{-1}(A) := \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in A\}$ , est une sous-tribu de  $\mathcal{F}$ . C'est la plus petite sous-tribu de  $\mathcal{F}$  par rapport à laquelle les variables aléatoires  $X_1, \dots, X_n$  soient toutes mesurables.

L'exemple ci-dessus donne une interprétation importante de la notion de sous-tribu. En effet,  $\sigma(X)$  représente l'ensemble des événements qu'on est potentiellement capable de distinguer en mesurant la variable  $X$ . Autrement dit,  $\sigma(X)$  est l'information que  $X$  peut fournir sur l'espace probabilisé.

**Définition 3.1.2** (Filtration, processus adapté).

1. Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé. Une filtration de  $(\Omega, \mathcal{F}, \mathbb{P})$  est une suite croissante de sous-tribus

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{F}. \quad (3.1.4)$$

On dit alors que  $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$  est un espace probabilisé filtré.

2. Soit  $\{X_n\}_{n \in \mathbb{N}}$  un processus stochastique sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . On dit que le processus est adapté à la filtration  $\{\mathcal{F}_n\}$  si  $X_n$  est mesurable par rapport à  $\mathcal{F}_n$  pour tout  $n$ .

Un choix minimal de filtration adaptée est la *filtration canonique* (ou *naturelle*)

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n). \quad (3.1.5)$$

Dans ce cas,  $\mathcal{F}_n$  représente l'information disponible au temps  $n$ , si l'on observe le processus stochastique.

**Exemple 3.1.3.** Considérons la marche aléatoire symétrique sur  $\mathbb{Z}$ . Dans ce cas,  $E = \mathbb{Z}$  et  $\mathcal{E} = \mathcal{P}(\mathbb{Z})$ . Le choix de  $\Omega$  est arbitraire, il suffit de le prendre "assez grand" pour distinguer toutes les réalisations possibles de la marche. Un choix pratique est l'ensemble  $\Omega = \{-1, 1\}^{\mathbb{N}}$  des suites infinies de  $-1$  et de  $1$ , avec la convention que le  $n$ ième élément de la suite spécifie la direction du  $n$ ième pas de la marche. En d'autres termes,

$$X_n(\omega) = \sum_{i=1}^n \omega_i. \quad (3.1.6)$$

La tribu associée est  $\mathcal{F} = \mathcal{P}(\{-1, 1\})^{\otimes \mathbb{N}} = \{\emptyset, \{-1\}, \{1\}, \{-1, 1\}\}^{\otimes \mathbb{N}}$ .

Construisons maintenant la filtration naturelle. Tout d'abord,  $X_0 = 0$  n'est pas vraiment aléatoire. Nous avons pour tout  $A \subset \mathbb{Z}$

$$X_0^{-1}(A) = \{\omega \in \Omega : X_0 = 0 \in A\} = \begin{cases} \emptyset & \text{si } 0 \notin A, \\ \Omega & \text{si } 0 \in A, \end{cases} \quad (3.1.7)$$

de sorte que

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \quad (3.1.8)$$

est la tribu triviale.

Pour  $n = 1$ , on observe que  $(X_0, X_1)(\Omega) = \{(0, -1), (0, 1)\}$ . Il y a donc quatre cas à distinguer, selon qu'aucun, l'un ou l'autre, ou les deux points appartiennent à  $A \subset \mathbb{Z}^2$ . Ainsi,

$$(X_0, X_1)^{-1}(A) = \begin{cases} \emptyset & \text{si } (0, -1) \notin A \text{ et } (0, 1) \notin A, \\ \{\omega : \omega_1 = 1\} & \text{si } (0, -1) \notin A \text{ et } (0, 1) \in A, \\ \{\omega : \omega_1 = -1\} & \text{si } (0, -1) \in A \text{ et } (0, 1) \notin A, \\ \Omega & \text{si } (0, -1) \in A \text{ et } (0, 1) \in A, \end{cases} \quad (3.1.9)$$

et par conséquent

$$\mathcal{F}_1 = \{\emptyset, \{\omega : \omega_1 = 1\}, \{\omega : \omega_1 = -1\}, \Omega\}. \quad (3.1.10)$$

Ceci traduit bien le fait que  $\mathcal{F}_1$  contient l'information disponible au temps 1 : On sait distinguer tous les événements dépendant du premier pas de la marche. Les variables aléatoires mesurables par rapport à  $\mathcal{F}_1$  sont précisément celles qui ne dépendent que de  $\omega_1$ .

Jusqu'au temps  $n = 2$ , il y a quatre trajectoires possibles, puisque  $(X_0, X_1, X_2)(\Omega) = \{(0, -1, -2), (0, -1, 0), (0, 1, 0), (0, 1, 2)\}$ . Il suit par un raisonnement analogue que  $\mathcal{F}_2$  contient  $2^4 = 16$  éléments, qui se distinguent par quels points parmi ces quatre sont contenus dans  $A$ . Les variables aléatoires mesurables par rapport à  $\mathcal{F}_2$  sont précisément celles qui ne dépendent que de  $\omega_1$  et  $\omega_2$ .

Il est maintenant facile de généraliser à des  $n$  quelconques.

## 3.2 Espérance conditionnelle

Dans cette section, nous fixons un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  et une sous-tribu  $\mathcal{F}_1 \subset \mathcal{F}$ . Nous avons vu que  $\mathcal{F}_1$  représente une information partielle sur l'espace, obtenue par exemple en observant une variable aléatoire  $X_1$ . L'espérance conditionnelle d'une variable aléatoire  $X$  par rapport à  $\mathcal{F}_1$  représente la meilleure estimation que l'on puisse faire de la valeur de  $X$  à l'aide de l'information contenue dans  $\mathcal{F}_1$ .

**Définition 3.2.1** (Espérance conditionnelle). *Soit  $X$  une variable aléatoire réelle sur  $(\Omega, \mathcal{F}, \mathbb{P})$  telle que  $\mathbb{E}(|X|) < \infty$ . On appelle espérance conditionnelle de  $X$  sachant  $\mathcal{F}_1$ , et on note  $\mathbb{E}(X|\mathcal{F}_1)$ , toute variable aléatoire  $Y$  satisfaisant les deux conditions*

1.  $Y \subseteq \mathcal{F}_1$ , c'est-à-dire  $Y$  est  $\mathcal{F}_1$ -mesurable;
2. pour tout  $A \in \mathcal{F}_1$ , on a

$$\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P} . \quad (3.2.1)$$

Si  $Z$  est une variable aléatoire réelle sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , nous abrégons  $\mathbb{E}(X|\sigma(Z))$  par  $\mathbb{E}(X|Z)$ .

En fait, toute variable aléatoire  $Y$  satisfaisant la définition est appelée une *version* de  $\mathbb{E}(X|\mathcal{F}_1)$ . Le résultat suivant tranche la question de l'existence et de l'unicité de l'espérance conditionnelle.

### Théorème 3.2.2.

1. L'espérance conditionnelle  $\mathbb{E}(X|\mathcal{F}_1)$  existe.
2. L'espérance conditionnelle est unique dans le sens que si  $Y$  et  $Y'$  sont deux versions de  $\mathbb{E}(X|\mathcal{F}_1)$ , alors  $Y = Y'$  presque sûrement.
3. On a  $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)) = \mathbb{E}(X)$  et  $\mathbb{E}(|\mathbb{E}(X|\mathcal{F}_1)|) \leq \mathbb{E}(|X|)$ .

**DÉMONSTRATION.** Commençons par prouver les assertions du point 3. La première est un cas particulier de (3.2.1) avec  $A = \Omega$ . Pour montrer la seconde, soit  $Y = \mathbb{E}(X|\mathcal{F}_1)$  et  $A = \{Y > 0\} := \{\omega \in \Omega : Y(\omega) > 0\}$ . On a  $A \in \mathcal{F}_1$  par mesurabilité de  $Y$ . Or par (3.2.1),

$$\begin{aligned} \int_A Y \, d\mathbb{P} &= \int_A X \, d\mathbb{P} \leq \int_A |X| \, d\mathbb{P} , \\ \int_{A^c} -Y \, d\mathbb{P} &= \int_{A^c} -X \, d\mathbb{P} \leq \int_{A^c} |X| \, d\mathbb{P} . \end{aligned} \quad (3.2.2)$$

Le résultat suit en ajoutant ces deux inégalités.

Montrons l'unicité. Si  $Y$  et  $Y'$  sont deux versions de  $\mathbb{E}(X|\mathcal{F}_1)$ , alors pour tout  $A \in \mathcal{F}_1$

$$\int_A Y \, d\mathbb{P} = \int_A Y' \, d\mathbb{P} . \quad (3.2.3)$$

Prenons  $A = \{Y - Y' \geq \varepsilon\}$  pour  $\varepsilon > 0$ . Alors

$$0 = \int_A (Y - Y') \, d\mathbb{P} \geq \varepsilon \mathbb{P}(A) = \varepsilon \mathbb{P}\{Y - Y' \geq \varepsilon\}. \quad (3.2.4)$$

Comme c'est vrai pour tout  $\varepsilon$ , on a  $Y \leq Y'$  presque sûrement. En interchangeant les rôles de  $Y$  et  $Y'$ , on obtient l'inégalité inverse, d'où on déduit l'égalité presque sûre.

Montrons finalement l'existence. Rappelons qu'une mesure  $\nu$  sur  $(\Omega, \mathcal{F}_1)$  est dite *absolument continue par rapport à la mesure  $\mu$*  si  $\mu(A) = 0$  implique  $\nu(A) = 0$  pour tout  $A \in \mathcal{F}_1$ . On écrit alors  $\nu \ll \mu$ . Le théorème de Radon–Nikodym affirme qu'il existe alors une fonction  $f \subseteq \mathcal{F}_1$  telle que

$$\int_A f \, d\mu = \nu(A) \quad \forall A \in \mathcal{F}_1. \quad (3.2.5)$$

La fonction  $f$  est appelée *dérivée de Radon–Nikodym* et notée  $d\nu/d\mu$ .

Supposons d'abord que  $X \geq 0$ . Posons  $\mu = \mathbb{P}$  et définissons  $\nu$  par

$$\nu(A) = \int_A X \, d\mathbb{P} \quad \forall A \in \mathcal{F}_1, \quad (3.2.6)$$

de sorte que  $\nu \ll \mu$ . Nous avons  $d\nu/d\mu \subseteq \mathcal{F}_1$  et pour tout  $A \in \mathcal{F}_1$

$$\int_A X \, d\mathbb{P} = \nu(A) = \int_A \frac{d\nu}{d\mu} \, d\mathbb{P}. \quad (3.2.7)$$

Ceci montre que  $d\nu/d\mu$  satisfait (3.2.1). De plus, prenant  $A = \Omega$  on voit que  $d\nu/d\mu$  est intégrable. Par conséquent  $d\nu/d\mu$  est une version de  $\mathbb{E}(X|\mathcal{F}_1)$ .

Finalement, un  $X$  général peut se décomposer  $X = X^+ - X^-$  avec  $X^+, X^- \geq 0$ . Soit  $Y_1 = \mathbb{E}(X^+|\mathcal{F}_1)$  et  $Y_2 = \mathbb{E}(X^-|\mathcal{F}_1)$ . Alors  $Y_1 - Y_2$  est  $\mathcal{F}_1$ -mesurable et intégrable et on a pour tout  $A \in \mathcal{F}_1$

$$\int_A X \, d\mathbb{P} = \int_A X^+ \, d\mathbb{P} - \int_A X^- \, d\mathbb{P} = \int_A Y_1 \, d\mathbb{P} - \int_A Y_2 \, d\mathbb{P} = \int_A (Y_1 - Y_2) \, d\mathbb{P}. \quad (3.2.8)$$

Ceci montre que  $Y_1 - Y_2$  est une version de  $\mathbb{E}(X|\mathcal{F}_1)$ .  $\square$

Dans la suite, nous allons en général ignorer l'existence de plusieurs versions de l'espérance conditionnelle, puisque des variables aléatoires égales presque partout sont indistinguables en pratique.

### Exemple 3.2.3.

1. Supposons que  $X$  soit  $\mathcal{F}_1$ -mesurable. Alors  $\mathbb{E}(X|\mathcal{F}_1) = X$  vérifie la définition. Cela traduit le fait que  $\mathcal{F}_1$  contient déjà toute l'information sur  $X$ .
2. L'autre extrême est le cas de l'indépendance. Rappelons que  $X$  est indépendante de  $\mathcal{F}_1$  si pour tout  $A \in \mathcal{F}_1$  et tout borélien  $B \subset \mathbb{R}$ ,

$$\mathbb{P}(\{X \in B\} \cap A) = \mathbb{P}\{X \in B\}\mathbb{P}(A). \quad (3.2.9)$$

Dans ce cas nous avons  $\mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}(X)$ , c'est-à-dire qu'en l'absence de toute information, la meilleure estimation que l'on puisse faire de  $X$  est son espérance. En particulier, on notera que toute variable aléatoire est indépendante de la tribu triviale  $\mathcal{F}_0$ , et que par conséquent on aura toujours  $\mathbb{E}(X|\mathcal{F}_0) = \mathbb{E}(X)$ .

Pour vérifier la première condition de la définition, il suffit d'observer que  $\mathbb{E}(X)$ , étant une constante, est mesurable par rapport à la tribu triviale  $\mathcal{F}_0$ , donc aussi par rapport à  $\mathcal{F}_1$ . Pour vérifier la seconde assertion, prenons  $A \in \mathcal{F}_1$ . Alors  $1_A \subseteq \mathcal{F}_1$  et par indépendance

$$\int_A X \, d\mathbb{P} = \mathbb{E}(X1_A) = \mathbb{E}(X)\mathbb{E}(1_A) = \mathbb{E}(X)\mathbb{P}(A) = \int_A \mathbb{E}(X) \, d\mathbb{P}. \quad (3.2.10)$$

3. Soit  $\Omega_1, \Omega_2, \dots$  une partition de  $\Omega$  telle que  $\mathbb{P}(\Omega_i)$  soit strictement positif pour tout  $i$ . Soit  $\mathcal{F}_1 = \sigma(\Omega_1, \Omega_2, \dots)$  la tribu engendrée par les  $\Omega_i$ . Alors

$$\mathbb{E}(X|\mathcal{F}_1)(\omega) = \frac{\mathbb{E}(X1_{\Omega_i})}{\mathbb{P}(\Omega_i)} = \frac{1}{\mathbb{P}(\Omega_i)} \int_{\Omega_i} X \, d\mathbb{P} \quad \forall \omega \in \Omega_i. \quad (3.2.11)$$

Dans ce cas, l'information contenue dans  $\mathcal{F}_1$  spécifie dans quel  $\Omega_i$  on se trouve, et la meilleure estimation de  $X$  est donc sa moyenne sur  $\Omega_i$ .

Pour le vérifier, observons d'abord que comme  $\mathbb{E}(X|\mathcal{F}_1)$  est constante sur chaque  $\Omega_i$ , elle est mesurable par rapport à  $\mathcal{F}_1$ . De plus,

$$\int_{\Omega_i} \mathbb{E}(X|\mathcal{F}_1) \, d\mathbb{P} = \mathbb{E}(X1_{\Omega_i}) = \int_{\Omega_i} X \, d\mathbb{P}. \quad (3.2.12)$$

Comme  $\mathcal{F}_1$  est engendrée par les  $\Omega_i$ , le résultat suit par  $\sigma$ -additivité.

4. Considérons une chaîne de Markov  $\{X_n\}_{n \in \mathbb{N}}$  sur un ensemble  $\mathcal{X}$ , de matrice de transition  $P = (p_{i,j})$ , et soit  $f : \mathcal{X} \rightarrow \mathbb{R}$  une fonction mesurable. Nous prétendons que pour tout  $n \in \mathbb{N}$ ,

$$\mathbb{E}(f(X_{n+1})|X_n) = \sum_{j \in \mathcal{X}} f(j)p_{X_n,j}. \quad (3.2.13)$$

En effet, cette expression étant constante sur tout ensemble où  $X_n$  est constant, elle est mesurable par rapport à  $\sigma(X_n)$ . De plus, en appliquant (3.2.11) avec la partition donnée par  $\Omega_i = \{\omega : X_n = i\}$ , on a pour  $\omega \in \Omega_i$

$$\begin{aligned} \mathbb{E}(f(X_{n+1})|X_n)(\omega) &= \frac{\mathbb{E}(f(X_{n+1})1_{\{X_n=i\}})}{\mathbb{P}\{X_n=i\}} \\ &= \sum_{j \in \mathcal{X}} f(j) \frac{\mathbb{P}\{X_{n+1}=j, X_n=i\}}{\mathbb{P}\{X_n=i\}} \\ &= \sum_{j \in \mathcal{X}} f(j) \mathbb{P}\{X_{n+1}=j|X_n=i\} = \sum_{j \in \mathcal{X}} f(j)p_{i,j}. \end{aligned} \quad (3.2.14)$$

**Proposition 3.2.4.** *L'espérance conditionnelle a les propriétés suivantes :*

1. *Linéarité* :  $\mathbb{E}(aX + Y|\mathcal{F}_1) = a\mathbb{E}(X|\mathcal{F}_1) + \mathbb{E}(Y|\mathcal{F}_1)$ .
2. *Monotonie* : Si  $X \leq Y$  alors  $\mathbb{E}(X|\mathcal{F}_1) \leq \mathbb{E}(Y|\mathcal{F}_1)$ .
3. *Convergence monotone* : Si  $X_n \geq 0$  est une suite croissante telle que  $X_n \nearrow X$  avec  $\mathbb{E}(X) < \infty$  alors  $\mathbb{E}(X_n|\mathcal{F}_1) \nearrow \mathbb{E}(X|\mathcal{F}_1)$ .
4. *Inégalité de Jensen* : Si  $\varphi$  est convexe et  $\mathbb{E}(|X|)$  et  $\mathbb{E}(|\varphi(X)|)$  sont finies alors

$$\varphi(\mathbb{E}(X|\mathcal{F}_1)) \leq \mathbb{E}(\varphi(X)|\mathcal{F}_1). \quad (3.2.15)$$

5. *Contraction dans  $L^p$  pour  $p \geq 1$*  :  $\mathbb{E}(|\mathbb{E}(X|\mathcal{F}_1)|^p) \leq \mathbb{E}(|X|^p)$ .

Nous laissons la preuve en exercice. Le résultat suivant décrit comment les espérances conditionnelles se comportent par rapport à des sous-tribus. Il dit en résumé que c'est toujours la tribu la plus grossière qui l'emporte.

**Proposition 3.2.5.** *Si  $\mathcal{F}_1 \subset \mathcal{F}_2$ , alors*

1.  $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1)$ ;
2.  $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$ .

DÉMONSTRATION. Pour montrer la première identité, il suffit de noter que  $\mathbb{E}(X|\mathcal{F}_1) \subseteq \mathcal{F}_2$  et d'appliquer le premier point de l'exemple 3.2.3. Pour la seconde relation, on observe que pour tout  $A \in \mathcal{F}_1 \subset \mathcal{F}_2$ ,

$$\int_A \mathbb{E}(X|\mathcal{F}_1) \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A \mathbb{E}(X|\mathcal{F}_2) \, d\mathbb{P}, \quad (3.2.16)$$

la première égalité suivant de  $\mathbb{E}(X|\mathcal{F}_1) \subseteq \mathcal{F}_1$  et la seconde de  $\mathbb{E}(X|\mathcal{F}_2) \subseteq \mathcal{F}_2$ .  $\square$

Le résultat suivant montre que les variables aléatoires  $\mathcal{F}_1$ -mesurables se comportent comme des constantes relativement aux espérances conditionnelles par rapport à  $\mathcal{F}_1$ .

**Théorème 3.2.6.** *Si  $X \subseteq \mathcal{F}_1$  et  $\mathbb{E}(|Y|), \mathbb{E}(|XY|) < \infty$ , alors*

$$\mathbb{E}(XY|\mathcal{F}_1) = X\mathbb{E}(Y|\mathcal{F}_1). \quad (3.2.17)$$

DÉMONSTRATION. Le membre de droite étant  $\mathcal{F}_1$ -mesurable, la première condition de la définition est vérifiée. Pour vérifier la seconde condition, nous commençons par considérer le cas  $X = 1_B$  avec  $B \in \mathcal{F}_1$ . Alors pour tout  $A \in \mathcal{F}_1$  on a

$$\int_A 1_B \mathbb{E}(Y|\mathcal{F}_1) \, d\mathbb{P} = \int_{A \cap B} \mathbb{E}(Y|\mathcal{F}_1) \, d\mathbb{P} = \int_{A \cap B} Y \, d\mathbb{P} = \int_A 1_B Y \, d\mathbb{P}. \quad (3.2.18)$$

Ceci montre que la seconde condition est vérifiée pour des fonctions indicatrices. Par linéarité, elle est aussi vraie pour des fonctions étagées  $\sum a_i 1_{B_i}$ . Le théorème de convergence monotone permet d'étendre le résultat aux variables  $X, Y \geq 0$ . Enfin, pour le cas général, il suffit de décomposer  $X$  et  $Y$  en leurs parties positive et négative.  $\square$

Enfin, le résultat suivant montre que l'espérance conditionnelle peut être considérée comme une projection de  $L^2(\mathcal{F}) = \{Y \subseteq \mathcal{F} : \mathbb{E}(Y^2) < \infty\}$  dans  $L^2(\mathcal{F}_1)$ .

**Théorème 3.2.7.** *Si  $\mathbb{E}(X^2) < \infty$ , alors  $\mathbb{E}(X|\mathcal{F}_1)$  est la variable  $Y \subseteq \mathcal{F}_1$  qui minimise  $\mathbb{E}((X - Y)^2)$ .*

DÉMONSTRATION. Pour tout  $Z \in L^2(\mathcal{F}_1)$ , le théorème précédent montre que  $Z\mathbb{E}(X|\mathcal{F}_1) = \mathbb{E}(ZX|\mathcal{F}_1)$ . Prenant l'espérance, on obtient

$$\mathbb{E}(Z\mathbb{E}(X|\mathcal{F}_1)) = \mathbb{E}(\mathbb{E}(ZX|\mathcal{F}_1)) = \mathbb{E}(ZX), \quad (3.2.19)$$

ou encore

$$\mathbb{E}(Z[X - \mathbb{E}(X|\mathcal{F}_1)]) = 0. \quad (3.2.20)$$

Si  $Y = \mathbb{E}(X|\mathcal{F}_1) + Z \subseteq \mathcal{F}_1$ , alors

$$\mathbb{E}((X - Y)^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{F}_1) - Z)^2) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{F}_1))^2) + \mathbb{E}(Z^2), \quad (3.2.21)$$

qui est minimal quand  $Z = 0$ .  $\square$

### 3.3 Exercices

**Exercice 3.1.** Dans une expérience consistant à jeter deux tétraèdres parfaitement symétriques, dont les faces sont numérotées de 1 à 4, on considère les variables aléatoires  $X$ , égale à la somme des points, et  $Y$ , égale à leur différence (en valeur absolue).

1. Spécifier un espace probabilisé permettant de décrire cette expérience.
2. Déterminer la loi conjointe de  $X$  et  $Y$  ainsi que leurs espérances.
3. Calculer  $\mathbb{E}(X|Y)$  et  $\mathbb{E}(Y|X)$ .

**Exercice 3.2.** On jette un dé symétrique, puis on jette une pièce de monnaie autant de fois que le dé indique de points. Soit  $X$  le nombre de Pile obtenus. Déterminer  $\mathbb{E}(X)$ .

**Exercice 3.3.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, et  $\mathcal{F}_1$  une sous-tribu de  $\mathcal{F}$ . Pour tout  $A \in \mathcal{F}$ , on pose  $\mathbb{P}(A|\mathcal{F}_1) = \mathbb{E}(1_A|\mathcal{F}_1)$ . Montrer l'inégalité de Bienaymé–Chebyshev

$$\mathbb{P}(\{|X| \geq a\} | \mathcal{F}_1) \leq \frac{1}{a^2} \mathbb{E}(X^2 | \mathcal{F}_1) .$$

**Exercice 3.4.** Soient  $X, Y$  des variables aléatoires réelles intégrables telles que  $XY$  soit également intégrable. Montrer les implications :

$$X, Y \text{ indépendantes} \Rightarrow \mathbb{E}(Y|X) = \mathbb{E}(Y) \Rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

**Indication :** Commencer par considérer des fonctions indicatrices.

**Exercice 3.5.** Soient  $X$  et  $Y$  des variables aléatoires à valeurs dans  $\{-1, 0, 1\}$ . Exprimer les trois conditions de l'exercice 3.4 à l'aide de la loi conjointe de  $X$  et  $Y$ . Donner des contre-exemples aux implications inverses.

**Indication :** On peut supposer  $\mathbb{E}(Y) = 0$ .

**Exercice 3.6.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, et  $\mathcal{F}_1 \subset \mathcal{F}_2$  des sous-tribus de  $\mathcal{F}$ .

1. Montrer que

$$\mathbb{E}([X - \mathbb{E}(X|\mathcal{F}_2)]^2) + \mathbb{E}([\mathbb{E}(X|\mathcal{F}_2) - \mathbb{E}(X|\mathcal{F}_1)]^2) = \mathbb{E}([X - \mathbb{E}(X|\mathcal{F}_1)]^2)$$

2. On pose  $\text{Var}(X|\mathcal{F}_1) = \mathbb{E}(X^2|\mathcal{F}_1) - \mathbb{E}(X|\mathcal{F}_1)^2$ . Montrer que

$$\text{Var}(X) = \mathbb{E}(\text{Var}(X|\mathcal{F}_1)) + \text{Var}(\mathbb{E}(X|\mathcal{F}_1)) .$$

3. Soit  $Y_1, Y_2, \dots$  une suite de variables aléatoires i.i.d. d'espérance  $\mu$  et de variance  $\sigma^2$ . Soit  $N$  une variable aléatoire à valeurs dans  $\mathbb{N}$ , indépendante de tous les  $Y_i$ . Soit finalement  $X = Y_1 + Y_2 + \dots + Y_N$ . Montrer que

$$\text{Var}(X) = \sigma^2 \mathbb{E}(N) + \mu^2 \text{Var}(N) .$$

4. Déterminer la variance de la variable aléatoire  $X$  de l'exercice 3.2.

**Exercice 3.7.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé, et  $\mathcal{G}$  une sous-tribu de  $\mathcal{F}$ . On considère deux variables aléatoires  $X$  et  $Y$  telles que

$$\mathbb{E}(Y|\mathcal{G}) = X \quad \text{et} \quad \mathbb{E}(X^2) = \mathbb{E}(Y^2)$$

1. Calculer  $\text{Var}(Y - X|\mathcal{G})$ .
2. En déduire  $\text{Var}(Y - X)$ .
3. Que peut-on en déduire sur la relation entre  $X$  et  $Y$ ?

**Exercice 3.8.** Un processus ponctuel de Poisson d'intensité  $\lambda > 0$  est un processus  $\{N_t\}_{t \geq 0}$  tel que

- i.  $N_0 = 0$ ;
- ii. pour  $t > s \geq 0$ ,  $N_t - N_s$  est indépendant de  $N_s$ ;
- iii. pour  $t > s \geq 0$ ,  $N_t - N_s$  suit une loi de Poisson de paramètre  $\lambda(t - s)$ :

$$\mathbb{P}\{N_t - N_s = k\} = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}, \quad k \in \mathbb{N} .$$

On se donne des variables aléatoires i.i.d.  $\{\xi_k\}_{k \in \mathbb{N}}$  à valeurs dans  $\mathbb{N}$  et de carré intégrables. Soit

$$X_t = \sum_{k=1}^{N_t} \xi_k .$$

1. Calculer  $\mathbb{E}(X_t)$ .
2. Calculer  $\text{Var}(X_t)$ .

# Chapitre 4

## Martingales

Le nom *martingale* est synonyme de *jeu équitable*, c'est-à-dire d'un jeu où le gain que l'on peut espérer faire en tout temps ultérieur est égal à la somme gagnée au moment présent. En probabilités, on appelle donc martingale un processus stochastique  $\{X_n\}_n$  tel que l'espérance conditionnelle  $\mathbb{E}(X_m|X_n)$  est égale à  $X_n$  pour tout  $m \geq n$ . Les martingales, ainsi que leurs variantes les sous-martingales et les surmartingales, jouissent de nombreuses propriétés qui les rendent très utiles dans l'étude de processus stochastiques plus généraux. Nous allons voir quelques-unes de ces propriétés dans ce chapitre et les suivants.

### 4.1 Définitions et exemples

**Définition 4.1.1** (Martingale, sous- et surmartingale). Soit  $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, \mathbb{P})$  un espace probabilisé filtré. Une martingale par rapport à la filtration  $\{\mathcal{F}_n\}_n$  est un processus stochastique  $\{X_n\}_{n \in \mathbb{N}}$  tel que

1.  $\mathbb{E}(|X_n|) < \infty$  pour tout  $n \in \mathbb{N}$ ;
2.  $\{X_n\}_n$  est adapté à la filtration  $\{\mathcal{F}_n\}_n$ ;
3.  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$  pour tout  $n \in \mathbb{N}$ .

Si la dernière condition est remplacée par  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$  on dit que  $\{X_n\}_n$  est une surmartingale, et si elle est remplacée par  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$  on dit que c'est une sous-martingale.

On notera qu'en termes de jeu, une surmartingale est défavorable au joueur, alors qu'une sous-martingale lui est favorable (la terminologie vient de la notion de fonction sous-harmonique).

#### Exemple 4.1.2.

1. Soit  $\{X_n\}_n$  la marche aléatoire symétrique sur  $\mathbb{Z}$ , et soit  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$  la filtration canonique. On remarque que  $X_{n+1} - X_n$  est indépendant de  $\mathcal{F}_n$ , ce qui permet d'écrire

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}(X_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = X_n + \mathbb{E}(X_{n+1} - X_n) = X_n, \quad (4.1.1)$$

montrant que la marche aléatoire symétrique est une martingale. Cela traduit le fait que si l'on sait où se trouve la marche au temps  $n$ , alors la meilleure estimation de sa position aux temps ultérieurs est donnée par celle au temps  $n$ .

2. Soit  $\mathcal{X} \subset \mathbb{R}$  un ensemble dénombrable, et soit  $\{X_n\}_n$  une chaîne de Markov sur  $\mathcal{X}$ , de matrice de transition  $P = (p_{i,j})_{i,j \in \mathcal{X}}$ . Alors l'équation (3.2.13) implique

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \sum_{j \in \mathcal{X}} j p_{X_n, j}. \quad (4.1.2)$$

Par conséquent la chaîne est une martingale si  $\sum_j j p_{i,j} = i$  pour tout  $i \in \mathcal{X}$ , une surmartingale si  $\sum_j j p_{i,j} \leq i$  pour tout  $i \in \mathcal{X}$ , et une sous-martingale si  $\sum_j j p_{i,j} \geq i$  pour tout  $i \in \mathcal{X}$ . Comme  $\sum_j p_{i,j} = 1$ , la condition d'être une martingale peut aussi s'écrire  $\sum_j (j - i) p_{i,j} = 0$  (respectivement  $\leq 0$  pour une surmartingale et  $\geq 0$  pour une sous-martingale).

3. **Urne de Polya :** On considère une urne contenant  $r$  boules rouges et  $v$  boules vertes. De manière répétée, on tire une boule de l'urne, puis on la remet en ajoutant un nombre fixé  $c$  de boules de la même couleur. Soit  $r_n$  le nombre de boules rouges après le  $n$ ème tirage,  $v_n$  le nombre de boules vertes, et  $X_n = r_n/(r_n + v_n)$  la proportion de boules rouges. On aura donc

$$X_{n+1} = \begin{cases} \frac{r_n + c}{r_n + v_n + c} & \text{avec probabilité } \frac{r_n}{r_n + v_n} = X_n, \\ \frac{r_n}{r_n + v_n + c} & \text{avec probabilité } \frac{v_n}{r_n + v_n} = 1 - X_n, \end{cases} \quad (4.1.3)$$

de sorte que

$$\mathbb{E}(X_{n+1}|X_n) = \frac{r_n + c}{r_n + v_n + c} \frac{r_n}{r_n + v_n} + \frac{r_n}{r_n + v_n + c} \frac{v_n}{r_n + v_n} = \frac{r_n}{r_n + v_n} = X_n. \quad (4.1.4)$$

La suite  $\{X_n\}_n$  est donc une martingale (par rapport à la filtration canonique).

## 4.2 Inégalités

Commençons par vérifier que le fait de ne faire intervenir que des temps consécutifs  $n$  et  $n + 1$  dans la définition n'est pas une restriction, mais que la monotonie s'étend à tous les temps.

**Proposition 4.2.1.** *Si  $\{X_n\}_n$  est une surmartingale (respectivement une sous-martingale, une martingale), alors*

$$\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m \quad \forall n > m \geq 0 \quad (4.2.1)$$

(respectivement  $\mathbb{E}(X_n|\mathcal{F}_m) \geq X_m$ ,  $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$ ).

DÉMONSTRATION. Considérons le premier cas. Le résultat suit de la définition si  $n = m + 1$ . Si  $n = m + k$  avec  $k \geq 2$ , alors par la Proposition 3.2.5,

$$\mathbb{E}(X_{m+k}|\mathcal{F}_m) = \mathbb{E}(\mathbb{E}(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \leq \mathbb{E}(X_{m+k-1}|\mathcal{F}_m). \quad (4.2.2)$$

Le résultat suit alors par récurrence. Si  $\{X_n\}_n$  est une sous-martingale, il suffit d'observer que  $\{-X_n\}_n$  est une surmartingale et d'appliquer la linéarité. Enfin, si  $\{X_n\}_n$  est une martingale, alors c'est à la fois une surmartingale et une sous-martingale, d'où la conclusion (ce raisonnement en trois pas est typique pour les martingales).  $\square$

Un deuxième type important d'inégalité s'obtient en appliquant une fonction convexe à un processus. Nous énoncerons simplement les résultats dans l'un des trois cas sur-sous- ou martingale, mais souvent d'autres inégalités s'obtiennent dans les autres cas par linéarité.

**Proposition 4.2.2.**

1. Soit  $X_n$  une martingale par rapport à  $\mathcal{F}_n$ , et soit  $\varphi$  une fonction convexe telle que  $\mathbb{E}(|\varphi(X_n)|) < \infty$  pour tout  $n$ . Alors  $\varphi(X_n)$  est une sous-martingale par rapport à  $\mathcal{F}_n$ .
2. Soit  $X_n$  une sous-martingale par rapport à  $\mathcal{F}_n$  et  $\varphi$  une fonction convexe croissante telle que  $\mathbb{E}(|\varphi(X_n)|) < \infty$  pour tout  $n$ . Alors  $\varphi(X_n)$  est une sous-martingale par rapport à  $\mathcal{F}_n$ .

DÉMONSTRATION.

1. Par l'inégalité de Jensen,  $\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n)$ .
2. Par l'inégalité de Jensen,  $\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) \geq \varphi(X_n)$ . □

Plusieurs cas particuliers de fonctions  $\varphi$  joueront un rôle dans la suite. Pour deux nombres réels  $a$  et  $b$ , nous notons  $a \wedge b$  leur minimum et  $a \vee b$  leur maximum. De plus,  $a^+ = a \vee 0$  dénote la partie positive de  $a$  et  $a^- = (-a) \vee 0$  sa partie négative.

**Corollaire 4.2.3.**

1. Si  $p \geq 1$  et  $X_n$  est une martingale telle que  $\mathbb{E}(|X_n|^p) < \infty$  pour tout  $n$ , alors  $|X_n|^p$  est une sous-martingale.
2. Si  $X_n$  est une sous-martingale, alors  $(X_n - a)^+$  est une sous-martingale.
3. Si  $X_n$  est une surmartingale, alors  $X_n \wedge a$  est une surmartingale.

### 4.3 Décomposition de Doob, processus croissant

**Définition 4.3.1** (Processus prévisible). Soit  $\{\mathcal{F}_n\}_{n \geq 0}$  une filtration. Une suite  $\{H_n\}_{n \geq 0}$  est un processus prévisible si  $H_n \subseteq \mathcal{F}_{n-1}$  pour tout  $n \geq 1$ .

La notion de prévisibilité se comprend facilement dans le cadre de la théorie des jeux (et par conséquent également dans celle des marchés financiers, ce qui est essentiellement pareil). Supposons en effet qu'un joueur mise de manière répétée sur le résultat d'une expérience aléatoire, telle que le jet d'une pièce de monnaie. Une stratégie est une manière de décider la somme mise à chaque tour, en fonction des gains précédents. Par exemple, le joueur peut décider de doubler la mise à chaque tour, ou de miser une proportion fixée de la somme qu'il a gagnée. Toute stratégie doit être prévisible, car elle ne peut pas dépendre de résultats futurs du jeu. De même, un investisseur décide de la manière de placer ses capitaux en fonction de l'information disponible au temps présent, à moins de commettre une délit d'initié.

Considérons le cas particulier où le joueur gagne un euro pour chaque euro misé si la pièce tombe sur Pile, et perd chaque euro misé si elle tombe sur Face. Soit  $X_n$  la somme totale qu'aurait gagnée au temps  $n$  un joueur misant un Euro à chaque coup. Un joueur suivant la stratégie  $H$  aura alors gagné au temps  $n$  la somme

$$(H \cdot X)_n := \sum_{m=1}^n H_m (X_m - X_{m-1}), \quad (4.3.1)$$

puisque  $X_m - X_{m-1}$  vaut 1 ou  $-1$  selon que la pièce est tombée sur Pile au Face lors du  $m$ ème jet. Le résultat suivant affirme qu'il n'existe pas de stratégie gagnante dans ce jeu.

**Proposition 4.3.2.** *Soit  $\{X_n\}_{n \geq 0}$  une surmartingale et  $H_n$  un processus prévisible, non-négatif et borné pour tout  $n$ . Alors  $(H \cdot X)_n$  est une surmartingale.*

DÉMONSTRATION. Comme  $H_{n+1} \subseteq \mathcal{F}_n$  et  $(H \cdot X)_n \subseteq \mathcal{F}_n$ , la linéarité de l'espérance conditionnelle implique

$$\begin{aligned} \mathbb{E}((H \cdot X)_{n+1} | \mathcal{F}_n) &= (H \cdot X)_n + \mathbb{E}(H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n) \\ &= (H \cdot X)_n + H_{n+1} \mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) \leq (H \cdot X)_n, \end{aligned} \quad (4.3.2)$$

puisque  $H_{n+1} \geq 0$  et  $\mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) = \mathbb{E}(X_{n+1} | \mathcal{F}_n) - X_n \leq 0$ .  $\square$

La décomposition de Doob permet d'écrire une sous-martingale comme la somme d'une martingale et d'un processus prévisible.

**Proposition 4.3.3** (Décomposition de Doob). *Toute sous-martingale  $\{X_n\}_{n \geq 0}$  peut être écrite d'une manière unique comme  $X_n = M_n + A_n$ , où  $M_n$  est une martingale et  $A_n$  est un processus prévisible croissant tel que  $A_0 = 0$ .*

DÉMONSTRATION. La preuve est parfaitement constructive. Supposons d'abord que la décomposition existe, avec  $\mathbb{E}(M_n | \mathcal{F}_{n-1}) = M_{n-1}$  et  $A_n \subseteq \mathcal{F}_{n-1}$ . Alors nécessairement

$$\begin{aligned} \mathbb{E}(X_n | \mathcal{F}_{n-1}) &= \mathbb{E}(M_n | \mathcal{F}_{n-1}) + \mathbb{E}(A_n | \mathcal{F}_{n-1}) \\ &= M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n. \end{aligned} \quad (4.3.3)$$

Il en résulte les deux relations

$$A_n = A_{n-1} + \mathbb{E}(X_n | \mathcal{F}_{n-1}) - X_{n-1} \quad (4.3.4)$$

et

$$M_n = X_n - A_n. \quad (4.3.5)$$

Ces deux relations définissent  $M_n$  et  $A_n$  univoquement. En effet,  $A_0 = 0$  et donc  $M_0 = X_0$  par hypothèse, de sorte que tous les  $M_n$  et  $A_n$  sont définis par récurrence. Le fait que  $X_n$  est une sous-martingale implique que  $A_n \geq A_{n-1} \geq 0$ . Par récurrence,  $A_n \subseteq \mathcal{F}_{n-1}$ . Enfin,

$$\begin{aligned} \mathbb{E}(M_n | \mathcal{F}_{n-1}) &= \mathbb{E}(X_n - A_n | \mathcal{F}_{n-1}) \\ &= \mathbb{E}(X_n | \mathcal{F}_{n-1}) - A_n = X_{n-1} - A_{n-1} = M_{n-1}, \end{aligned} \quad (4.3.6)$$

ce qui montre que  $M_n$  est bien une martingale.  $\square$

Un cas particulier important se présente si  $X_n$  est une martingale telle que  $X_0 = 0$  et  $\mathbb{E}(X_n^2) < \infty$  pour tout  $n$ . Dans ce cas, le Corollaire 4.2.3 implique que  $X_n^2$  est une sous-martingale. La décomposition de Doob de  $X_n^2$  s'écrit alors

$$X_n^2 = M_n + A_n, \quad (4.3.7)$$

où  $M_n$  est une martingale, et (4.3.4) implique

$$A_n = \sum_{m=1}^n \mathbb{E}(X_m^2 | \mathcal{F}_{m-1}) - X_{m-1}^2 = \sum_{m=1}^n \mathbb{E}((X_m - X_{m-1})^2 | \mathcal{F}_{m-1}). \quad (4.3.8)$$

**Définition 4.3.4** (Processus croissant). *Le processus prévisible (4.3.8) est appelé le processus croissant ou le crochet, ou encore le compensateur de  $X_n$ , et noté  $\langle X \rangle_n$ .*

Le processus croissant  $\langle X \rangle_n$  est une mesure de la variance de la trajectoire jusqu'au temps  $n$ . Nous verrons que sous certaines conditions,  $\langle X \rangle_n$  converge vers une variable aléatoire  $\langle X \rangle_\infty$ , qui mesure alors la variance totale des trajectoires.

## 4.4 Exercices

**Exercice 4.1.** Soient  $Y_1, Y_2, \dots$  des variables i.i.d., et soit  $X_n = \prod_{m=1}^n Y_m$ . Sous quelle condition la suite  $X_n$  est-elle une surmartingale? Une sous-martingale? Une martingale?

**Exercice 4.2.** Soit  $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, \mathbb{P})$  un espace probabilisé filtré, et  $X_n = \sum_{m=1}^n 1_{B_m}$ , avec  $B_n \in \mathcal{F}_n \forall n$ .

1. Montrer que  $X_n$  est une sous-martingale.
2. Donner la décomposition de Doob de  $X_n$ .
3. Particulariser au cas  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ .

**Exercice 4.3.** On considère le modèle d'urne de Polya, avec paramètres  $(r, v, c)$ .

1. Déterminer la loi de la proportion de boules vertes  $X_n$  dans le cas  $r = v = c = 1$ .
2. Dans le cas général, exprimer le processus croissant  $\langle X \rangle_n$  en fonction des  $X_m$  et du nombre total de boules  $N_m$  aux temps  $m \leq n$ .
3. Montrer que  $\lim_{n \rightarrow \infty} \langle X \rangle_n < \infty$ .

**Exercice 4.4** (Le processus de Galton–Watson). On se donne des variables aléatoires i.i.d.  $\{\xi_{n,i}\}_{n,i \geq 1}$  à valeurs dans  $\mathbb{N}$ . On note leur distribution  $p_k = \mathbb{P}\{\xi_{n,i} = k\}$ , leur espérance  $\mu > 0$  et leur variance  $\sigma^2$ . On notera  $\mathcal{F}_n$  la filtration  $\mathcal{F}_n = \sigma\{\xi_{i,m}, m \leq n\}$ . On définit un processus  $\{Z_n\}_{n \geq 0}$  par  $Z_0 = 1$  et pour  $n \geq 0$

$$Z_{n+1} = \begin{cases} \xi_{1,n+1} + \dots + \xi_{Z_n,n+1} & \text{si } Z_n > 0, \\ 0 & \text{sinon.} \end{cases}$$

Ce processus modélise l'évolution d'une population avec initialement  $Z_0 = 1$  individu, et dans laquelle chaque individu  $i$  donne naissance au temps  $n$  à un nombre aléatoire  $\xi_{n,i}$  d'enfants, indépendamment et avec la même loi que tous les autres individus.

1. Montrer que  $X_n = Z_n/\mu^n$  est une martingale par rapport à  $\mathcal{F}_n$ .
2. Montrer que si  $\mu < 1$ , alors  $\mathbb{P}\{Z_n = 0\} \rightarrow 1$  lorsque  $n \rightarrow \infty$ , et donc  $X_n \rightarrow 0$ .
3. On suppose  $\mu > 1$ . Soit  $\varphi(s) = \mathbb{E}(s^{\xi_{i,n}}) = \sum_{k=0}^{\infty} p_k s^k$  la fonction génératrice de la distribution d'enfants.
  - (a) Montrer que  $\varphi$  est croissante et convexe sur  $[0, 1]$ .
  - (b) Soit  $\theta_m = \mathbb{P}\{Z_m = 0\}$ . Montrer que  $\mathbb{P}\{Z_m = 0 | Z_1 = k\} = \theta_{m-1}^k$  et en déduire que  $\theta_m = \varphi(\theta_{m-1})$ .
  - (c) Montrer que  $\varphi$  admet un unique point fixe  $\rho$  sur  $[0, 1]$ .
  - (d) Montrer que  $\theta_m \nearrow \rho$  lorsque  $m \rightarrow \infty$ .  
En déduire que  $\mathbb{P}\{Z_n > 0 \forall n\} = 1 - \rho > 0$ .
4. Galton et Watson on introduit leur modèle afin de décrire la survie de noms de famille. Au XVIIIe siècle, ces noms n'étaient transmis que par les enfants de sexe masculin. On suppose que chaque famille a trois enfants, dont le sexe est déterminé par une loi de Bernoulli de paramètre  $1/2$ . Le nombre de descendants mâles est donc décrit par un processus de Galton–Watson de loi binomiale  $p_0 = p_3 = 1/8$ ,  $p_1 = p_2 = 3/8$ . Déterminer la probabilité de survie du nom de famille.
5. Déterminer le processus croissant  $\langle X \rangle_n$  de  $X_n$ . Calculer  $\lim_{n \rightarrow \infty} \langle X \rangle_n$ .



# Chapitre 5

## Temps d'arrêt

Dans un jeu de hasard, un temps d'arrêt est un temps lors duquel le joueur décide d'arrêter de jouer, selon un critère ne dépendant que du passé et du présent. Il peut par exemple décider d'arrêter de jouer dès qu'il a dépensé tout son capital, dès qu'il a gagné une certaine somme, dès qu'il a gagné un certain nombre de fois successives, ou selon toute combinaison de ces critères. Les temps d'arrêt ont donc deux propriétés importantes : ils sont aléatoires, puisqu'ils dépendent du déroulement antérieur du jeu, et ils ne peuvent pas dépendre du futur, puisque le joueur doit à tout moment pouvoir décider s'il arrête ou non.

### 5.1 Définition et exemples

**Définition 5.1.1** (Temps d'arrêt). Soit  $\{X_n\}_{n \in \mathbb{N}}$  un processus stochastique et soit  $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$  la filtration canonique. Une variable aléatoire  $N$  à valeurs dans  $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$  est un temps d'arrêt si  $\{N = n\} \in \mathcal{F}_n$  pour tout  $n \in \mathbb{N}$ .

La condition  $\{N = n\} \in \mathcal{F}_n$  signifie qu'avec l'information disponible au temps  $n$ , on doit pouvoir décider si oui ou non l'événement  $\{N = n\}$  est réalisé. En d'autres termes, un temps d'arrêt ne "peut pas voir dans le futur".

#### Exemple 5.1.2.

1. Si  $N$  est constante presque sûrement, alors c'est un temps d'arrêt.
2. Supposons que  $X_n$  soit réel, et soit  $B \subset \mathbb{R}$  un borélien. Alors le *temps de première atteinte* de  $B$

$$N_B = \inf\{n \in \mathbb{N} : X_n \in B\} \quad (5.1.1)$$

est un temps d'arrêt (par convention, on pose  $N_B = \infty$  si le processus n'atteint jamais l'ensemble  $B$ ). Pour le vérifier, il suffit d'observer qu'on a la décomposition

$$\{N_B = n\} = \bigcap_{k=0}^{n-1} \{X_k \in B^c\} \cap \{X_n \in B\} \in \mathcal{F}_n. \quad (5.1.2)$$

3. Le temps de dernier passage dans  $B$  avant un temps fixé  $m$ ,

$$\sup\{n \leq m : X_n \in B\} \quad (5.1.3)$$

n'est pas un temps d'arrêt. En effet, si par exemple  $X_n \in B$  pour un  $n < m$ , l'information disponible au temps  $n$  ne permet pas de dire si oui ou non le processus

repassera dans l'ensemble  $B$  jusqu'au temps  $m$ . Un joueur ne peut pas décider qu'il jouera jusqu'au dernier tour avant le centième lors duquel il possédera au moins 100 Euros!

**Proposition 5.1.3.**

1. Soient  $N$  et  $M$  des temps d'arrêt. Alors  $N \wedge M$  et  $N \vee M$  sont des temps d'arrêt.
2. Soit  $N_k$ ,  $k \in \mathbb{N}$ , une suite de temps d'arrêt telle que  $N_k \nearrow N$ . Alors  $N$  est un temps d'arrêt.

Nous laissons la preuve en exercice.

**Définition 5.1.4** (Tribu des événements antérieurs). Soit  $N$  un temps d'arrêt. Alors la tribu

$$\mathcal{F}_N = \{A \in \mathcal{F} : A \cap \{N = n\} \in \mathcal{F}_n \forall n < \infty\} \quad (5.1.4)$$

est appelée la tribu des événements antérieurs à  $N$ .

**Proposition 5.1.5.** Soient  $N$  et  $M$  deux temps d'arrêt tels que  $M \leq N$ . Alors  $\mathcal{F}_M \subset \mathcal{F}_N$ .

DÉMONSTRATION. Soit  $A \in \mathcal{F}_M$ . Alors pour tout  $n \geq 0$ ,  $N = n$  implique  $M \leq n$  d'où

$$A \cap \{N = n\} = A \cap \{M \leq n\} \cap \{N = n\} \in \mathcal{F}_n, \quad (5.1.5)$$

puisque  $A \cap \{M \leq n\} = \bigcup_{k=0}^n (A \cap \{M = k\}) \in \mathcal{F}_n$  et  $\{N = n\} \in \mathcal{F}_n$ .  $\square$

## 5.2 Processus arrêté, inégalité de Doob

**Définition 5.2.1** (Processus arrêté). Soit  $\{\mathcal{F}_n\}_n$  une filtration,  $\{X_n\}_n$  un processus adapté à la filtration et  $N$  un temps d'arrêt. On appelle processus arrêté en  $N$  le processus  $X^N$  défini par

$$X_n^N = X_{N \wedge n}. \quad (5.2.1)$$

Si par exemple  $N$  est le temps de première atteinte d'un ensemble  $B$ , alors  $X^N$  est le processus obtenu en "gelant"  $X$  à l'endroit où il atteint  $B$  pour la première fois.

**Proposition 5.2.2.** Si  $N$  est un temps d'arrêt et  $X_n$  est une surmartingale (par rapport à la même filtration  $\mathcal{F}_n$ ), alors le processus arrêté  $X_{N \wedge n}$  est une surmartingale.

DÉMONSTRATION. Considérons le processus  $H_n = 1_{\{N \geq n\}}$ . Puisque l'on a  $\{N \geq n\} = \{N \leq n-1\}^c \in \mathcal{F}_{n-1}$ , le processus  $H_n$  est prévisible. De plus, il est évidemment borné et non-négatif. Par la proposition 4.3.2,  $(H \cdot X)_n$  est une surmartingale. Or nous avons

$$(H \cdot X)_n = \sum_{m=1}^n 1_{\{N \geq m\}} (X_m - X_{m-1}) = \sum_{m=1}^{N \wedge n} (X_m - X_{m-1}) = X_{N \wedge n} - X_0, \quad (5.2.2)$$

donc  $X_{N \wedge n} = X_0 + (H \cdot X)_n$  est une surmartingale.  $\square$

Il suit directement de ce résultat que si  $X_n$  est une sous-martingale (respectivement une martingale), alors  $X_{N \wedge n}$  est une sous-martingale (respectivement une martingale).

**Corollaire 5.2.3.** Soit  $X_n$  une sous-martingale et soit  $N$  un temps d'arrêt satisfaisant  $\mathbb{P}\{N \leq k\} = 1$  pour un  $k \in \mathbb{N}$ . Alors

$$\mathbb{E}(X_0) \leq \mathbb{E}(X_N) \leq \mathbb{E}(X_k). \quad (5.2.3)$$

DÉMONSTRATION. Comme  $X_{N \wedge n}$  est une sous-martingale, on a

$$\mathbb{E}(X_0) = \mathbb{E}(X_{N \wedge 0}) \leq \mathbb{E}(X_{N \wedge k}) = \mathbb{E}(X_N). \quad (5.2.4)$$

Pour prouver la seconde inégalité, nous imitons la preuve de la proposition en utilisant  $K_n = 1 - H_n$ , c'est-à-dire  $K_n = 1_{\{N < n\}} = 1_{\{N \leq n-1\}}$ . C'est également un processus prévisible, donc  $(K \cdot X)_n$  est une sous-martingale (car  $-X_n$  est une surmartingale). Or

$$(K \cdot X)_n = \sum_{m=1}^n 1_{\{N \leq m-1\}}(X_m - X_{m-1}) = \sum_{m=N+1}^n (X_m - X_{m-1}) = X_n - X_{N \wedge n}, \quad (5.2.5)$$

donc, comme  $N \wedge k = N$  presque sûrement,

$$\mathbb{E}(X_k) - \mathbb{E}(X_N) = \mathbb{E}((K \cdot X)_k) \geq \mathbb{E}((K \cdot X)_0) = 0, \quad (5.2.6)$$

d'où la seconde inégalité.  $\square$

Nous sommes maintenant en mesure de prouver l'inégalité de Doob, qui est très utile, notamment pour estimer le supremum d'un processus stochastique. Dans la suite, nous posons

$$\bar{X}_n = \max_{0 \leq m \leq n} X_m^+ = \max_{0 \leq m \leq n} X_m \vee 0. \quad (5.2.7)$$

**Théorème 5.2.4** (Inégalité de Doob). *Soit  $X_n$  une sous-martingale. Alors pour tout  $\lambda > 0$ , on a*

$$\mathbb{P}\{\bar{X}_n \geq \lambda\} \leq \frac{1}{\lambda} \mathbb{E}(X_n 1_{\{\bar{X}_n \geq \lambda\}}) \leq \frac{1}{\lambda} \mathbb{E}(X_n^+). \quad (5.2.8)$$

DÉMONSTRATION. Soit  $A = \{\bar{X}_n \geq \lambda\}$  et  $N = \inf\{m : X_m \geq \lambda\} \wedge n$ . Alors

$$\lambda \mathbb{P}(A) = \mathbb{E}(\lambda 1_A) \leq \mathbb{E}(X_N 1_A) \leq \mathbb{E}(X_n 1_A) \leq \mathbb{E}(X_n^+). \quad (5.2.9)$$

La première inégalité suit du fait que  $X_N \geq \lambda$  dans  $A$ . La seconde vient du fait que  $X_N = X_n$  sur  $A^c$  et que  $\mathbb{E}(X_N) \leq \mathbb{E}(X_n)$  par le corollaire 5.2.3. La dernière inégalité est triviale.  $\square$

L'intérêt de cette inégalité est qu'elle permet de majorer une quantité faisant intervenir tout le processus jusqu'au temps  $n$  par une quantité ne dépendant que de  $X_n$ , qui est souvent beaucoup plus simple à estimer.

**Exemple 5.2.5.** Soient  $\xi_1, \xi_2, \dots$  des variables indépendantes d'espérance nulle et variance finie, et soit  $X_n = \sum_{m=1}^n \xi_m$ . Alors  $X_n$  est une martingale (la preuve est la même que pour la marche aléatoire symétrique sur  $\mathbb{Z}$  dans l'exemple 4.1.2), donc  $X_n^2$  est une sous-martingale. Une application de l'inégalité de Doob donne l'*inégalité de Kolmogorov*

$$\mathbb{P}\left\{\max_{1 \leq m \leq n} |X_m| \geq \lambda\right\} = \mathbb{P}\left\{\max_{1 \leq m \leq n} X_m^2 \geq \lambda^2\right\} \leq \frac{1}{\lambda^2} \mathbb{E}(X_n^2) = \frac{1}{\lambda^2} \text{Var}(X_n). \quad (5.2.10)$$

Une autre application de l'inégalité de Doob est l'inégalité du maximum  $L^p$  :

**Théorème 5.2.6.** *Si  $X_n$  est une sous-martingale, alors pour tout  $p > 1$*

$$\mathbb{E}(\bar{X}_n^p) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}([X_n^+]^p). \quad (5.2.11)$$

DÉMONSTRATION. Soit  $M > 0$  une constante, que nous allons faire tendre vers l'infini à la fin de la preuve. Alors, par intégration par parties,

$$\begin{aligned} \mathbb{E}([\bar{X}_n \wedge M]^p) &= \int_0^\infty p\lambda^{p-1} \mathbb{P}\{\bar{X}_n \wedge M \geq \lambda\} \, d\lambda \\ &\leq \int_0^\infty p\lambda^{p-1} \frac{1}{\lambda} \mathbb{E}(X_n^+ 1_{\{\bar{X}_n \wedge M \geq \lambda\}}) \, d\lambda \\ &= \int_0^\infty p\lambda^{p-2} \int X_n^+ 1_{\{\bar{X}_n \wedge M \geq \lambda\}} \, d\mathbb{P} \, d\lambda \\ &= \int X_n^+ \int_0^{\bar{X}_n \wedge M} p\lambda^{p-2} \, d\lambda \, d\mathbb{P} \\ &= \frac{p}{p-1} \int X_n^+ [\bar{X}_n \wedge M]^{p-1} \, d\mathbb{P} \\ &= \frac{p}{p-1} \mathbb{E}(X_n^+ [\bar{X}_n \wedge M]^{p-1}) \\ &\leq \frac{p}{p-1} [\mathbb{E}(|X_n^+|^p)]^{1/p} [\mathbb{E}(|\bar{X}_n \wedge M|^p)]^{(p-1)/p}. \end{aligned} \quad (5.2.12)$$

La dernière inégalité provient de l'inégalité de Hölder. Divisant les deux côtés de l'inégalité par  $[\mathbb{E}(|\bar{X}_n \wedge M|^p)]^{(p-1)/p}$  et élevant à la puissance  $p$ , on trouve

$$\mathbb{E}(|\bar{X}_n \wedge M|^p) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}([X_n^+]^p), \quad (5.2.13)$$

et le résultat suit du théorème de la convergence dominée, en faisant tendre  $M$  vers l'infini.  $\square$

**Corollaire 5.2.7.** *Si  $Y_n$  est une martingale, alors*

$$\mathbb{E}\left(\left[\max_{0 \leq m \leq n} |Y_m|\right]^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|Y_n|^p). \quad (5.2.14)$$

### 5.3 Exercices

**Exercice 5.1.** Démontrer la Proposition 5.1.3 :

1. Si  $N$  et  $M$  sont des temps d'arrêt, alors  $N \wedge M$  et  $N \vee M$  sont des temps d'arrêt.
2. Si  $N_k$ ,  $k \in \mathbb{N}$ , est une suite de temps d'arrêt telle que  $N_k \nearrow N$ , alors  $N$  est un temps d'arrêt.

**Exercice 5.2.** Soient  $\xi_1, \xi_2, \dots$  des variables aléatoires i.i.d. telles que  $\mathbb{E}(\xi_m) = 0$ , et soit la martingale  $X_n = \sum_{m=1}^n \xi_m$ . On se donne  $\lambda > 0$ , et soit

$$P_n(\lambda) = \mathbb{P}\left\{\max_{1 \leq m \leq n} |X_m| \geq \lambda\right\}.$$

1. On suppose les  $\xi_m$  de variance finie. Donner une majoration de  $P_n(\lambda)$  en appliquant l'inégalité de Doob à  $X_n^2$ .
2. Améliorer la borne précédente en appliquant l'inégalité de Doob à  $(X_n + c)^2$  et en optimisant sur  $c$ .
3. On suppose que les  $\xi_m$  suivent une loi normale centrée réduite. Majorer  $P_n(\lambda)$  en appliquant l'inégalité de Doob à  $e^{cX_n^2}$  et en optimisant sur  $c$ .
4. Pour des  $\xi_m$  normales centrées réduites, majorer  $\mathbb{P}\{\max_{1 \leq m \leq n} X_m \geq \lambda\}$  en appliquant l'inégalité de Doob à  $e^{cX_n}$  et en optimisant sur  $c$ .



# Chapitre 6

## Théorèmes de convergence

Une propriété importante des suites non-décroissantes de réels est que si elles sont bornées, alors elles convergent. Les sous-martingales sont un analogue stochastique de telles suites : si elles sont bornées, alors elles convergent dans un sens approprié. Ceci donne une méthode très simple pour étudier le comportement asymptotique de certains processus stochastiques.

Il faut cependant faire attention au fait qu'il existe différentes notions de convergence de suites de variables aléatoires. Dans cette section, nous allons d'abord examiner la convergence de sous-martingales au sens presque sûr, qui est le plus fort. Ensuite nous verrons ce qui se passe dans le cas plus faible de la convergence  $L^p$ . Il s'avère que le cas  $p > 1$  est relativement simple à contrôler, alors que le cas  $p = 1$  est plus difficile.

### 6.1 Rappel: Notions de convergence

Commençons par rappeler trois des notions de convergence d'une suite de variables aléatoires (une quatrième notion, celle de convergence en loi, ne jouera pas de rôle important ici).

**Définition 6.1.1** (Convergence d'une suite de variables aléatoires). *Soient  $X$  et  $\{X_n\}_{n \geq 0}$  des variables aléatoires réelles, définies sur un même espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ .*

1. *On dit que  $X_n$  converge presque sûrement vers  $X$  si*

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1 .$$

2. *Si  $p > 0$ , on dénote par  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  l'ensemble des variables aléatoires  $X$  telles que  $\mathbb{E}(|X|^p) < \infty$ . Si  $X_n, X \in L^p$ , on dit que  $X_n$  converge dans  $L^p$  vers  $X$  si*

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0 .$$

3. *On dit que  $X_n$  converge en probabilité vers  $X$  si*

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0$$

*pour tout  $\varepsilon > 0$ .*

Les principaux liens entre ces trois notions de convergence sont résumées dans la proposition suivante.

**Proposition 6.1.2.**

1. La convergence presque sûre implique la convergence en probabilité.
2. La convergence dans  $L^p$  implique la convergence en probabilité.
3. La convergence dans  $L^p$  implique la convergence dans  $L^q$  pour tout  $q < p$ .
4. Si  $X_n \rightarrow X$  presque sûrement et  $|X_n| \leq Y \forall n$  avec  $Y \in L^p$ , alors  $X_n \rightarrow X$  dans  $L^p$ .

DÉMONSTRATION.

1. Soit  $Y_n = 1_{\{|X_n - X| \geq \varepsilon\}}$ . Si  $X_n \rightarrow X$  presque sûrement, alors  $X_n \rightarrow X$  presque partout, donc  $Y_n \rightarrow 0$  presque sûrement. Par le théorème de la convergence dominée (qui s'applique car  $Y_n \leq 1$ ), on a  $\mathbb{P}\{|X_n - X| \geq \varepsilon\} = \mathbb{E}(Y_n) \rightarrow 0$ .
2. Par l'inégalité de Markov,  $\mathbb{P}\{|X_n - X| \geq \varepsilon\} \leq \varepsilon^{-p} \mathbb{E}(|X_n - X|^p) \rightarrow 0$ .
3. Si  $q < p$ , l'inégalité de Hölder implique  $\mathbb{E}(|X_n - X|^q) \leq \mathbb{E}(|X_n - X|^p)^{q/p} \rightarrow 0$ .
4. On a  $|X_n - X|^p \leq (|X_n| + |X|)^p \leq (2Y)^p \leq 2^p Y^p \in L^1$ . Comme  $|X_n - X|^p \rightarrow 0$  presque sûrement, le théorème de la convergence dominée permet d'écrire

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = \mathbb{E}\left(\lim_{n \rightarrow \infty} |X_n - X|^p\right) = 0. \quad (6.1.1)$$

□

Voici deux contre-exemples classiques à l'équivalence entre ces différentes notions de convergence :

**Exemple 6.1.3.** Soit  $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ , où  $\mathcal{B}_{[0,1]}$  est la tribu des boréliens et  $\lambda$  la mesure de Lebesgue.

1. Soit  $X_n = n^{1/p} 1_{[0, 1/n]}$ . Alors  $X_n(\omega) \rightarrow 0$  pour tout  $\omega > 0$ , donc  $X_n \rightarrow 0$  presque sûrement. De plus, pour tout  $\varepsilon > 0$  on a  $\mathbb{P}\{|X_n| > \varepsilon\} \leq 1/n$ , donc  $X_n \rightarrow 0$  en probabilité. Par contre, on a  $\mathbb{E}(|X_n|^p) = \int_0^{1/n} (n^{1/p})^p dx = 1$ , donc  $X_n$  ne tend pas vers 0 dans  $L^p$ .
2. Décomposons tout entier  $n$  comme  $n = 2^m + k$  avec  $m \in \mathbb{N}$  et  $0 \leq k < 2^m$ , et soit  $X_n = 1_{[k2^{-m}, (k+1)2^{-m}]}$ . Alors  $\mathbb{P}\{|X_n| > \varepsilon\} \leq 2^{-m}$  pour tout  $\varepsilon > 0$ , donc  $X_n \rightarrow 0$  en probabilité. De plus, pour tout  $p > 0$ ,  $\mathbb{E}(|X_n|^p) \leq 2^{-m}$ , donc  $X_n \rightarrow 0$  dans  $L^p$ . Par contre,  $X_n(\omega)$  ne converge pour aucun  $\omega$  (on peut trouver des sous-suites de  $n$  telles que  $X_n(\omega) = 1$  pour tout  $n$ , et aussi telles que  $X_n(\omega) = 0$  pour tout  $n$ ), donc  $X_n$  ne converge pas presque sûrement.

## 6.2 Convergence presque sûre

Considérons un investisseur suivant la stratégie suivante. Il se fixe deux seuils  $a < b$ . Il achète un nombre fixé de parts d'action lorsque leur prix est inférieur à  $a$ . Puis il attend que le prix de l'action atteigne ou dépasse la valeur  $b$  pour revendre ses parts. Ensuite, il attend que le prix retombe au-dessous de  $a$  pour racheter des parts, et ainsi de suite. Si  $X_n$  désigne le prix de l'action au temps  $n$ , la stratégie peut être décrite formellement comme suit. On pose  $N_0 = -1$  et, pour tout  $k \geq 1$ , on introduit les temps d'arrêt

$$\begin{aligned} N_{2k-1} &= \inf\{m > N_{2k-2} : X_m \leq a\}, \\ N_{2k} &= \inf\{m > N_{2k-1} : X_m \geq b\}. \end{aligned} \quad (6.2.1)$$

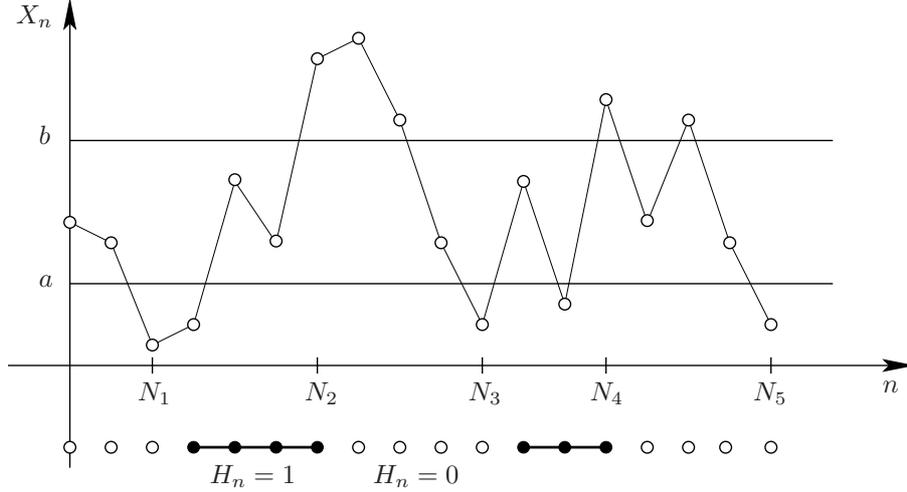


FIGURE 6.1. Prix de l'action et stratégie de l'investisseur.

La stratégie correspond à la suite prévisible

$$H_m = \begin{cases} 1 & \text{s'il existe } k \text{ tel que } N_{2k-1} < m \leq N_{2k}, \\ 0 & \text{sinon.} \end{cases} \quad (6.2.2)$$

Jusqu'au temps  $n$ , l'investisseur aura revendu ses actions un nombre  $U_n$  de fois, donné par

$$U_n = \sup\{k : N_{2k} \leq n\}. \quad (6.2.3)$$

Au temps  $n$ , le gain de l'investisseur sera donc supérieur ou égal à  $(b-a)U_n$ . Le résultat suivant donne une borne supérieure sur le gain moyen dans le cas où le processus est favorable ou équitable.

**Proposition 6.2.1.** *Si  $X_n$  est une sous-martingale, alors*

$$(b-a)\mathbb{E}(U_n) \leq \mathbb{E}((X_n - a)^+) - \mathbb{E}((X_0 - a)^+). \quad (6.2.4)$$

**DÉMONSTRATION.** Soit  $Y_n = a + (X_n - a)^+ = X_n \vee a$ . Par le corollaire 4.2.3,  $Y_n$  est aussi une sous-martingale. De plus,  $Y_n$  et  $X_n$  effectuent le même nombre de transitions de  $a$  vers  $b$ . On a

$$(b-a)U_n \leq (H \cdot Y)_n = \sum_{m=1}^n H_m(Y_m - Y_{m-1}). \quad (6.2.5)$$

En effet, le membre de droite est supérieur ou égal au profit fait au temps  $n$ . Soit alors  $K_m = 1 - H_m$ . Comme  $Y_n - Y_0 = (H \cdot Y)_n + (K \cdot Y)_n$ , et que  $(K \cdot Y)_n$  est une sous-martingale en vertu de la Proposition 4.3.2, on a

$$\mathbb{E}((K \cdot Y)_n) \geq \mathbb{E}((K \cdot Y)_0) = 0, \quad (6.2.6)$$

et donc  $\mathbb{E}((H \cdot Y)_n) \leq \mathbb{E}(Y_n - Y_0)$ , d'où le résultat.  $\square$

Cette majoration nous permet de démontrer un premier résultat de convergence.

**Théorème 6.2.2** (Convergence presque sûre). *Soit  $X_n$  une sous-martingale telle que  $\sup_n \mathbb{E}(X_n^+) < \infty$ . Alors  $X_n$  converge presque sûrement, lorsque  $n \rightarrow \infty$ , vers une variable aléatoire  $X$  satisfaisant  $\mathbb{E}(|X|) < \infty$ .*

DÉMONSTRATION. Comme  $(X - a)^+ \leq X^+ + |a|$ , la proposition précédente implique

$$\mathbb{E}(U_n) \leq \frac{|a| + \mathbb{E}(X_n^+)}{b - a}. \quad (6.2.7)$$

Lorsque  $n \rightarrow \infty$ , la suite croissante  $U_n$  tend vers le nombre total  $U$  de passages de  $a$  à  $b$  de la suite des  $X_n$ . Si  $\mathbb{E}(X_n^+)$  est borné, il suit que  $\mathbb{E}(U) < \infty$ , et donc que  $U$  est fini presque sûrement. Ceci étant vrai pour tout choix de rationnels  $a < b$ , l'événement

$$\bigcup_{a,b \in \mathbb{Q}} \{\omega : \liminf X_n(\omega) < a < b < \limsup X_n(\omega)\} \quad (6.2.8)$$

a une probabilité nulle. Il suit que  $\liminf X_n = \limsup X_n$  presque sûrement, et donc que  $\lim_{n \rightarrow \infty} X_n =: X$  existe presque sûrement. Le lemme de Fatou nous assure que  $\mathbb{E}(X^+) \leq \liminf \mathbb{E}(X_n^+) < \infty$ , et donc  $X < \infty$  presque sûrement. Enfin, pour montrer que  $X > -\infty$  presque sûrement, on décompose

$$\mathbb{E}(X_n^-) = \mathbb{E}(X_n^+) - \mathbb{E}(X_n) \leq \mathbb{E}(X_n^+) - \mathbb{E}(X_0), \quad (6.2.9)$$

et on applique à nouveau le lemme de Fatou pour montrer que  $\mathbb{E}(X^-) < \infty$ .  $\square$

**Corollaire 6.2.3.** *Soit  $X_n$  une surmartingale positive. Alors  $X_n$  converge presque sûrement, lorsque  $n \rightarrow \infty$ , vers une variable aléatoire  $X$  avec  $\mathbb{E}(X) \leq \mathbb{E}(X_0)$ .*

DÉMONSTRATION.  $Y_n = -X_n$  est une sous-martingale bornée supérieurement par 0, avec  $\mathbb{E}(Y_n^+) = 0$ , donc elle converge. Comme  $\mathbb{E}(X_0) \geq \mathbb{E}(X_n)$ , l'inégalité suit du lemme de Fatou.  $\square$

#### Exemple 6.2.4.

1. Considérons l'urne de Polya introduite dans l'exemple 4.1.2. La proportion  $X_n$  de boules rouges est une martingale positive, qui est bornée supérieurement par 1. Par le théorème et le corollaire,  $X_n$  converge presque sûrement vers une variable aléatoire  $X$ , d'espérance inférieure ou égale à 1. On peut montrer que selon le nombre initial  $r$  et  $v$  de boules et le paramètre  $c$ , la loi de  $X$  est soit uniforme, soit une loi bêta, de densité proportionnelle à  $x^{r/c-1}(1-x)^{v/c-1}$ .
2. Soit  $S_n$  une marche aléatoire symétrique sur  $\mathbb{Z}$ ,  $N = N_0$  le temps de première atteinte de 0, et  $X_n = S_{n \wedge N}$  la marche arrêtée en 0. Par la Proposition 5.2.2,  $X_n$  est une martingale non-négative. Par le corollaire, elle converge vers une limite  $X$  finie presque sûrement. Cette limite doit être nulle : En effet, si  $X_n = k > 0$ , alors  $X_{n+1} = k \pm 1$ , ce qui contredirait la convergence. On a donc montré que la marche aléatoire partant de 1 atteint 0 presque sûrement. On notera cependant que comme  $X_n$  est une martingale, on a  $\mathbb{E}(X_n) = \mathbb{E}(X_0) = 1$ , donc que  $X_n$  ne converge pas dans  $L^1$ .

### 6.3 Convergence dans $L^p$ , $p > 1$

La convergence d'une martingale dans  $L^p$  suit de manière simple de l'inégalité du maximum  $L^p$  :

**Théorème 6.3.1** (Convergence d'une martingale dans  $L^p$ ). *Soit  $X_n$  une martingale telle que  $\sup_n \mathbb{E}(|X_n|^p) < \infty$  pour un  $p > 1$ . Alors  $X_n$  converge vers une variable aléatoire  $X$  presque sûrement et dans  $L^p$ .*

DÉMONSTRATION. On a  $(\mathbb{E}(X_n^+))^p \leq (\mathbb{E}(|X_n|))^p \leq \mathbb{E}(|X_n|^p)$ . Donc par le Théorème 6.2.2,  $X_n$  converge presque sûrement vers une variable  $X$ . Par le corollaire (5.2.7),

$$\mathbb{E}\left(\left[\sup_{0 \leq m \leq n} |X_m|\right]^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_n|^p). \quad (6.3.1)$$

Faisant tendre  $n$  vers l'infini, le théorème de la convergence monotone montre que  $\sup_n |X_n|$  est dans  $L^p$ . Comme  $|X_n - X|^p \leq (2 \sup_n |X_n|)^p$ , le théorème de la convergence dominée montre que  $\mathbb{E}(|X_n - X|^p) \rightarrow 0$ .  $\square$

Dans la suite, nous considérons plus particulièrement le cas  $p = 2$ . Rappelons que si une martingale  $X_n$  est dans  $L^2$ , on peut définir son processus croissant

$$\langle X \rangle_n = \sum_{m=1}^n \mathbb{E}((X_m - X_{m-1})^2 | \mathcal{F}_{m-1}). \quad (6.3.2)$$

La croissance implique que

$$\lim_{n \rightarrow \infty} \langle X \rangle_n =: \langle X \rangle_\infty \quad (6.3.3)$$

existe dans  $\mathbb{R}_+ \cup \{\infty\}$ . Cette quantité s'interprète comme la variance totale de la trajectoire  $X_n(\omega)$ .

**Proposition 6.3.2.** *Soit  $X_n$  une martingale dans  $L^2$  telle que  $X_0 = 0$ . Alors*

$$\mathbb{E}\left(\sup_n X_n^2\right) \leq 4\mathbb{E}(\langle X \rangle_\infty). \quad (6.3.4)$$

DÉMONSTRATION. L'inégalité du maximum  $L^2$  donne

$$\mathbb{E}\left(\sup_{0 \leq m \leq n} X_m^2\right) \leq 4\mathbb{E}(X_n^2) = 4\mathbb{E}(\langle X \rangle_n), \quad (6.3.5)$$

puisque  $\mathbb{E}(X_n^2) = \mathbb{E}(M_n) + \mathbb{E}(\langle X \rangle_n)$  et  $\mathbb{E}(M_n) = \mathbb{E}(M_0) = \mathbb{E}(X_0^2) = 0$ . Le résultat suit alors du théorème de convergence monotone.  $\square$

**Théorème 6.3.3.** *La limite  $\lim_{n \rightarrow \infty} X_n(\omega)$  existe et est finie presque sûrement sur l'ensemble  $\{\omega : \langle X \rangle_\infty(\omega) < \infty\}$ .*

DÉMONSTRATION. Soit  $a > 0$ . Comme  $\langle X \rangle_{n+1} \subseteq \mathcal{F}_n$ ,  $N = \inf\{n : \langle X \rangle_{n+1} > a^2\}$  est un temps d'arrêt. Comme  $\langle X \rangle_{N \wedge n} < a^2$ , la proposition ci-dessus appliquée à  $X_{N \wedge n}$  donne

$$\mathbb{E}\left(\sup_n |X_{N \wedge n}|^2\right) \leq 4a^2. \quad (6.3.6)$$

Par conséquent, le théorème 6.3.1 avec  $p = 2$  implique que la limite de  $X_{N \wedge n}$  existe et est finie presque sûrement. Le résultat suit alors du fait que  $a$  est arbitraire.  $\square$

## 6.4 Convergence dans $L^1$

La discussion de la convergence d'une martingale dans  $L^1$  nécessite la notion d'intégrabilité uniforme.

**Définition 6.4.1** (Intégrabilité uniforme). *Une collection  $\{X_i\}_{i \in I}$  de variables aléatoires est dite uniformément intégrable si*

$$\lim_{M \rightarrow \infty} \left( \sup_{i \in I} \mathbb{E}(|X_i| 1_{\{|X_i| > M\}}) \right) = 0. \quad (6.4.1)$$

On remarque qu'en prenant  $M$  assez grand pour que le supremum soit inférieur à 1, on obtient

$$\sup_{i \in I} \mathbb{E}(|X_i|) \leq \sup_{i \in I} \mathbb{E}(|X_i| 1_{\{|X_i| \leq M\}}) + \sup_{i \in I} \mathbb{E}(|X_i| 1_{\{|X_i| > M\}}) \leq M + 1 < \infty. \quad (6.4.2)$$

Nous commençons par un résultat général montrant qu'il existe de très vastes familles de variables uniformément intégrables.

**Proposition 6.4.2.** *Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé et  $X \in L^1$ . Alors la famille*

$$\{\mathbb{E}(X|\mathcal{F}_i) : \mathcal{F}_i \text{ sous-tribu de } \mathcal{F}\} \quad (6.4.3)$$

*est uniformément intégrable.*

DÉMONSTRATION. Pour tout  $\varepsilon > 0$ , le théorème de la convergence dominée implique qu'on peut trouver  $\delta > 0$  tel que si  $\mathbb{P}(A) \leq \delta$ , alors  $\mathbb{E}(|X|1_A) \leq \varepsilon$ . Soit  $M = \mathbb{E}(|X|)/\delta$  et  $A_i = \{\mathbb{E}(|X|\mathcal{F}_i) > M\} \in \mathcal{F}_i$ . Par l'inégalité de Markov,

$$\mathbb{P}(A_i) \leq \frac{1}{M} \mathbb{E}(\mathbb{E}(|X|\mathcal{F}_i)) = \frac{1}{M} \mathbb{E}(|X|) \leq \delta.$$

Par l'inégalité de Jensen, il suit

$$\mathbb{E}\left(|\mathbb{E}(X|\mathcal{F}_i)| 1_{\{|\mathbb{E}(X|\mathcal{F}_i)| > M\}}\right) \leq \mathbb{E}(\mathbb{E}(|X|\mathcal{F}_i) 1_{A_i}) = \mathbb{E}(|X| 1_{A_i}) \leq \varepsilon.$$

Comme  $\varepsilon$  est arbitraire (et  $M$  ne dépend pas de  $i$ ), le résultat est prouvé.  $\square$

Le lien entre intégrabilité uniforme et convergence dans  $L^1$  est expliqué par le résultat suivant.

**Théorème 6.4.3.** *Si  $X_n \rightarrow X$  en probabilité, alors les trois conditions suivantes sont équivalentes :*

1. *La famille  $\{X_n : n \geq 0\}$  est uniformément intégrable;*
2.  *$X_n \rightarrow X$  dans  $L^1$ ;*
3.  *$\mathbb{E}(|X_n|) \rightarrow \mathbb{E}(|X|) < \infty$ .*

DÉMONSTRATION.

1.  $\Rightarrow$  2. Soit

$$\varphi_M(x) = \begin{cases} M & \text{si } x \geq M, \\ x & \text{si } |x| \leq M, \\ -M & \text{si } x \leq -M. \end{cases}$$

L'inégalité triangulaire implique

$$|X_n - X| \leq |X_n - \varphi_M(X_n)| + |\varphi_M(X_n) - \varphi_M(X)| + |\varphi_M(X) - X|.$$

Prenant l'espérance, et utilisant le fait que  $|\varphi_M(Y) - Y| \leq |Y|1_{\{|Y| \geq M\}}$ , il vient

$$\mathbb{E}(|X_n - X|) \leq \mathbb{E}(|\varphi_M(X_n) - \varphi_M(X)|) + \mathbb{E}(|X_n|1_{\{|X_n| \geq M\}}) + \mathbb{E}(|X|1_{\{|X| \geq M\}}).$$

Le premier terme tend vers 0 lorsque  $n \rightarrow \infty$  par convergence dominée et le fait que  $\varphi_M(X_n) \rightarrow \varphi_M(X)$  en probabilité. L'intégrabilité uniforme implique que pour tout  $\varepsilon > 0$ , le second terme est plus petit que  $\varepsilon$  pour  $M$  assez grand. L'intégrabilité uniforme implique aussi que  $\sup_n \mathbb{E}(|X_n|) < \infty$ , donc par le lemme de Fatou que  $\mathbb{E}(|X|) < \infty$ . Le troisième terme peut donc également être rendu plus petit que  $\varepsilon$  en prenant  $M$  assez grand, ce qui prouve 2.

2.  $\Rightarrow$  3. Par l'inégalité de Jensen,

$$|\mathbb{E}(|X_n|) - \mathbb{E}(|X|)| \leq \mathbb{E}(|X_n| - |X|) \leq \mathbb{E}(|X_n - X|) \rightarrow 0.$$

3.  $\Rightarrow$  1. Soit

$$\psi_M(x) = \begin{cases} x & \text{si } 0 \leq x \leq M-1, \\ (M-1)(M-x) & \text{si } M-1 \leq x \leq M, \\ 0 & \text{si } x \geq M. \end{cases}$$

Pour tout  $\varepsilon > 0$ , le théorème de la convergence dominée montre que  $\mathbb{E}(|X|) - \mathbb{E}(\psi_M(|X|)) \leq \varepsilon/2$  pour  $M$  assez grand. De plus,  $\mathbb{E}(\psi_M(|X_n|)) \rightarrow \mathbb{E}(\psi_M(|X|))$ , donc par 3.

$$\mathbb{E}(|X_n|1_{\{|X_n| > M\}}) \leq \mathbb{E}(|X_n|) - \mathbb{E}(\psi_M(|X_n|)) \leq \mathbb{E}(|X|) - \mathbb{E}(\psi_M(|X|)) + \frac{\varepsilon}{2} < \varepsilon$$

pour les  $n$  plus grands qu'un  $n_0(\varepsilon)$ . En augmentant encore  $M$ , on peut rendre  $\mathbb{E}(|X_n|1_{\{|X_n| > M\}})$  inférieur à  $\varepsilon$  pour les  $n \leq n_0(\varepsilon)$  également, ce qui montre l'intégrabilité uniforme.  $\square$

Il est maintenant aisé d'appliquer ce résultat au cas des sous-martingales.

**Théorème 6.4.4** (Convergence d'une sous-martingale dans  $L^1$ ). *Si  $X_n$  est une sous-martingale, alors les trois conditions suivantes sont équivalentes :*

1.  $\{X_n\}_{n \geq 0}$  est uniformément intégrable;
2.  $X_n$  converge presque sûrement et dans  $L^1$ ;
3.  $X_n$  converge dans  $L^1$ .

DÉMONSTRATION.

1.  $\Rightarrow$  2. On a  $\sup_n \mathbb{E}(X_n) < \infty$ , donc  $X_n \rightarrow X$  presque sûrement par le théorème 6.2.2, et le théorème 6.4.3 implique que  $X_n \rightarrow X$  dans  $L^1$ .

2.  $\Rightarrow$  3. Trivial.

3.  $\Rightarrow$  1. La convergence dans  $L^1$  implique la convergence en probabilité, et donc le théorème 6.4.3 permet de conclure.  $\square$

Dans le cas particulier des martingales, on a

**Théorème 6.4.5** (Convergence d'une martingale dans  $L^1$ ). *Si  $X_n$  est une martingale, alors les quatre conditions suivantes sont équivalentes :*

1.  $\{X_n\}_{n \geq 0}$  est uniformément intégrable;
2.  $X_n$  converge presque sûrement et dans  $L^1$ ;
3.  $X_n$  converge dans  $L^1$ ;
4. il existe une variable aléatoire intégrable  $X$  telle que  $X_n = \mathbb{E}(X|\mathcal{F}_n)$ .

Dans ce cas  $X$  est la limite de  $X_n$  dans  $L^1$ .

DÉMONSTRATION.

1.  $\Rightarrow$  2. Suit du théorème précédent.
2.  $\Rightarrow$  3. Trivial.
3.  $\Rightarrow$  4. Si  $X_n \rightarrow X$  dans  $L^1$ , alors pour tout  $A \in \mathcal{F}$ ,  $\mathbb{E}(X_n 1_A) \rightarrow \mathbb{E}(X 1_A)$  puisque

$$|\mathbb{E}(X_m 1_A) - \mathbb{E}(X 1_A)| \leq \mathbb{E}(|X_m 1_A - X 1_A|) \leq \mathbb{E}(|X_m - X|) \rightarrow 0.$$

En particulier, si  $A \in \mathcal{F}_n$ , alors pour tout  $m > n$ ,

$$\mathbb{E}(X_n 1_A) = \mathbb{E}(\mathbb{E}(X_m|\mathcal{F}_n) 1_A) = \mathbb{E}(X_m 1_A).$$

Comme  $\mathbb{E}(X_m 1_A) \rightarrow \mathbb{E}(X 1_A)$ , on a en fait  $\mathbb{E}(X_n 1_A) = \mathbb{E}(X 1_A)$  pour tout  $A \in \mathcal{F}_n$ .

Mais ceci est la définition de  $X_n = \mathbb{E}(X|\mathcal{F}_n)$ .

4.  $\Rightarrow$  1. Suit de la proposition 6.4.2.  $\square$

## 6.5 Loi 0 – 1 de Lévy

Les lois 0 – 1 jouent un rôle important pour les suites de variables aléatoires, et ont parfois des conséquences surprenantes. La plus connue est la loi 0 – 1 de Kolmogorov. Soit  $X_n$  une suite de variables aléatoires indépendantes,  $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \dots)$ , et  $\mathcal{T} = \bigcap_n \mathcal{F}'_n$  la tribu terminale. Intuitivement, les événements de  $\mathcal{T}$  sont ceux dont l'occurrence n'est pas affectée par la modification d'un nombre fini de  $X_n$ , comme par exemple l'événement  $\{\sum_{i=1}^{\infty} X_i \text{ existe}\}$ . La loi 0 – 1 de Kolmogorov affirme alors que si  $A \in \mathcal{T}$ , alors  $\mathbb{P}(A) = 0$  ou 1. Ceci permet notamment de prouver la loi forte des grands nombres.

Nous écrivons  $\mathcal{F}_n \nearrow \mathcal{F}_\infty$  si  $\mathcal{F}_n$  est une filtration telle que  $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$ .

**Théorème 6.5.1.** *Si  $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ , alors*

$$\mathbb{E}(X|\mathcal{F}_n) \rightarrow \mathbb{E}(X|\mathcal{F}_\infty) \tag{6.5.1}$$

*presque sûrement et dans  $L^1$  lorsque  $n \rightarrow \infty$ .*

DÉMONSTRATION. Comme pour  $m > n$ ,  $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_m)|\mathcal{F}_n) = \mathbb{E}(X|\mathcal{F}_n)$ ,  $Y_n = \mathbb{E}(X|\mathcal{F}_n)$  est une martingale. La Proposition 6.4.2 montre que  $Y_n$  est uniformément intégrable, donc par le Théorème 6.4.5 elle converge presque sûrement et dans  $L^1$  vers une variable aléatoire  $Y_\infty$ , telle que  $Y_n = \mathbb{E}(Y_\infty|\mathcal{F}_n)$ . Par définition de l'espérance conditionnelle,

$$\int_A X \, d\mathbb{P} = \int_A Y_\infty \, d\mathbb{P}$$

pour tout  $A \in \mathcal{F}_n$ . En prenant la limite  $n \rightarrow \infty$ , justifiée par le théorème  $\pi$ - $\lambda$  de théorie de l'intégration, on obtient  $\mathbb{E}(X|\mathcal{F}_\infty) = Y_\infty$  puisque  $Y_\infty \subseteq \mathcal{F}_\infty$ .  $\square$

**Corollaire 6.5.2** (Loi 0 – 1 de Lévy). *Si  $\mathcal{F}_n \nearrow \mathcal{F}_\infty$  et  $A \in \mathcal{F}_\infty$  alors  $\mathbb{E}(1_A|\mathcal{F}_n) \rightarrow 1_A$  presque sûrement.*

**Exemple 6.5.3.** Supposons que  $X_n$  est une suite de variables aléatoires indépendantes, et soit  $A \in \mathcal{T}$  un événement de la tribu terminale. Pour tout  $n$ ,  $A$  est indépendant de  $\mathcal{F}_n$ , donc  $\mathbb{E}(1_A|\mathcal{F}_n) = \mathbb{E}(1_A) = \mathbb{P}(A)$ . Faisant tendre  $n$  vers l'infini, on obtient  $1_A = \mathbb{P}(A)$ , d'où  $\mathbb{P}(A) = 0$  ou  $1$ , ce qui n'est autre que la loi 0 – 1 de Kolmogorov.

## 6.6 Exercices

**Exercice 6.1** (Le processus de Galton–Watson). On considère le processus de Galton–Watson  $Z_n$  introduit dans l'exercice 4.4.

1. Montrer que  $X_n = Z_n/\mu^n$  converge dans  $L^2$  vers une variable aléatoire  $X$  d'espérance égale à 1.
2. On suppose  $p_0 > 0$ . A l'aide de la loi 0 – 1 de Lévy, montrer que  $Z_n$  converge presque sûrement vers une variable aléatoire  $Z_\infty : \Omega \rightarrow \{0, +\infty\}$ .

**Indications :** Soit  $A = \{\lim_{n \rightarrow \infty} Z_n < \infty\}$  et  $B = \{\exists n : Z_n = 0\}$ . Montrer que dans  $A$ , on a  $\lim_{n \rightarrow \infty} \mathbb{E}(1_B|\mathcal{F}_n) = 1_B$ , et en déduire que  $A \subset B$ . Conclure en établissant que  $Z_\infty = \infty$  dans  $A^c$  et  $Z_\infty = 0$  dans  $B$ .

**Exercice 6.2.** Un joueur dispose initialement de la somme  $X_0 = 1$ . Il joue à un jeu de hasard, dans lequel il mise à chaque tour une proportion  $\lambda$  de son capital, avec  $0 < \lambda \leq 1$ . Il a une chance sur deux de gagner le double de sa mise, sinon il perd sa mise.

L'évolution du capital  $X_n$  en fonction du temps  $n$  est décrite par

$$X_{n+1} = (1 - \lambda)X_n + \lambda X_n \xi_n \quad (n \geq 0)$$

où les  $\xi_n$  sont i.i.d., avec  $\mathbb{P}\{\xi_n = 2\} = \mathbb{P}\{\xi_n = 0\} = 1/2$ .

1. Montrer que  $X_n$  est une martingale.
2. Calculer  $\mathbb{E}(X_n)$ .
3. Discuter la convergence presque sûre de  $X_n$  lorsque  $n \rightarrow \infty$ .
4. Calculer  $\mathbb{E}(X_n^2)$  par récurrence sur  $n$ .
5. Que peut-on en déduire sur la convergence dans  $L^2$  de  $X_n$ ?
6. Déterminer le processus croissant  $\langle X \rangle_n$ .
7. On suppose que le joueur mise à chaque tour la totalité de son capital, c'est-à-dire  $\lambda = 1$ .
  - (a) Calculer explicitement la loi de  $X_n$ .
  - (b) Déterminer la limite presque sûre de  $X_n$ .
  - (c) Discuter la convergence de  $X_n$  dans  $L^1$ .  
Les  $X_n$  sont-ils uniformément intégrables?
  - (d) Commenter ces résultats - est-ce que vous joueriez à ce jeu?

**Exercice 6.3.** Soit  $X_0 = 1$ . On définit une suite  $\{X_n\}_{n \in \mathbb{N}}$  récursivement en posant que pour tout  $n \geq 1$ ,  $X_n$  suit la loi uniforme sur  $]0, 2X_{n-1}]$  (c-à-d  $X_n = 2U_n X_{n-1}$  où les  $U_n$  sont i.i.d. de loi uniforme sur  $]0, 1]$ ).

1. Montrer que  $X_n$  est une martingale.
2. Calculer le processus croissant  $\langle X \rangle_n$  et discuter la convergence de  $X_n$  dans  $L^2$ .

3. Discuter la convergence presque sûre de  $X_n$ .
4. Déterminer la limite presque sûre de  $X_n$ .  
*Indication* : Considérer  $Y_n = \log(X_n)$  ainsi que  $Z_n = Y_n - \mathbb{E}(Y_n)$ .
5. Discuter la convergence de  $X_n$  dans  $L^1$ .

**Exercice 6.4.** Soit  $U$  une variable aléatoire uniforme sur  $[0, 1]$ . Soit  $f : [0, 1] \rightarrow \mathbb{R}$  une fonction uniformément lipschitzienne. Soit  $I_{k,n} = [k2^{-n}, (k+1)2^{-n}[$  et soit  $\mathcal{F}_n$  la tribu sur  $\Omega = [0, 1]$  engendrée par les  $I_{k,n}$  pour  $k = 0, \dots, 2^n - 1$ .

On pose

$$X_n = \sum_{k=0}^{2^n-1} \frac{f((k+1)2^{-n}) - f(k2^{-n})}{2^{-n}} 1_{\{U \in I_{k,n}\}}.$$

1. Montrer que  $X_n$  est une martingale par rapport à  $\mathcal{F}_n$ .
2. Montrer que  $X_n$  converge presque sûrement. On note la limite  $X_\infty$ .
3. Discuter la convergence de  $X_n$  dans  $L^1$ .
4. Montrer que pour tout  $0 \leq a < b \leq 1$ ,

$$\mathbb{E}(X_\infty 1_{\{U \in [a,b]\}}) = f(b) - f(a).$$

5. On suppose  $f$  de classe  $C^1$ . A l'aide de 4., expliciter  $X_\infty(\omega)$  (si l'on note  $U(\omega) = \omega$  pour tout  $\omega \in \Omega$ ).

## Partie II

# Processus en temps continu



# Chapitre 7

## Le mouvement Brownien

Le mouvement Brownien joue un rôle fondamental dans la théorie des processus stochastiques en temps continu. Bien qu'il soit relativement simple à définir, il jouit de nombreuses propriétés remarquables, et parfois surprenantes. Le mouvement Brownien a été étudié de manière très approfondie, et bien qu'il existe encore des questions ouvertes, c'est l'un des processus les mieux compris. Il joue un rôle important dans la construction de processus stochastiques plus généraux.

Dans ce chapitre, nous présenterons une construction du mouvement Brownien, et une petite sélection de ses nombreuses propriétés, plus particulièrement celles en rapport avec les martingales.

### 7.1 Limite d'échelle d'une marche aléatoire

Nous commençons par introduire le mouvement Brownien de manière heuristique. Considérons pour cela la marche aléatoire symétrique  $X_n$  sur  $\mathbb{Z}$ . Elle peut être représentée sous la forme

$$X_n = \sum_{i=1}^n \xi_i, \quad (7.1.1)$$

où les  $\xi_i$  sont des variables aléatoires i.i.d., prenant valeurs 1 et  $-1$  avec probabilité  $1/2$ . On en déduit aisément les propriétés suivantes:

1.  $\mathbb{E}(X_n) = 0$  pour tout  $n$ ;
2.  $\text{Var}(X_n) = n$ ;
3.  $X_n$  prend ses valeurs dans  $\{-n, -n+2, \dots, n-2, n\}$  avec

$$\mathbb{P}\{X_n = k\} = \frac{1}{2^n} \frac{n!}{\left(\frac{n+k}{2}\right)! \left(\frac{n-k}{2}\right)!}. \quad (7.1.2)$$

4. *Propriétés des incréments indépendants*: pour tout  $n > m \geq 0$ ,  $X_n - X_m$  est indépendant de  $X_1, \dots, X_m$ ;
5. *Propriétés des incréments stationnaires*: pour tout  $n > m \geq 0$ ,  $X_n - X_m$  a la même loi que  $X_{n-m}$ .

Considérons alors la suite de processus

$$B_t^{(n)} = \frac{1}{\sqrt{n}} X_{[nt]}, \quad t \in \mathbb{R}_+, \quad n \in \mathbb{N}^*. \quad (7.1.3)$$

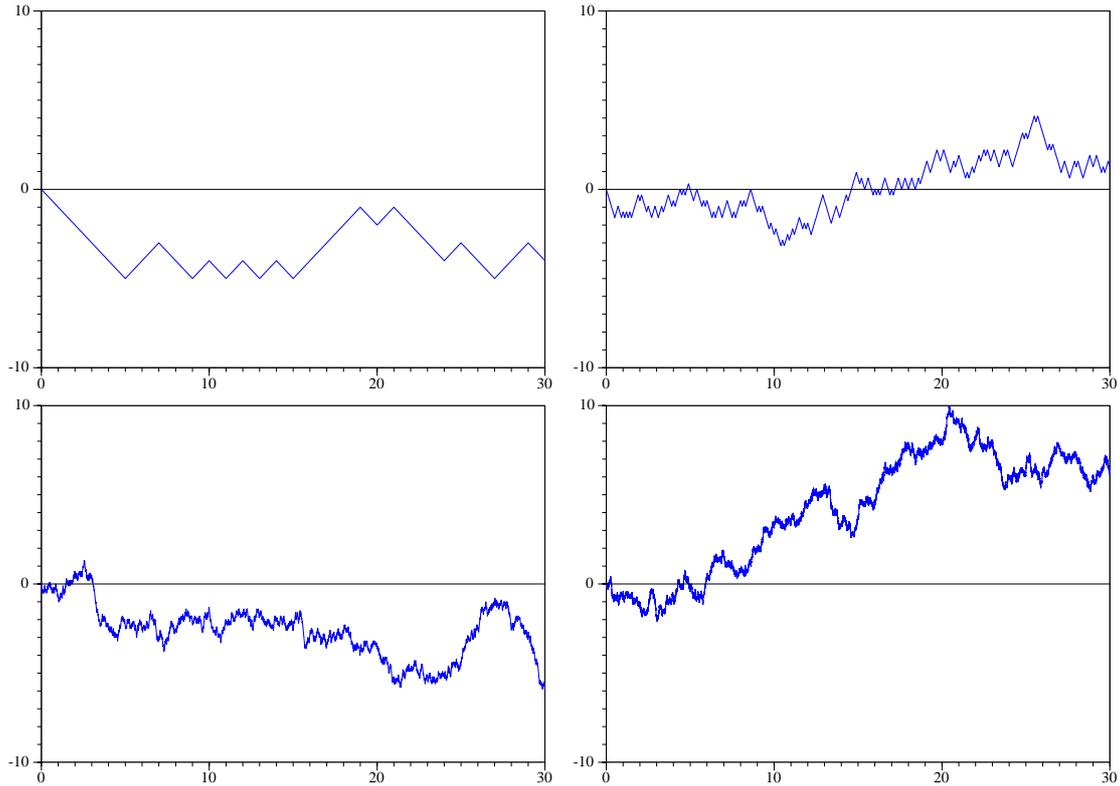


FIGURE 7.1. Les versions rééchelonnées  $B_t^{(1)}$ ,  $B_t^{(10)}$ ,  $B_t^{(100)}$  et  $B_t^{(1000)}$  d'une même réalisation  $\omega$  d'une marche aléatoire.

Cela signifie que l'on accélère le temps d'un facteur  $n$ , tout en comprimant l'espace d'un facteur  $\sqrt{n}$ , de sorte que  $B_t^{(n)}$  effectue des pas de  $\pm 1/\sqrt{n}$  sur des intervalles de temps de longueur  $1/n$  (Figure 7.1).

Soit  $B_t$  le processus obtenu en prenant la limite de  $B_t^{(n)}$  lorsque  $n \rightarrow \infty$ , au sens des distributions finies. Autrement dit,  $B_t$  est défini par le fait que pour toute partition  $0 \leq t_1 < t_2 < \dots < t_k = t$  de  $[0, t]$  et tout  $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ ,

$$\mathbb{P}\{B_{t_1} \leq x_1, \dots, B_{t_k} \leq x_k\} = \lim_{n \rightarrow \infty} \mathbb{P}\{B_{t_1}^{(n)} \leq x_1, \dots, B_{t_k}^{(n)} \leq x_k\}. \quad (7.1.4)$$

Supposons pour l'instant que cette limite existe. Le processus  $B_t$  aura les propriétés suivantes:

1.  $\mathbb{E}(B_t) = 0$  pour tout  $t \geq 0$ ;
2. La variance de  $B_t$  satisfait

$$\text{Var}(B_t) = \lim_{n \rightarrow \infty} \left( \frac{1}{\sqrt{n}} \right)^2 [nt] = t. \quad (7.1.5)$$

3. Par le théorème de la limite centrale (ou la formule de Moivre–Laplace appliquée à (7.1.2)),  $X_{[nt]}/\sqrt{[nt]}$  converge en loi vers une variable normale centrée réduite. Par conséquent,  $B_t$  suit une loi normale  $\mathcal{N}(0, t)$ .
4. *Propriétés des incréments indépendants*: pour tout  $t > s \geq 0$ ,  $B_t - B_s$  est indépendant de  $\{B_u\}_{0 \leq u \leq s}$ ;
5. *Propriétés des incréments stationnaires*: pour tout  $t > s \geq 0$ ,  $B_t - B_s$  a la même loi que  $B_{t-s}$ .

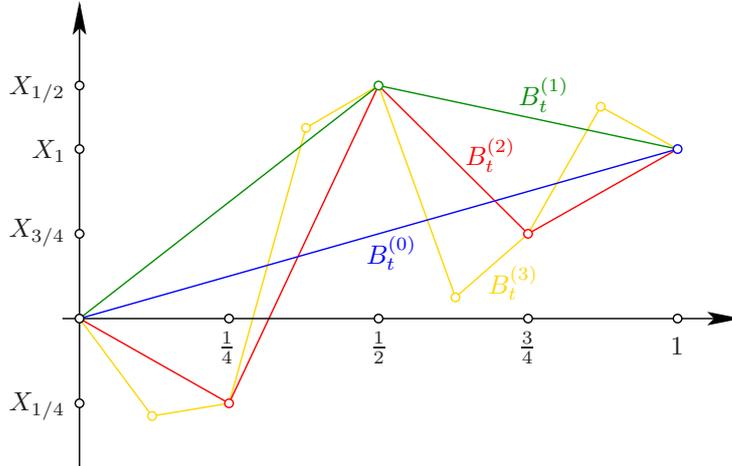


FIGURE 7.2. Construction du mouvement Brownien par interpolation.

## 7.2 Construction du mouvement Brownien

La construction heuristique ci-dessus motive la définition suivante.

**Définition 7.2.1.** *Le mouvement Brownien standard ou processus de Wiener standard est le processus stochastique  $\{B_t\}_{t \geq 0}$  satisfaisant :*

1.  $B_0 = 0$ ;
2. *Incréments indépendants: pour tout  $t > s \geq 0$ ,  $B_t - B_s$  est indépendant de  $\{B_u\}_{u \leq s}$ ;*
3. *Incréments gaussiens: pour tout  $t > s \geq 0$ ,  $B_t - B_s$  suit une loi normale  $\mathcal{N}(0, t - s)$ .*

Nous allons maintenant donner une démonstration de l'existence de ce processus. De plus, nous montrons qu'il admet une version continue, c'est-à-dire que ses trajectoires sont presque sûrement continues, et qu'on peut donc admettre qu'elles sont toutes continues quitte à modifier le processus sur un ensemble de mesure nulle.

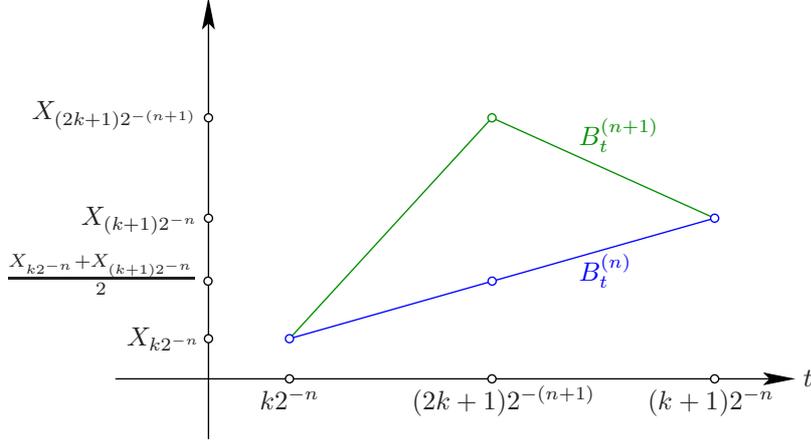
**Théorème 7.2.2.** *Il existe un processus stochastique  $\{B_t\}_{t \geq 0}$  satisfaisant la définition 7.2.1, et dont les trajectoires  $t \mapsto B_t(\omega)$  sont continues.*

DÉMONSTRATION.

1. Nous allons d'abord construire  $\{B_t\}_{0 \leq t \leq 1}$  à partir d'une collection de variables aléatoires gaussiennes indépendantes  $V_1, V_{1/2}, V_{1/4}, V_{3/4}, V_{1/8}, \dots$ , toutes centrées, et avec  $V_1$  et  $V_{1/2}$  de variance 1 et  $V_{k2^{-n}}$  de variance  $2^{-(n-1)}$  ( $k < 2^n$  impair).

Montrons d'abord que si  $X_s$  et  $X_t$  sont deux variables aléatoires telles que  $X_t - X_s$  soit gaussienne centrée de variance  $t - s$ , alors il existe une variable aléatoire  $X_{(t+s)/2}$  telle que les variables  $X_t - X_{(t+s)/2}$  et  $X_{(t+s)/2} - X_s$  soient i.i.d. de loi  $\mathcal{N}(0, (t-s)/2)$ . Si  $U = X_t - X_s$  et  $V$  est indépendante de  $U$ , de même distribution, il suffit de définir  $X_{(t+s)/2}$  par

$$\begin{aligned} X_t - X_{(t+s)/2} &= \frac{U + V}{2} \\ X_{(t+s)/2} - X_s &= \frac{U - V}{2}. \end{aligned} \tag{7.2.1}$$

FIGURE 7.3. Calcul de  $\Delta^{(n)}$ .

En effet, il est aisé de vérifier que ces variables ont la distribution souhaitée, et qu'elles sont indépendantes, puisque  $\mathbb{E}((U+V)(U-V)) = \mathbb{E}(U^2) - \mathbb{E}(V^2) = 0$ , et que des variables aléatoires normales sont indépendantes si et seulement si elles sont non corrélées. Posons alors  $X_0 = 0$ ,  $X_1 = V_1$ , et construisons  $X_{1/2}$  à l'aide de la procédure ci-dessus, avec  $V = V_{1/2}$ . Puis nous construisons  $X_{1/4}$  à l'aide de  $X_0$ ,  $X_{1/2}$  et  $V_{1/4}$ , et ainsi de suite, pour obtenir une familles de variables  $\{X_t\}_{t=k2^{-n}, n \geq 1, k < 2^n}$  telles que pour  $t > s$ ,  $X_t - X_s$  soit indépendante de  $X_s$  et de loi  $\mathcal{N}(0, t - s)$ .

2. Pour  $n \geq 0$ , soit  $\{B_t^{(n)}\}_{0 \leq t \leq 1}$  le processus stochastique à trajectoires linéaires par morceaux sur les intervalles  $[k2^{-n}, (k+1)2^{-n}]$ ,  $k < 2^n$ , et tel que  $B_{k2^{-n}}^{(n)} = X_{k2^{-n}}$  (Figure 7.2). Nous voulons montrer que la suite des  $B^{(n)}(\omega)$  converge uniformément sur  $[0, 1]$  pour toute réalisation  $\omega$  des  $V_i$ . Il nous faut donc estimer

$$\begin{aligned} \Delta^{(n)}(\omega) &= \sup_{0 \leq t \leq 1} |B_t^{(n+1)}(\omega) - B_t^{(n)}(\omega)| \\ &= \max_{0 \leq k \leq 2^{n-1}} \max_{k2^{-n} \leq t \leq (k+1)2^{-n}} |B_t^{(n+1)}(\omega) - B_t^{(n)}(\omega)| \\ &= \max_{0 \leq k \leq 2^{n-1}} \left| X_{(2k+1)2^{-(n+1)}}(\omega) - \frac{1}{2}(X_{k2^{-n}}(\omega) + X_{(k+1)2^{-n}}(\omega)) \right| \end{aligned} \quad (7.2.2)$$

(voir Figure 7.3). Le terme en valeur absolue vaut  $\frac{1}{2}V_{(2k+1)2^{-(n+1)}}$  par construction, c.f. (7.2.1), qui est gaussienne de variance  $2^{-n}$ . Il suit que

$$\begin{aligned} \mathbb{P}\{\Delta^{(n)} > \sqrt{n2^{-n}}\} &= \mathbb{P}\left\{ \max_{0 \leq k \leq 2^{n-1}} |V_{(2k+1)2^{-(n+1)}}| \geq 2\sqrt{n2^{-n}} \right\} \\ &\leq 2 \cdot 2^n \int_{2\sqrt{n2^{-n}}}^{\infty} e^{-x^2/2 \cdot 2^{-n}} \frac{dx}{\sqrt{2\pi}2^{-n}} \\ &= 2 \cdot 2^n \int_{2\sqrt{n}}^{\infty} e^{-y^2/2} \frac{dx}{\sqrt{2\pi}} \leq \text{const } 2^n e^{-2n}, \end{aligned} \quad (7.2.3)$$

et donc

$$\sum_{n \geq 0} \mathbb{P}\{\Delta^{(n)} > \sqrt{n2^{-n}}\} \leq \text{const} \sum_{n \geq 0} (2e^{-2})^n < \infty. \quad (7.2.4)$$

Le lemme de Borel–Cantelli nous permet de conclure qu'avec probabilité 1, il n'existe

qu'un nombre fini de  $n$  pour lesquels  $\Delta^{(n)} > \sqrt{n2^{-n}}$ . Par conséquent,

$$\mathbb{P}\left\{\sum_{n \geq 0} \Delta^{(n)} < \infty\right\} = 1. \quad (7.2.5)$$

La suite des  $\{B_t^{(n)}\}_{0 \leq t \leq 1}$  est donc une suite de Cauchy pour la norme sup avec probabilité 1, et alors elle converge uniformément. Nous posons pour  $t \in [0, 1]$

$$B_t^0 = \begin{cases} \lim_{n \rightarrow \infty} B_t^{(n)} & \text{si la suite converge uniformément} \\ 0 & \text{sinon (avec probabilité 0).} \end{cases} \quad (7.2.6)$$

Il est facile de vérifier que  $B^0$  satisfait les trois propriétés de la définition.

3. Pour étendre le processus à des temps quelconques, nous fabriquons des copies indépendantes  $\{B^i\}_{i \geq 0}$  et posons

$$B_t = \begin{cases} B_t^0 & 0 \leq t < 1 \\ B_1^0 + B_{t-1}^1 & 1 \leq t < 2 \\ B_1^0 + B_1^1 + B_{t-2}^2 & 2 \leq t < 3 \\ \dots & \dots \end{cases} \quad (7.2.7)$$

Ceci conclut la démonstration. □

### 7.3 Propriétés de base

Les propriétés suivantes sont des conséquences directes de la définition 7.2.1 et nous les donnons sans démonstration détaillée.

1. *Propriété de Markov*: Pour tout borelien  $A \in \mathbb{R}$ ,

$$\mathbb{P}\{B_{t+s} \in A \mid B_t = x\} = \int_A p(y, t+s \mid x, t) dy, \quad (7.3.1)$$

indépendamment de  $\{B_u\}_{u < t}$ , avec des probabilités de transition gaussiennes

$$p(y, t+s \mid x, t) = \frac{e^{-(y-x)^2/2s}}{\sqrt{2\pi s}}. \quad (7.3.2)$$

La preuve suit directement du fait que l'on peut décomposer  $B_{t+s} = B_t + (B_{t+s} - B_t)$ , le deuxième terme étant indépendant du premier et de loi  $\mathcal{N}(0, s)$ . On vérifiera en particulier l'équation de Chapman–Kolmogorov: Pour  $t > u > s$ ,

$$p(y, t \mid x, s) = \int_{\mathbb{R}} p(y, t \mid z, u) p(z, u \mid x, s) du. \quad (7.3.3)$$

2. *Propriété différentielle*: Pour tout  $t \geq 0$ ,  $\{B_{t+s} - B_t\}_{s \geq 0}$  est un mouvement Brownien standard, indépendant de  $\{B_u\}_{u < t}$ .
3. *Propriété d'échelle*: Pour tout  $c > 0$ ,  $\{cB_{t/c^2}\}_{s \geq 0}$  est un mouvement Brownien standard.
4. *Symétrie*:  $\{-B_t\}_{t \geq 0}$  est un mouvement Brownien standard.

5. *Processus Gaussien*: Le processus de Wiener est Gaussien de moyenne nulle (c'est-à-dire que ses distributions jointes finies sont normales centrées), et il est caractérisé par sa covariance

$$\text{cov}\{B_t, B_s\} \equiv \mathbb{E}(B_t B_s) = s \wedge t \quad (7.3.4)$$

( $s \wedge t$  désigne le minimum de  $s$  et  $t$ ).

DÉMONSTRATION. Pour  $s < t$ , nous avons

$$\mathbb{E}(B_t B_s) = \mathbb{E}(B_s(B_s + B_t - B_s)) = \mathbb{E}(B_s^2) + \mathbb{E}(B_s(B_t - B_s)) = s, \quad (7.3.5)$$

puisque le deuxième terme est nul par la propriété d'incrémentés indépendants.  $\square$

En fait, un processus Gaussien centré, dont la covariance satisfait (7.3.4) est un processus de Wiener standard.

6. *Renversement du temps*: Le processus  $\{X_t\}_{t \geq 0}$  défini par

$$X_t = \begin{cases} 0 & \text{si } t = 0 \\ tB_{1/t} & \text{si } t > 0 \end{cases} \quad (7.3.6)$$

est un processus de Wiener standard.

DÉMONSTRATION.  $X_t$  est un processus Gaussien de moyenne nulle, et un calcul simple montre que sa covariance est bien  $\text{cov}\{X_t, X_s\} = s \wedge t$ . Le seul point non trivial est de montrer la continuité en zéro de  $X_t$ . Celle-ci est équivalente à la loi forte des grands nombres.  $\square$

Voici enfin une propriété importante de non-régularité des trajectoires du mouvement Brownien.

**Théorème 7.3.1.** *Les trajectoires  $t \mapsto B_t(\omega)$  sont presque sûrement nulle part lipschitziennes, donc aussi nulle part différentiables.*

DÉMONSTRATION. Fixons un  $C < \infty$  et introduisons, pour  $n \geq 1$ ,

$$A_n = \left\{ \omega : \exists s \in [0, 1] \text{ t.q. } |B_t(\omega) - B_s(\omega)| \leq C|t - s| \text{ si } |t - s| \leq \frac{3}{n} \right\}.$$

Il s'agit de montrer que  $\mathbb{P}(A_n) = 0$  pour tout  $n$ . Observons que si  $n$  augmente, la condition s'affaiblit, donc  $A_n \subset A_{n+1}$ . Soit encore, pour  $n \geq 3$  et  $1 \leq k \leq n-2$ ,

$$Y_{k,n}(\omega) = \max_{j=0,1,2} \left\{ \left| B_{(k+j)/n}(\omega) - B_{(k+j-1)/n}(\omega) \right| \right\},$$

$$B_n = \bigcup_{k=1}^{n-2} \left\{ \omega : Y_{k,n}(\omega) \leq \frac{5C}{n} \right\}.$$

L'inégalité triangulaire implique  $A_n \subset B_n$ . En effet, soit  $\omega \in A_n$ . Si par exemple  $s = 1$ , alors pour  $k = n-2$

$$\left| B_{(n-3)/n}(\omega) - B_{(n-2)/n}(\omega) \right| \leq \left| B_{(n-3)/n}(\omega) - B_1(\omega) \right| + \left| B_1(\omega) - B_{(n-2)/n}(\omega) \right| \leq C \left( \frac{3}{n} + \frac{2}{n} \right)$$

donc  $\omega \in B_n$ . Il suit des propriétés des incrémentés indépendants et d'échelle que

$$\mathbb{P}(A_n) \leq \mathbb{P}(B_n) \leq n \mathbb{P} \left( |B_{1/n}| \leq \frac{5C}{n} \right)^3 = n \mathbb{P} \left( |B_1| \leq \frac{5C}{\sqrt{n}} \right)^3 \leq n \left( \frac{10C}{\sqrt{2\pi n}} \right)^3.$$

Il suit que  $\mathbb{P}(A_n) \rightarrow 0$  pour  $n \rightarrow \infty$ . Mais comme  $\mathbb{P}(A_n) \leq \mathbb{P}(A_{n+1})$  pour tout  $n$ , ceci implique  $\mathbb{P}(A_n) = 0$  pour tout  $n$ .  $\square$

## 7.4 Temps d'arrêt

Soit  $\{B_t\}_{t \geq 0}$  un mouvement Brownien standard. Nous allons laisser de côté les détails techniques de la construction de l'espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  sur lequel  $B_t$  est défini (la construction utilise le théorème d'extension de Kolmogorov). Précisons simplement que l'univers  $\Omega$  peut être identifié à l'ensemble des fonctions continues  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  telles que  $f(0) = 0$ . Introduisons pour tout  $t \geq 0$  la tribu

$$\mathcal{F}_t = \sigma(\{B_s\}_{0 \leq s \leq t}), \quad (7.4.1)$$

contenant les événements ne dépendant que du mouvement Brownien jusqu'au temps  $t$ . La suite  $\{\mathcal{F}_t\}_{t \geq 0}$  est une suite croissante de sous-tribus de  $\mathcal{F}$ , c'est la *filtration canonique* engendrée par le mouvement Brownien.<sup>1</sup>

**Définition 7.4.1** (Temps d'arrêt). *On appelle temps d'arrêt une variable aléatoire  $\tau : \Omega \rightarrow [0, \infty]$  telle que  $\{\tau < t\} \in \mathcal{F}_t$  pour tout  $t \geq 0$ .*

Remarquons que si  $\{\tau \leq t\} \in \mathcal{F}_t$ , alors

$$\{\tau < t\} = \bigcup_{n \geq 1} \left\{ \tau \leq t - \frac{1}{n} \right\} \in \mathcal{F}_t. \quad (7.4.2)$$

La continuité à droite de la filtration montre en plus que si  $\{\tau < t\} \in \mathcal{F}_t$ , alors

$$\{\tau \leq t\} = \bigcap_{n \geq 1} \left\{ \tau < t + \frac{1}{n} \right\} \in \mathcal{F}_t. \quad (7.4.3)$$

On peut donc remplacer sans autres  $\tau < t$  par  $\tau \leq t$  dans la définition des temps d'arrêt (ce qui n'est pas le cas pour les processus en temps discret).

Commençons par énoncer quelques propriétés de base permettant de construire des exemples de temps d'arrêt.

### Proposition 7.4.2.

1. Soit  $A$  un ouvert. Alors le temps de premier passage dans  $A$ ,  $\tau_A = \inf\{t \geq 0 : B_t \in A\}$  est un temps d'arrêt.
2. Si  $\tau_n$  est une suite décroissante de temps d'arrêt telle que  $\tau_n \rightarrow \tau$ , alors  $\tau$  est un temps d'arrêt.
3. Si  $\tau_n$  est une suite croissante de temps d'arrêt telle que  $\tau_n \rightarrow \tau$ , alors  $\tau$  est un temps d'arrêt.
4. Soit  $K$  un fermé. Alors le temps de premier passage dans  $K$  est un temps d'arrêt.

DÉMONSTRATION.

1. Par continuité de  $B_t$ ,  $\{\tau_A < t\} = \bigcup_{q \in \mathbb{Q}, q < t} \{B_q \in A\} \in \mathcal{F}_t$ .
2. On a  $\{\tau < t\} = \bigcup_{n \geq 1} \{\tau_n < t\} \in \mathcal{F}_t$ .
3. On a  $\{\tau \leq t\} = \bigcap_{n \geq 1} \{\tau_n \leq t\} \in \mathcal{F}_t$ .

<sup>1</sup>Pour être précis, on préfère considérer  $\mathcal{F}'_t = \bigcap_{s > t} \mathcal{F}_s$ , qui est continue à droite :  $\mathcal{F}'_t = \bigcap_{s > t} \mathcal{F}'_s$ . Pour des raisons techniques, on complète en général cette tribu par les ensembles de mesure nulle (c.f. [Dur96, Section 7.2]).

4. Soit  $A_n = \bigcup_{x \in K} (x - 1/n, x + 1/n)$  et  $\tau_n$  le temps de premier passage dans  $A_n$ . Comme  $A_n$  est ouvert,  $\tau_n$  est un temps d'arrêt par 1. Les  $A_n$  étant décroissants, les  $\tau_n$  sont croissants. Soit  $\sigma = \lim_{n \rightarrow \infty} \tau_n \in [0, \infty]$ . Comme  $\tau_K \geq \tau_n$  pour tout  $n$ , on a  $\sigma \leq \tau_K$ . Si  $\sigma = \infty$ , alors  $\tau_K = \infty = \sigma$ . Supposons que  $\sigma < \infty$ . Comme  $B_{\tau_n} \in \bar{A}_n$  et  $B_{\tau_n} \rightarrow B_\sigma$ , il suit que  $B_\sigma \in K$  et donc  $\tau_K \leq \sigma$ . Ainsi  $\tau_K = \sigma$  et le résultat suit de 3.  $\square$

Comme dans le cas à temps discret, nous introduisons

**Définition 7.4.3** (Tribu de événements antérieurs). *Soit  $\tau$  un temps d'arrêt. Alors la tribu*

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \forall t \geq 0\} \quad (7.4.4)$$

*est appelée la tribu des événements antérieurs à  $\tau$ .*

**Proposition 7.4.4.**

1. Si  $\sigma$  et  $\tau$  sont des temps d'arrêt tels que  $\sigma \leq \tau$  alors  $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$ .
2. Si  $\tau_n \rightarrow \tau$  est une suite décroissante de temps d'arrêt, alors  $\mathcal{F}_\tau = \bigcap_n \mathcal{F}_{\tau_n}$ .

DÉMONSTRATION.

1. Voir la preuve de la Proposition 5.1.5.
2. D'une part par 1., on a  $\mathcal{F}_{\tau_n} \supset \mathcal{F}_\tau$  pour tout  $n$ . D'autre part, soit  $A \in \bigcap_n \mathcal{F}_{\tau_n}$ . Comme  $A \cap \{\tau_n < t\} \in \mathcal{F}_t$  et  $\tau_n$  décroît vers  $\tau$ , on a  $A \cap \{\tau < t\} \in \mathcal{F}_t$ , donc  $\bigcap_n \mathcal{F}_{\tau_n} \subset \mathcal{F}_\tau$ .  $\square$

L'une des propriétés les plus importantes du mouvement Brownien, liée aux temps d'arrêt, est la propriété de Markov forte.

**Théorème 7.4.5** (Propriété de Markov forte du mouvement Brownien). *Si  $\tau$  est un temps d'arrêt tel que  $\mathbb{P}\{\tau < \infty\} = 1$ , alors  $\{B_{\tau+t} - B_\tau\}_{t \geq 0}$  est un mouvement Brownien standard, indépendant de  $\mathcal{F}_\tau$ .*

DÉMONSTRATION. Voir [Dur96, Section 7.3] ou [McK69], p. 10.  $\square$

**Corollaire 7.4.6** (Principe de réflexion). *Pour  $L > 0$  et  $\tau = \inf\{t \geq 0 : B_t \geq L\}$ , le processus*

$$B_t^* = \begin{cases} B_t & \text{si } t \leq \tau \\ 2L - B_t & \text{si } t > \tau \end{cases} \quad (7.4.5)$$

*est un mouvement Brownien standard.*

Voici une application classique du principe de réflexion.

**Corollaire 7.4.7.** *Pour tout  $L \geq 0$ ,*

$$\mathbb{P}\left\{\sup_{0 \leq s \leq t} B_s \geq L\right\} = 2\mathbb{P}\{B_t \geq L\}. \quad (7.4.6)$$

DÉMONSTRATION. Si  $\tau = \inf\{t \geq 0 : B_t \geq L\}$ , alors  $\{\sup_{0 \leq s \leq t} B_s \geq L\} = \{\tau \leq t\}$ . Les trajectoires du mouvement Brownien étant continues,  $B_t \geq L$  implique  $\tau \leq t$ . Ainsi,

$$\begin{aligned} \mathbb{P}\{B_t \geq L\} &= \mathbb{P}\{B_t \geq L, \tau \leq t\} \\ &= \mathbb{P}\{2L - B_t \leq L, \tau \leq t\} \\ &= \mathbb{P}\{B_t^* \leq L, \tau \leq t\} \\ &= \mathbb{P}\{B_t \leq L, \tau \leq t\}. \end{aligned} \quad (7.4.7)$$

Le résultat suit de la décomposition  $\mathbb{P}\{\tau \leq t\} = \mathbb{P}\{B_t \leq L, \tau \leq t\} + \mathbb{P}\{B_t > L, \tau \leq t\}$ .  $\square$

On remarquera que lorsque  $t \rightarrow \infty$ , l'expression (7.4.6) tend vers 1, donc le mouvement Brownien atteint presque sûrement tout niveau  $L$ .

## 7.5 Mouvement Brownien et martingales

Le mouvement Brownien, ainsi que toute une série de processus dérivés, sont des martingales. Les martingales, sur- et sous-martingales sont définies comme dans le cas discret, sauf qu'on considère tous les temps  $t > s$ , pour lesquels  $\mathcal{F}_t \supset \mathcal{F}_s$ . Commençons par considérer le mouvement Brownien.

**Théorème 7.5.1** (Propriété de martingale du mouvement Brownien). *Le mouvement Brownien est une martingale par rapport à la filtration canonique  $\{\mathcal{F}_t\}_{t \geq 0}$ .*

DÉMONSTRATION. Pour tout  $t > s \geq 0$ , on a

$$\mathbb{E}(B_t | \mathcal{F}_s) = \mathbb{E}(B_t - B_s | \mathcal{F}_s) + \mathbb{E}(B_s | \mathcal{F}_s) = \mathbb{E}(B_{t-s}) + B_s = B_s,$$

en vertu de la propriété différentielle et de la propriété des incréments indépendants.  $\square$

L'inégalité de Jensen implique que  $B_t^2$  et  $e^{\gamma B_t}$  pour  $\gamma > 0$  sont des sous-martingales. Le résultat suivant montre qu'en les modifiant de manière déterministe, on obtient des martingales :

**Proposition 7.5.2.**

1.  $B_t^2 - t$  est une martingale.
2. Pour tout  $\gamma \in \mathbb{R}$ ,  $\exp(\gamma B_t - \gamma^2 t / 2)$  est une martingale.

DÉMONSTRATION.

1.  $\mathbb{E}(B_t^2 | \mathcal{F}_s) = \mathbb{E}(B_s^2 + 2B_s(B_t - B_s) + (B_t - B_s)^2 | \mathcal{F}_s) = B_s^2 + 0 + (t - s)$ .
2.  $\mathbb{E}(e^{\gamma B_t} | \mathcal{F}_s) = e^{\gamma B_s} \mathbb{E}(e^{\gamma(B_t - B_s)} | \mathcal{F}_s) = e^{\gamma B_s} \mathbb{E}(e^{\gamma B_{t-s}})$ , et

$$\mathbb{E}(e^{\gamma B_{t-s}}) = \int_{-\infty}^{\infty} e^{\gamma x} \frac{e^{-x^2/2(t-s)}}{\sqrt{2\pi(t-s)}} dx = e^{\gamma^2(t-s)/2}$$

par complétion du carré.  $\square$

On peut donc écrire

$$B_t^2 = t + M_t, \quad M_t \text{ martingale}, \quad (7.5.1)$$

ce qui est l'analogie de la décomposition de Doob du cas en temps discret. Le processus croissant associé au mouvement Brownien est donc donné par

$$\langle B \rangle_t = t. \quad (7.5.2)$$

Ceci peut être vu intuitivement comme conséquence du fait que

$$\mathbb{E}((B_t - B_s)^2 | \mathcal{F}_s) = \mathbb{E}((B_t - B_s)^2) = \mathbb{E}(B_{t-s}^2) = t - s, \quad (7.5.3)$$

et donc que l'accroissement infinitésimal de  $B_t^2$  est en moyenne  $\mathbb{E}((B_{s+ds} - B_s)^2 | \mathcal{F}_s) = ds$ .

Nous donnons deux types d'applications de la propriété de martingale. Un premier type d'applications suit du fait qu'une martingale arrêtée est encore une martingale (c.f. Section 5.2).

**Proposition 7.5.3.** *Soit  $X_t$  une martingale continue à droite, adaptée à une filtration continue à droite. Si  $\tau$  est un temps d'arrêt borné, alors  $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$ .*

DÉMONSTRATION. Soit  $n$  un entier tel que  $\mathbb{P}\{\tau \leq n-1\} = 1$ , et  $\tau_m = (\lfloor 2^m \tau \rfloor + 1)/2^m$ . Le processus discrétisé  $\{Y_k^m\}_{k \geq 1} = X_{k2^{-m}}$  est une martingale pour la filtration  $\mathcal{F}_k^m = \mathcal{F}_{k2^{-m}}$ , et  $\sigma_m = 2^m \tau_m$  est un temps d'arrêt pour cette filtration. Il suit que

$$X_{\tau_m} = Y_{\sigma_m}^m = \mathbb{E}(Y_{n2^m}^m | \mathcal{F}_{\sigma_m}^m) = \mathbb{E}(X_n | \mathcal{F}_{\tau_m}). \quad (7.5.4)$$

Lorsque  $m \rightarrow \infty$ ,  $X_{\tau_m}$  tend vers  $X_\tau$  par continuité à droite, et  $\mathcal{F}_{\tau_m} \searrow \mathcal{F}_\tau$  par la Proposition 7.4.4. Ceci implique que  $X_\tau = \mathbb{E}(X_n | \mathcal{F}_\tau)$ . Par la propriété de martingale, il suit que  $\mathbb{E}(X_\tau) = \mathbb{E}(X_n) = \mathbb{E}(X_0)$ .  $\square$

Voici quelques applications de ce résultat.

**Corollaire 7.5.4.**

1. *Soit  $\tau_a = \inf\{t \geq 0 : x + B_t = a\}$ . Alors pour  $a < x < b$  on a*

$$\mathbb{P}\{\tau_a < \tau_b\} = \frac{b-x}{b-a}. \quad (7.5.5)$$

2. *Soit  $a < 0 < b$ , et  $\tau = \inf\{t \geq 0 : B_t \notin (a, b)\}$ . Alors*

$$\mathbb{E}(\tau) = |ab|. \quad (7.5.6)$$

3. *Soit  $a > 0$  et  $\tau = \inf\{t \geq 0 : B_t \notin (-a, a)\}$ . Alors pour  $\lambda > 0$ ,*

$$\mathbb{E}(e^{-\lambda\tau}) = \frac{1}{\cosh(a\sqrt{2\lambda})}. \quad (7.5.7)$$

DÉMONSTRATION.

1. Soit  $\tau = \tau_a \wedge \tau_b$ . Alors  $\mathbb{E}(x + B_{\tau \wedge t}) = x$ . Le Corollaire 7.4.7 montre que  $\tau_a$  et  $\tau_b$  sont finis presque sûrement, donc aussi  $\tau$ . Faisant tendre  $t$  vers l'infini, on obtient

$$x = \mathbb{E}(x + B_\tau) = a\mathbb{P}\{\tau_a < \tau_b\} + b\mathbb{P}\{\tau_b < \tau_a\},$$

et le résultat suit en résolvant par rapport à  $\mathbb{P}\{\tau_a < \tau_b\}$ .

2. Comme  $B_t^2 - t$  est une martingale, on a  $\mathbb{E}(B_{\tau \wedge t}^2) = \mathbb{E}(\tau \wedge t)$ . Le théorème de convergence dominée et (7.5.5) permettent d'écrire

$$\mathbb{E}(\tau) = \lim_{t \rightarrow \infty} \mathbb{E}(\tau \wedge t) = \lim_{t \rightarrow \infty} \mathbb{E}(B_{\tau \wedge t}^2) = \mathbb{E}(B_\tau^2) = a^2 \frac{b}{b-a} + b^2 \left(1 - \frac{b}{b-a}\right) = -ab.$$

3. Comme  $e^{\gamma B_t - \gamma^2 t/2}$  est une martingale, on a  $\mathbb{E}(e^{\gamma B_{\tau \wedge t} - \gamma^2 (\tau \wedge t)/2}) = 1$ . Faisant tendre  $t$  vers l'infini et invoquant le théorème de convergence dominée, on obtient

$$\mathbb{E}(e^{\gamma B_\tau - \gamma^2 \tau/2}) = 1.$$

Par symétrie, on a  $\mathbb{P}\{B_\tau = -a\} = \mathbb{P}\{B_\tau = a\} = 1/2$  et  $B_\tau$  est indépendant de  $\tau$ . Par conséquent,

$$\mathbb{E}(e^{\gamma B_\tau - \gamma^2 \tau/2}) = \cosh(\gamma a) \mathbb{E}(e^{-\gamma^2 \tau/2}),$$

et le résultat suit en prenant  $\gamma = \sqrt{2\lambda}$ .  $\square$

Remarquons que  $\mathbb{E}(e^{-\lambda\tau})$  est la transformée de Laplace de la densité de  $\tau$ , donc l'expression (7.5.7) permet de trouver cette densité par transformée de Laplace inverse.

Le second type d'applications que nous considérons suit de l'inégalité de Doob.

**Théorème 7.5.5** (Inégalité de Doob). *Soit  $X_t$  une sous-martingale à trajectoires continues. Alors pour tout  $\lambda > 0$ ,*

$$\mathbb{P}\left\{\sup_{0 \leq s \leq t} X_s \geq \lambda\right\} \leq \frac{1}{\lambda} \mathbb{E}(X_t^+). \quad (7.5.8)$$

DÉMONSTRATION. Voir la preuve du Théorème 5.2.4. □

**Corollaire 7.5.6.** *Pour tout  $\gamma, \lambda > 0$ ,*

$$\mathbb{P}\left\{\sup_{0 \leq s \leq t} \left[B_s - \gamma \frac{s}{2}\right] > \lambda\right\} \leq e^{-\gamma\lambda}. \quad (7.5.9)$$

## 7.6 Exercices

**Exercice 7.1.** Montrer que pour tout  $\gamma \in \mathbb{R}$ , le processus  $X_t = e^{-\gamma^2 t/2} \cosh(\gamma B_t)$  est une martingale.

En déduire une autre preuve du fait que  $\tau = \inf\{t \geq 0 : B_t \notin (-a, a)\}$  satisfait

$$\mathbb{E}(e^{-\lambda\tau}) = \frac{1}{\cosh(a\sqrt{2\lambda})}.$$

**Exercice 7.2.**

1. Montrer que si  $X_t = f(B_t, t, \gamma)$  est une martingale, alors (sous des conditions de régularité qu'on précisera)  $\frac{d}{d\gamma} f(B_t, t, \gamma)$  est également une martingale.
2. Soit  $f(x, t, \gamma) = e^{\gamma x - \gamma^2 t/2}$ . Calculer le développement limité de  $f$  en  $\gamma = 0$  jusqu'à l'ordre  $\gamma^4$ . En déduire deux nouvelles martingales dérivées du mouvement Brownien.
3. Soit  $\tau = \inf\{t \geq 0 : B_t \notin (-a, a)\}$ . Calculer  $\mathbb{E}(\tau^2)$  à l'aide du résultat précédent.

**Exercice 7.3.** Pour  $a, b > 0$ , on pose  $X_t = B_t - bt$  et  $\tau = \inf\{t \geq 0 : X_t = a\}$ .

1. En utilisant la martingale  $e^{\gamma B_t - \gamma^2 t/2}$ , où  $\gamma$  est la solution positive de  $\gamma^2 - 2b\gamma - 2\lambda = 0$ , calculer

$$\mathbb{E}(e^{-\lambda\tau} 1_{\{\tau < \infty\}}).$$

2. Particulariser au cas  $b = 0$ .
3. Soit  $b > 0$ . En choisissant une valeur convenable de  $\lambda$ , déterminer  $\mathbb{P}\{\tau < \infty\}$ .



# Chapitre 8

## L'intégrale d'Itô

Le but de l'intégrale d'Itô est de donner un sens à des équations de la forme

$$\frac{dX}{dt} = f(X) + g(X) \frac{dB_t}{dt} . \quad (8.0.1)$$

Par exemple, si  $f \equiv 0$  et  $g \equiv 1$ , on devrait retrouver  $X_t = X_0 + B_t$ , décrivant le mouvement suramorti d'une particule Brownienne.

Le problème est que, comme nous l'avons mentionné, les trajectoires du processus de Wiener ne sont pas différentiables, ni même à variations bornées.

Comme dans le cas des équations différentielles ordinaires, on interprète une solution de l'équation différentielle (8.0.1) comme une solution de l'équation intégrale

$$X_t = X_0 + \int_0^t f(X_s) ds + \int_0^t g(X_s) dB_s . \quad (8.0.2)$$

C'est à la deuxième intégrale qu'il s'agit de donner un sens mathématique. Si  $s \mapsto g(X_s)$  était différentiable, on pourrait le faire à l'aide d'une intégration par parties, mais ce n'est en général pas le cas. Itô a donné une autre définition de l'intégrale stochastique, qui s'applique à une classe beaucoup plus vaste d'intégrands (et donne le même résultat que l'intégration par parties dans le cas différentiable).

### 8.1 Définition

Notre but est de définir l'intégrale stochastique

$$\int_0^t X_s dB_s \quad (8.1.1)$$

simultanément pour tous les  $t \in [0, T]$ , où  $X_t$  est lui-même un processus stochastique. Plus précisément, nous supposons que  $X_t$  est une *fonctionnelle Brownienne non-anticipative*, c'est-à-dire ( $\{\mathcal{F}_t\}_{t \geq 0}$  désignant la filtration canonique engendrée par  $\{B_t\}_{t \geq 0}$ )

1.  $X$  est mesurable par rapport à  $\mathcal{F}$ ;
2.  $X_t$  est adapté à  $\mathcal{F}_t$ , c'est-à-dire mesurable par rapport à  $\mathcal{F}_t$  pour tout  $t \in [0, T]$ .

Ceci revient à exiger que  $X_t$  ne dépende que de l'histoire du processus de Wiener jusqu'au temps  $t$ , ce qui est raisonnable au vu de (8.0.2). En outre, nous allons supposer que

$$\mathbb{P} \left\{ \int_0^T X_t^2 dt < \infty \right\} = 1 . \quad (8.1.2)$$

**Remarque 8.1.1.** On peut admettre que  $X_t$  dépende de variables aléatoires supplémentaires, indépendantes de  $B_t$ ; par exemple, la condition initiale peut être aléatoire. Il convient alors d'étendre les tribus  $\mathcal{F}$  et  $\mathcal{F}_t$  dans la définition ci-dessus à des tribus plus grandes  $\mathcal{A}$  et  $\mathcal{A}_t$ , où  $\mathcal{A}_t$  ne doit pas dépendre de la tribu engendrée par  $\{B_{t+s} - B_t\}_{s \geq 0}$ .

Dans un premier temps, nous allons définir l'intégrale stochastique pour un intégrant simple.

**Définition 8.1.2.** Une fonctionnelle Brownienne non-anticipative  $\{e_t\}_{t \in [0, T]}$  est dite simple ou élémentaire s'il existe une partition  $0 = t_0 < t_1 < \dots < t_N = T$  de  $[0, T]$  telle que

$$e_t = \sum_{k=1}^N e_{t_{k-1}} 1_{[t_{k-1}, t_k)}(t). \quad (8.1.3)$$

Pour une telle fonctionnelle, nous définissons l'intégrale stochastique par

$$\int_0^t e_s dB_s = \sum_{k=1}^m e_{t_{k-1}} [B_{t_k} - B_{t_{k-1}}] + e_{t_m} [B_t - B_{t_m}], \quad (8.1.4)$$

où  $m$  est tel que  $t \in [t_m, t_{m+1})$ .

Il est aisé de vérifier les propriétés suivantes:

1. Pour deux fonctionnelles simples  $e^{(1)}$  et  $e^{(2)}$ ,

$$\int_0^t (e_s^{(1)} + e_s^{(2)}) dB_s = \int_0^t e_s^{(1)} dB_s + \int_0^t e_s^{(2)} dB_s. \quad (8.1.5)$$

2. Pour toute constante  $c$ ,

$$\int_0^t (ce_s) dB_s = c \int_0^t e_s dB_s. \quad (8.1.6)$$

3. L'intégrale (8.1.4) est une fonction continue de  $t$ .
4. Si  $\int_0^t \mathbb{E}(|e_s|) ds < \infty$ , alors

$$\mathbb{E}\left(\int_0^t e_s dB_s\right) = 0. \quad (8.1.7)$$

DÉMONSTRATION. Posons  $t_{m+1} = t$ . On a

$$\begin{aligned} \mathbb{E}\left(\int_0^t e_s dB_s\right) &= \mathbb{E}\left(\sum_{k=1}^{m+1} e_{t_{k-1}} (B_{t_k} - B_{t_{k-1}})\right) \\ &= \sum_{k=1}^{m+1} \mathbb{E}(e_{t_{k-1}}) \underbrace{\mathbb{E}(B_{t_k} - B_{t_{k-1}})}_0 = 0, \end{aligned}$$

en vertu des propriétés des incréments indépendants et gaussiens.  $\square$

5. Si  $\int_0^t \mathbb{E}(e_s^2) ds < \infty$ , on a l'isométrie d'Itô

$$\mathbb{E}\left(\left(\int_0^t e_s dB_s\right)^2\right) = \int_0^t \mathbb{E}(e_s^2) ds. \quad (8.1.8)$$

DÉMONSTRATION. Posons  $t_{m+1} = t$ . On a

$$\begin{aligned} \mathbb{E}\left(\left(\int_0^t e_s dB_s\right)^2\right) &= \mathbb{E}\left(\sum_{k,l=1}^{m+1} e_{t_{k-1}} e_{t_{l-1}} (B_{t_k} - B_{t_{k-1}})(B_{t_l} - B_{t_{l-1}})\right) \\ &= \sum_{k=1}^{m+1} \mathbb{E}(e_{t_{k-1}}^2) \underbrace{\mathbb{E}((B_{t_k} - B_{t_{k-1}})^2)}_{t_k - t_{k-1}} \\ &= \int_0^t \mathbb{E}(e_s^2) ds. \end{aligned} \quad (8.1.9)$$

Nous avons utilisé la propriété des incréments indépendants afin d'éliminer les termes  $k \neq l$  de la double somme, et le fait que  $e_s$  est nonanticipative.  $\square$

L'idée d'Itô pour définir l'intégrale stochastique d'une fonctionnelle non-anticipative générale  $X$  est de trouver une suite de fonctionnelles simples  $e^{(n)}$  approchant  $X$  dans  $L^2(\mathbb{P})$ , c'est-à-dire

$$\lim_{n \rightarrow \infty} \int_0^T \mathbb{E}((X_s - e_s^{(n)})^2) ds = 0. \quad (8.1.10)$$

L'isométrie (8.1.8) nous permet alors d'affirmer que la limite suivante existe dans  $L^2(\mathbb{P})$ :

$$\lim_{n \rightarrow \infty} \int_0^t e_s^{(n)} dB_s =: \int_0^t X_s dB_s. \quad (8.1.11)$$

C'est par définition l'intégrale d'Itô de  $X_s$ .

Nous allons maintenant prouver que cette construction est bien possible, indépendante de la suite des  $e^{(n)}$ , et que l'intégrale résultante est une fonction continue de  $t$ . Nous commençons par énoncer deux lemmes préparatoires. Dans ce qui suit, nous utilisons la notation abrégée

$$\mathbb{P}\{A_n, n \rightarrow \infty\} = 1 \quad \Leftrightarrow \quad \mathbb{P}\left\{\sum_{n=1}^{\infty} 1_{A_n^c} < \infty\right\} = 1 \quad \Leftrightarrow \quad \mathbb{P}\left\{\limsup_n A_n\right\} = 0. \quad (8.1.12)$$

Autrement dit, avec probabilité 1, une infinité de  $A_n$  sont réalisés, c'est-à-dire que pour presque tout  $\omega \in \Omega$ , il existe  $n_0(\omega) < \infty$  tel que  $A_n \ni \omega$  pour tout  $n \geq n_0(\omega)$ . Le lemme de Borel–Cantelli affirme que c'est le cas si la somme des  $\mathbb{P}\{A_n^c\}$  converge.

**Lemme 8.1.3.** *Pour toute fonctionnelle non-anticipative  $X$  satisfaisant (8.1.2), il existe une suite  $\{e^{(n)}\}_{n \geq 1}$  de fonctionnelles simples telles que*

$$\mathbb{P}\left\{\int_0^T (X_t - e_t^{(n)})^2 dt \leq 2^{-n}, n \rightarrow \infty\right\} = 1. \quad (8.1.13)$$

DÉMONSTRATION. Considérons les fonctionnelles simples

$$e_t^{(m,k)} = 2^k \int_{(2^{-m} \lfloor 2^m t \rfloor - 2^{-k}) \vee 0}^{2^{-m} \lfloor 2^m t \rfloor} X_s ds. \quad (8.1.14)$$

Faisant tendre d'abord  $m$ , puis  $k$  vers l'infini, on s'aperçoit que  $\int_0^T (X_t - e_t^{(m,k)})^2 dt \rightarrow 0$ . On peut donc trouver des suites  $m_n$  et  $k_n$  telles que  $e^{(n)} = e^{(m_n, k_n)}$  satisfasse

$$\mathbb{P}\left\{\int_0^T (X_t - e_t^{(n)})^2 dt > 2^{-n}\right\} \leq 2^{-n}. \quad (8.1.15)$$

La relation (8.1.13) suit alors du lemme de Borel–Cantelli.  $\square$

**Lemme 8.1.4.** *Pour une suite  $\{f^{(n)}\}_{n \geq 1}$  de fonctionnelles nonanticipatives simples satisfaisant  $\mathbb{P}\{\int_0^T (f_t^{(n)})^2 dt \leq 2^{-n}, n \rightarrow \infty\} = 1$  et tout  $\theta > 1$ ,*

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} \left| \int_0^t f_s^{(n)} dB_s \right| < \theta \sqrt{\frac{\log n}{2^{n-1}}}, n \rightarrow \infty\right\} = 1. \quad (8.1.16)$$

DÉMONSTRATION. Pour tout  $\gamma \in \mathbb{R}$ , le processus stochastique

$$M_t^{(n)} = \exp\left\{\gamma \int_0^t f_s^{(n)} dB_s - \frac{\gamma^2}{2} \int_0^t (f_s^{(n)})^2 ds\right\} \quad (8.1.17)$$

est une martingale par rapport à  $\{B_t\}_{t \geq 0}$ . En effet, nous avons démontré cette propriété pour  $f^{(n)} \equiv 1$  dans la proposition 7.5.2. La même démonstration montre que si  $f^{(n)} \equiv c$  où  $c$  est une variable aléatoire mesurable par rapport à  $\mathcal{F}_s$ , alors  $\mathbb{E}(M_t | \mathcal{F}_s) = M_s$ . Le cas d'un  $f^{(n)}$  simple arbitraire est alors traité par récurrence.

L'inégalité de Doob implique

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} \left( \int_0^t f_s^{(n)} dB_s - \frac{\gamma}{2} \int_0^t (f_s^{(n)})^2 ds \right) > L\right\} \leq e^{-\gamma L}. \quad (8.1.18)$$

Posons alors  $\gamma = \sqrt{2^{n+1} \log n}$  et  $L = \theta \sqrt{2^{-(n+1)} \log n}$ . Utilisant l'hypothèse sur les  $f^{(n)}$ , nous obtenons

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} \left( \int_0^t f_s^{(n)} dB_s \right) > (1 + \theta) \sqrt{2^{-(n+1)} \log n}\right\} \leq e^{-\theta \log n} = n^{-\theta}. \quad (8.1.19)$$

Le lemme de Borel–Cantelli nous permet alors de conclure.  $\square$

**Théorème 8.1.5.** *La limite (8.1.11) existe, est indépendante de la suite des  $\{e^{(n)}\}_{n \geq 1}$  convergeant vers  $X$ , et est une fonction continue de  $t$ .*

DÉMONSTRATION. Quitte à passer à une sous-suite, nous pouvons choisir les  $e^{(n)}$  de manière à satisfaire (8.1.13). Mais ceci implique aussi que

$$\mathbb{P}\left\{\int_0^T (e_t^{(n)} - e_t^{(n-1)})^2 dt \leq \text{const } 2^{-n}, n \rightarrow \infty\right\} = 1. \quad (8.1.20)$$

Le lemme 8.1.4 montre alors que

$$\mathbb{P}\left\{\sup_{0 \leq t \leq T} \left| \int_0^t (e_s^{(n)} - e_s^{(n-1)}) dB_s \right| < \text{const } \theta \sqrt{\frac{\log n}{2^{n-1}}}, n \rightarrow \infty\right\} = 1. \quad (8.1.21)$$

Ainsi la suite des  $\int_0^t e_s^{(n)} dB_s$  est de Cauchy presque sûrement, et dans ce cas elle converge uniformément.  $\square$

## 8.2 Propriétés élémentaires

Les propriétés suivantes sont prouvées aisément pour des processus  $X$  et  $Y$  nonanticipatifs satisfaisant la condition d'intégrabilité (8.1.2).

1. **Linéarité:**

$$\int_0^t (X_s + Y_s) dB_s = \int_0^t X_s dB_s + \int_0^t Y_s dB_s \quad (8.2.1)$$

et

$$\int_0^t (cX_s) dB_s = c \int_0^t X_s dB_s . \quad (8.2.2)$$

2. **Additivité:** Pour  $0 \leq s < u < t \leq T$ ,

$$\int_s^t X_v dB_v = \int_s^u X_v dB_v + \int_u^t X_v dB_v . \quad (8.2.3)$$

3. Pour un temps d'arrêt  $\tau$ ,

$$\int_0^{\tau \wedge T} X_t dB_t = \int_0^T 1_{\{t \leq \tau\}} X_t dB_t . \quad (8.2.4)$$

4. Si  $\int_0^T \mathbb{E}(X_t^2) dt < \infty$ , alors pour tout  $t \leq T$ ,

$$\mathbb{E}\left(\int_0^t X_s dB_s\right) = 0 \quad (8.2.5)$$

et

$$\mathbb{E}\left(\left(\int_0^t X_s dB_s\right)^2\right) = \int_0^t \mathbb{E}(X_s^2) ds . \quad (8.2.6)$$

De plus, le processus  $\{\int_0^t X_s dB_s\}_{t \geq 0}$  est une martingale.

## 5. Le processus

$$M_t = \exp\left\{\int_0^t X_s dB_s - \frac{1}{2} \int_0^t X_s^2 ds\right\} \quad (8.2.7)$$

est une surmartingale.

### 8.3 Un exemple

Nous donnons ici un exemple de calcul explicite d'une intégrale stochastique par la méthode d'Itô:

$$\int_0^t B_s dB_s = \frac{1}{2} B_t^2 - \frac{t}{2} . \quad (8.3.1)$$

Le résultat, quelque peu surprenant au premier abord, prendra tout son sens lorsque nous aurons vu la formule d'Itô.

Considérons la suite de fonctionnelles simples définies par  $e_t^{(n)} = B_{2^{-n} \lfloor 2^n t \rfloor}$ . Il suffit alors de vérifier que

$$\lim_{n \rightarrow \infty} \int_0^t e_s^{(n)} dB_s = \frac{1}{2} B_t^2 - \frac{t}{2} . \quad (8.3.2)$$

Notons  $t_k = k2^{-n}$  pour  $k \leq m = \lfloor 2^n t \rfloor$  et  $t_{m+1} = t$ . Il suit de la définition (8.1.4) que

$$\begin{aligned} 2 \int_0^t e_s^{(n)} dB_s &= 2 \sum_{k=1}^{m+1} B_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) \\ &= \sum_{k=1}^{m+1} \left[ B_{t_k}^2 - B_{t_{k-1}}^2 - (B_{t_k} - B_{t_{k-1}})^2 \right] \\ &= B_t^2 - \sum_{k=1}^{m+1} (B_{t_k} - B_{t_{k-1}})^2. \end{aligned} \quad (8.3.3)$$

Considérons la martingale

$$M_t^{(n)} = \sum_{k=1}^{m+1} (B_{t_k} - B_{t_{k-1}})^2 - t = \sum_{k=1}^{m+1} [(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})]. \quad (8.3.4)$$

Les termes de la somme sont indépendants et d'espérance nulle. Nous avons donc

$$\begin{aligned} \mathbb{E}((M_t^{(n)})^2) &= \sum_{k=1}^{m+1} \mathbb{E}([(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})]^2) \\ &\leq (m+1) \mathbb{E}([(B_{t_1} - B_{t_0})^2 - (t_1 - t_0)]^2) \\ &\leq \text{const } 2^n \mathbb{E}([(B_{2^{-n}})^2 - 2^{-n}]^2) \\ &= \text{const } 2^{-n} \mathbb{E}([(B_1)^2 - 1]^2) \\ &\leq \text{const } 2^{-n}, \end{aligned} \quad (8.3.5)$$

à cause de la propriété d'échelle.  $(M_t^{(n)})^2$  étant une sous-martingale, l'inégalité de Doob nous donne

$$\mathbb{P} \left\{ \sup_{0 \leq s \leq t} (M_s^{(n)})^2 > n^2 2^{-n} \right\} \leq 2^n n^{-2} \mathbb{E}((M_t^{(n)})^2) \leq \text{const } n^{-2}. \quad (8.3.6)$$

Le lemme de Borel–Cantelli nous permet alors de conclure que

$$\mathbb{P} \left\{ \sup_{0 \leq s \leq t} |M_s^{(n)}| < n 2^{-n/2}, n \rightarrow \infty \right\} = 1, \quad (8.3.7)$$

ce qui prouve (8.3.2).

## 8.4 La formule d'Itô

Considérons une intégrale stochastique de la forme

$$X_t = X_0 + \int_0^t f_s ds + \int_0^t g_s dB_s, \quad t \in [0, T] \quad (8.4.1)$$

où  $X_0$  est indépendante du mouvement Brownien, et  $f_s$  et  $g_s$  sont des fonctionnelles nonanticipatives satisfaisant

$$\begin{aligned} \mathbb{P} \left\{ \int_0^T |f_s| ds < \infty \right\} &= 1 \\ \mathbb{P} \left\{ \int_0^T g_s^2 ds < \infty \right\} &= 1. \end{aligned} \quad (8.4.2)$$

Le processus (8.4.1) s'écrit également sous forme différentielle

$$dX_t = f_t dt + g_t dB_t . \quad (8.4.3)$$

Par exemple, la relation (8.3.1) est équivalente à

$$d(B_t^2) = dt + 2B_t dB_t . \quad (8.4.4)$$

La formule d'Itô permet de déterminer de manière générale l'effet d'un changement de variables sur une différentielle stochastique.

**Lemme 8.4.1** (Formule d'Itô). *Soit  $u : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $(t, x) \mapsto u(t, x)$  une fonction continûment différentiable par rapport à  $t$  et deux fois continûment différentiable par rapport à  $x$ . Alors le processus stochastique  $Y_t = u(t, X_t)$  satisfait l'équation*

$$\begin{aligned} Y_t = Y_0 + \int_0^t \frac{\partial u}{\partial t}(s, X_s) ds + \int_0^t \frac{\partial u}{\partial x}(s, X_s) f_s ds + \int_0^t \frac{\partial u}{\partial x}(s, X_s) g_s dB_s \\ + \frac{1}{2} \int_0^t \frac{\partial^2 u}{\partial x^2}(s, X_s) g_s^2 ds . \end{aligned} \quad (8.4.5)$$

DÉMONSTRATION. Il suffit de prouver le résultat pour des fonctionnelles simples, et par l'additivité des intégrales, on peut se ramener au cas de fonctionnelles constantes. Mais alors  $X_t = f_0 t + g_0 B_t$  et  $Y_t = u(t, f_0 t + g_0 B_t)$  peut s'exprimer comme une fonction de  $(t, B_t)$ . Il suffit en définitive de considérer le cas  $X_t = B_t$ . Or pour une partition  $0 = t_0 < t_1 < \dots < t_n = t$ , on a

$$\begin{aligned} u(t, B_t) - u(0, 0) &= \sum_{k=1}^n [u(t_k, B_{t_k}) - u(t_{k-1}, B_{t_k})] + [u(t_{k-1}, B_{t_k}) - u(t_{k-1}, B_{t_{k-1}})] \\ &= \sum_{k=1}^n \frac{\partial u}{\partial t}(t_{k-1}, B_{t_k})(t_k - t_{k-1}) + \frac{\partial u}{\partial x}(t_{k-1}, B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}}) \\ &\quad + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t_{k-1}, B_{t_{k-1}})(B_{t_k} - B_{t_{k-1}})^2 + \mathcal{O}(t_k - t_{k-1}) + \mathcal{O}((B_{t_k} - B_{t_{k-1}})^2) \\ &= \int_0^t \frac{\partial u}{\partial t}(s, B_s) ds + \int_0^t \frac{\partial u}{\partial x}(s, B_s) dB_s + \frac{1}{2} \int_0^t \frac{\partial^2 u}{\partial x^2}(s, B_s) ds \\ &\quad + \sum_{k=1}^n \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t_{k-1}, B_{t_{k-1}})[(B_{t_k} - B_{t_{k-1}})^2 - (t_k - t_{k-1})] + \mathcal{O}(1) . \end{aligned} \quad (8.4.6)$$

La somme se traite comme  $M_t^{(n)}$  dans la section 8.3 lorsque  $t_k - t_{k-1} \rightarrow 0$ , c.f. (8.3.4).  $\square$

**Remarque 8.4.2.**

1. La formule d'Itô s'écrit aussi sous forme différentielle,

$$dY_t = \frac{\partial u}{\partial t}(t, X_t) dt + \frac{\partial u}{\partial x}(t, X_t) [f_t dt + g_t dB_t] + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, X_t) g_t^2 dt . \quad (8.4.7)$$

2. Un moyen mnémotechnique pour retrouver la formule est de l'écrire sous la forme

$$dY_t = \frac{\partial u}{\partial t} dt + \frac{\partial u}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 u}{\partial x^2} dX_t^2 , \quad (8.4.8)$$

où  $dX_t^2$  se calcule en utilisant les règles

$$dt^2 = dt dB_t = 0, \quad dB_t^2 = dt . \quad (8.4.9)$$

3. La formule se généralise à des fonctions  $u(t, X_t^{(1)}, \dots, X_t^{(n)})$ , dépendant de  $n$  processus définis par  $dX_t^{(i)} = f_t^{(i)} dt + g_t^{(i)} dB_t$ , en

$$dY_t = \frac{\partial u}{\partial t} dt + \sum_i \frac{\partial u}{\partial x_i} dX_t^{(i)} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j} dX_t^{(i)} dX_t^{(j)}, \quad (8.4.10)$$

où  $dX_t^{(i)} dX_t^{(j)} = g_t^{(i)} g_t^{(j)} dt$ .

**Exemple 8.4.3.**

1. Si  $X_t = B_t$  et  $u(x) = x^2$ , on retrouve la relation (8.4.4).
2. Si  $dX_t = g_t dB_t - \frac{1}{2}g_t^2 dt$  et  $u(x) = e^x$ , on obtient

$$d(e^{X_t}) = g_t e^{X_t} dB_t. \quad (8.4.11)$$

Ainsi la martingale  $M_t = \exp\{\gamma B_t - \frac{\gamma^2 t}{2}\}$  de la proposition 7.5.2 est la solution de l'équation  $dM_t = \gamma M_t dB_t$ .

## 8.5 Exercices

**Exercice 8.1.** On considère les deux processus stochastiques

$$X_t = \int_0^t e^s dB_s, \quad Y_t = e^{-t} X_t.$$

1. Déterminer  $\mathbb{E}(X_t)$ ,  $\text{Var}(X_t)$ ,  $\mathbb{E}(Y_t)$  et  $\text{Var}(Y_t)$ .
2. Spécifier la loi de  $X_t$  et de  $Y_t$ .
3. Montrer que  $Y_t$  converge en loi vers une variable  $Y_\infty$  lorsque  $t \rightarrow \infty$  et spécifier sa loi.
4. Exprimer  $dY_t$  en fonction de  $Y_t$  et de  $B_t$ .

**Exercice 8.2.** Soit

$$X_t = \int_0^t s dB_s.$$

1. Calculer  $\mathbb{E}(X_t)$  et  $\text{Var}(X_t)$ .
2. Quelle est la loi de  $X_t$ ?
3. Calculer  $d(tB_t)$  à l'aide de la formule d'Itô.
4. En déduire une relation entre  $X_t$  et

$$Y_t = \int_0^t B_s ds.$$

5. Calculer la variance de  $Y_t$ ,
  - (a) directement à partir de sa définition;
  - (b) en calculant d'abord la covariance de  $B_t$  et  $X_t$ , à l'aide d'une partition de  $[0, t]$ .
 En déduire la loi de  $Y_t$ .

**Exercice 8.3.**

1. Soit  $Y$  une variable aléatoire normale centrée, de variance  $\sigma^2$ . Montrer que

$$\mathbb{E}(e^Y) = e^{\sigma^2/2}$$

2. Soit  $B_t$  un mouvement Brownien standard, et  $\varphi : [0, T] \rightarrow \mathbb{R}$  une fonction indépendante de  $B_t$ . Pour  $t \in [0, T]$  on pose

$$X_t = \int_0^t \varphi(s) dB_s$$

Calculer  $\mathbb{E}(X_t)$  et  $\text{Var } X_t$ . On précisera les hypothèses faites sur la fonction  $\varphi$ .

3. Montrer que

$$M_t = \exp\left\{X_t - \frac{1}{2} \int_0^t \varphi(s)^2 ds\right\}$$

est une martingale.

4. Démontrer l'*inégalité de Bernstein* : Pour tout  $\lambda > 0$ ,

$$\mathbb{P}\left\{\sup_{0 \leq s \leq t} X_s > \lambda\right\} \leq \exp\left\{-\frac{\lambda^2}{2\Phi(t)}\right\}$$

$$\text{où } \Phi(t) = \int_0^t \varphi(s)^2 ds.$$

**Exercice 8.4.** Soit  $\{B_t\}_{t \in [0, T]}$  un mouvement Brownien standard. Soit  $0 = t_0 < t_1 < \dots < t_N = T$  une partition de  $[0, T]$ , et soit

$$e_t = \sum_{k=1}^N e_{t_{k-1}} 1_{[t_{k-1}, t_k)}(t)$$

une fonction simple, adaptée à la filtration canonique du mouvement Brownien.

L'intégrale de Stratonovich de  $e_t$  est définie par

$$\int_0^T e_t \circ dB_t = \sum_{k=1}^N \frac{e_{t_k} + e_{t_{k-1}}}{2} \Delta B_k \quad \text{où } \Delta B_k = B_{t_k} - B_{t_{k-1}}.$$

L'intégrale de Stratonovich  $\int_0^T X_t \circ dB_t$  d'un processus adapté  $X_t$  est définie comme la limite de la suite  $\int_0^T e_t^{(n)} \circ dB_t$ , où  $e^{(n)}$  est une suite de fonctions simples convergeant vers  $X_t$  dans  $L^2$ . On admettra que cette limite existe et est indépendante de la suite  $e^{(n)}$ .

1. Calculer

$$\int_0^T B_t \circ dB_t.$$

2. Soit  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction de classe  $C^2$ , et soit  $X_t$  un processus adapté satisfaisant

$$X_t = \int_0^t g(X_s) \circ dB_s \quad \forall t \in [0, T].$$

Soit  $Y_t$  l'intégrale d'Itô

$$Y_t = \int_0^t g(X_s) dB_s.$$

Montrer que

$$X_t - Y_t = \frac{1}{2} \int_0^t g'(X_s) g(X_s) ds \quad \forall t \in [0, T].$$



# Chapitre 9

## Equations différentielles stochastiques

Nous avons maintenant tous les éléments en main pour définir la notion de solution d'une équation différentielle stochastique (EDS), de la forme

$$dX_t = f(X_t, t) dt + g(X_t, t) dB_t, \quad (9.0.1)$$

où  $f, g : \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$  sont des fonctions déterministes mesurables. La fonction  $f$  est communément appelée *coefficient de dérive*, alors que  $g$  est appelée *coefficient de diffusion*.

Dans tout ce chapitre, nous supposons que

- Soit  $X_0 \in \mathbb{R}$  est une constante, et alors  $\{\mathcal{F}_t\}_{t \in [0, T]}$  désigne la filtration engendrée par le mouvement Brownien.
- Soit  $X_0 : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire, de carré intégrable, et indépendante du mouvement Brownien. Dans ce cas,  $\{\mathcal{F}_t\}_{t \in [0, T]}$  désignera la filtration engendrée par le mouvement Brownien et par  $X_0$ .

### 9.1 Solutions fortes

**Définition 9.1.1.** *Un processus stochastique  $\{X_t\}_{t \in [0, T]}$  est appelé une solution forte de l'EDS (9.0.1) avec condition initiale  $X_0$  si*

- $X_t$  est  $\mathcal{F}_t$ -mesurable pour tout  $t \in [0, T]$ ;
- on a les conditions de régularité

$$\mathbb{P}\left\{\int_0^T |f(X_s, s)| ds < \infty\right\} = \mathbb{P}\left\{\int_0^T g(X_s, s)^2 ds < \infty\right\} = 1; \quad (9.1.1)$$

- pour tout  $t \in [0, T]$ , on a

$$X_t = X_0 + \int_0^t f(X_s, s) ds + \int_0^t g(X_s, s) dB_s \quad (9.1.2)$$

avec probabilité 1.

**Remarque 9.1.2.** On définit de manière similaire la solution forte d'une EDS multidimensionnelle, c'est-à-dire qu'on peut supposer que  $X \in \mathbb{R}^n$ ,  $f(x, t)$  prend des valeurs dans  $\mathbb{R}^n$ , et, si  $B_t$  est un mouvement Brownien de dimension  $k$ ,  $g(x, t)$  prend des valeurs dans les matrices  $n \times k$ .

Avant de montrer l'existence de solutions fortes dans un cadre général, nous donnons quelques exemples d'équations pour lesquelles une solution peut être donnée explicitement sous forme d'intégrales. Il n'est pas surprenant que de tels cas solubles sont extrêmement rares.

**Exemple 9.1.3.** Considérons l'EDS linéaire avec "bruit additif"

$$dX_t = a(t)X_t dt + \sigma(t) dB_t, \quad (9.1.3)$$

où  $a$  et  $\sigma$  sont des fonctions déterministes. Dans le cas particulier  $\sigma \equiv 0$ , la solution peut s'écrire simplement

$$X_t = e^{\alpha(t)} X_0, \quad \alpha(t) = \int_0^t a(s) ds. \quad (9.1.4)$$

Ceci suggère d'appliquer la méthode de la variation de la constante, c'est-à-dire de chercher une solution de la forme  $X_t = e^{\alpha(t)} Y_t$ . La formule d'Itô appliquée à  $Y_t = u(X_t, t) = e^{-\alpha(t)} X_t$  nous donne

$$dY_t = -a(t) e^{-\alpha(t)} X_t dt + e^{-\alpha(t)} dX_t = e^{-\alpha(t)} \sigma(t) dB_t, \quad (9.1.5)$$

d'où en intégrant et en tenant compte du fait que  $Y_0 = X_0$ ,

$$Y_t = X_0 + \int_0^t e^{-\alpha(s)} \sigma(s) dB_s. \quad (9.1.6)$$

Ceci donne finalement la solution forte de l'équation (9.1.3)

$$X_t = X_0 e^{\alpha(t)} + \int_0^t e^{\alpha(t)-\alpha(s)} \sigma(s) dB_s. \quad (9.1.7)$$

On vérifie effectivement (9.1.2) en appliquant encore une fois la formule d'Itô. On notera que si la condition initiale  $X_0$  est déterministe, alors  $X_t$  suit une loi normale, d'espérance  $\mathbb{E}(X_t) = X_0 e^{\alpha(t)}$  et de variance

$$\text{Var}(X_t) = \int_0^t e^{2(\alpha(t)-\alpha(s))} \sigma(s)^2 ds, \quad (9.1.8)$$

en vertu de l'isométrie d'Itô.

**Exemple 9.1.4.** Soit l'EDS linéaire avec "bruit multiplicatif"

$$dX_t = a(t)X_t dt + \sigma(t)X_t dB_t, \quad (9.1.9)$$

avec à nouveau  $a$  et  $\sigma$  des fonctions déterministes. Nous pouvons alors écrire

$$\frac{dX_t}{X_t} = a(t) dt + \sigma(t) dB_t. \quad (9.1.10)$$

En intégrant le membre de gauche, on devrait trouver  $\log(X_t)$ , mais ceci est-il compatible avec le calcul d'Itô? Pour s'en assurer, posons  $Y_t = u(X_t) = \log(X_t)$ . Alors la formule d'Itô donne

$$\begin{aligned} dY_t &= \frac{1}{X_t} dX_t - \frac{1}{2X_t^2} dX_t^2 \\ &= a(t) dt + \sigma(t) dB_t - \frac{1}{2} \sigma(t)^2 dt. \end{aligned} \quad (9.1.11)$$

En intégrant et en reprenant l'exponentielle, on obtient donc la solution forte

$$X_t = X_0 \exp \left\{ \int_0^t \left[ a(s) - \frac{1}{2} \sigma(s)^2 \right] ds + \int_0^t \sigma(s) dB_s \right\}. \quad (9.1.12)$$

En particulier, si  $a \equiv 0$  et  $\sigma \equiv \gamma$ , on retrouve la martingale  $X_t = X_0 \exp\{\gamma B_t - \gamma^2 t/2\}$ , appelée *mouvement Brownien exponentiel*.

## 9.2 Existence et unicité de solutions

Nous donnons d'abord un résultat d'existence et d'unicité d'une solution forte sous des conditions un peu restrictives sur les coefficients  $f$  et  $g$ .

**Théorème 9.2.1.** *Supposons que les fonctions  $f$  et  $g$  satisfont les deux conditions suivantes:*

1. Condition de Lipschitz globale: *Il existe une constante  $K$  telle que*

$$|f(x, t) - f(y, t)| + |g(x, t) - g(y, t)| \leq K|x - y| \quad (9.2.1)$$

*pour tous les  $x, y \in \mathbb{R}$  et  $t \in [0, T]$ .*

2. Condition de croissance: *Il existe une constante  $L$  telle que*

$$|f(x, t)| + |g(x, t)| \leq L(1 + |x|) \quad (9.2.2)$$

*pour tous les  $x \in \mathbb{R}$  et  $t \in [0, T]$ .*

Alors l'EDS (9.0.1) admet, pour toute condition initiale  $X_0$  de carré intégrable, une solution forte  $\{X_t\}_{t \in [0, T]}$ , presque sûrement continue. Cette solution est unique dans le sens que si  $\{X_t\}_{t \in [0, T]}$  et  $\{Y_t\}_{t \in [0, T]}$  sont deux solutions presque sûrement continues, alors

$$\mathbb{P} \left\{ \sup_{0 \leq t \leq T} |X_t - Y_t| > 0 \right\} = 0. \quad (9.2.3)$$

**DÉMONSTRATION.** Commençons par démontrer l'unicité. Soient  $X_t$  et  $Y_t$  deux solutions fortes de condition initiale  $X_0$ . Posons  $\phi(t) = f(X_t, t) - f(Y_t, t)$  et  $\gamma(t) = g(X_t, t) - g(Y_t, t)$ . Alors en utilisant l'inégalité de Cauchy-Schwartz et l'isométrie d'Itô on trouve

$$\begin{aligned} \mathbb{E}(|X_t - Y_t|^2) &= \mathbb{E} \left[ \left( \int_0^t \phi(s) ds + \int_0^t \gamma(s) dB_s \right)^2 \right] \\ &\leq 2\mathbb{E} \left[ \left( \int_0^t \phi(s) ds \right)^2 \right] + 2\mathbb{E} \left[ \left( \int_0^t \gamma(s) dB_s \right)^2 \right] \\ &\leq 2\mathbb{E} \left[ t \int_0^t \phi(s)^2 ds \right] + 2\mathbb{E} \left[ \int_0^t \gamma(s)^2 ds \right] \\ &\leq 2(1+t)K^2 \int_0^t \mathbb{E}(|X_s - Y_s|^2) ds. \end{aligned}$$

Ainsi la fonction  $v(t) = \mathbb{E}(|X_t - Y_t|^2)$  satisfait l'inégalité

$$v(t) \leq 2(1+T)K^2 \int_0^t v(s) ds, \quad (9.2.4)$$

donc par l'inégalité de Gronwall,

$$v(t) \leq v(0) e^{2(1+T)K^2 t} = 0, \quad (9.2.5)$$

puisque  $v(0) = 0$ . Il suit que

$$\mathbb{P}\{|X_t - Y_t| = 0 \forall t \in \mathbb{Q} \cap [0, T]\} = 1, \quad (9.2.6)$$

ce qui implique (9.2.3) en vertu de la continuité des trajectoires.

Afin de prouver l'existence, nous posons  $X_t^{(0)} = X_0$  et définissons une suite  $\{X_t^{(k)}\}_{k \in \mathbb{N}}$  récursivement par

$$X_t^{(k+1)} = X_0 + \int_0^t f(X_s^{(k)}, s) ds + \int_0^t g(X_s^{(k)}, s) dB_s. \quad (9.2.7)$$

Un calcul similaire au calcul ci-dessus montre que

$$\mathbb{E}(|X_t^{(k+1)} - X_t^{(k)}|^2) \leq 2(1+T)K^2 \int_0^t \mathbb{E}(|X_s^{(k)} - X_s^{(k-1)}|^2) ds \quad (9.2.8)$$

pour tout  $k \geq 1$  et  $t \in [0, T]$ , alors que

$$\mathbb{E}(|X_t^{(1)} - X_t^{(0)}|^2) \leq A_1 t \quad (9.2.9)$$

pour une constante  $A_1 = A_1(L, T, X_0)$ . Par récurrence, nous avons donc

$$\mathbb{E}(|X_t^{(k+1)} - X_t^{(k)}|^2) \leq \frac{A_2^{k+1} t^{k+1}}{(k+1)!} \quad (9.2.10)$$

pour tout  $k \geq 1$  et  $t \in [0, T]$  et une constante  $A_2 = A_2(K, L, T, X_0)$ . Il est alors facile de vérifier que la suite des  $X^{(k)}$  est une suite de Cauchy dans  $L^2(\mathbb{P} \times \lambda)$ , c'est-à-dire par rapport au produit scalaire défini par la mesure  $\mathbb{P}$  sur  $\Omega$  et la mesure de Lebesgue  $\lambda$  sur  $[0, T]$ . Il existe donc un processus limite

$$X_t = \lim_{n \rightarrow \infty} X_t^{(n)} \in L^2(\mathbb{P} \times \lambda). \quad (9.2.11)$$

Il reste à montrer que c'est la solution forte cherchée. Faisons tendre  $k$  vers l'infini dans (9.2.7). Le membre de gauche tend vers  $X_t$ . La première intégrale dans le membre de droite tend dans  $L^2(\mathbb{P} \times \lambda)$  vers l'intégrale de  $f(X_s, s) ds$  par l'inégalité de Hölder. La seconde tend dans  $L^2(\mathbb{P} \times \lambda)$  vers l'intégrale de  $g(X_s, s) dB_s$  par l'isométrie d'Itô. Ceci montre que  $X_t$  satisfait (9.1.2). Les relations (9.1.1) sont satisfaites en conséquence de la condition de croissance et du fait que  $X_t$  est de carré intégrable. Finalement, le Théorème 8.1.5 implique la continuité presque sûre de  $X_t$ .  $\square$

Les conditions (9.2.1) et (9.2.2) sont en fait trop restrictives, et peuvent être remplacées par les conditions plus faibles suivantes :

1. *Condition de Lipschitz locale*: Pour tout compact  $\mathcal{K} \in \mathbb{R}$ , il existe une constante  $K = K(\mathcal{K})$  telle que

$$|f(x, t) - f(y, t)| + |g(x, t) - g(y, t)| \leq K|x - y| \quad (9.2.12)$$

pour  $x, y \in \mathcal{K}$  et  $t \in [0, T]$ .

2. *Condition de croissance*: Il existe une constante  $L$  telle que

$$xf(x, t) + g(x, t)^2 \leq L^2(1 + x^2) \quad (9.2.13)$$

pour tous  $x, t$ .

Par exemple, le coefficient de dérive  $f(x, t) = -x^2$  satisfait les conditions (9.2.12) et (9.2.13), alors qu'il ne satisfait pas les conditions (9.2.1) et (9.2.2). Par contre, la condition de Lipschitz locale est nécessaire pour garantir l'unicité, comme le montre le contre-exemple classique  $f(x, t) = 3x^{2/3}$  (toute fonction valant 0 jusqu'à un temps arbitraire  $a$ , puis  $(t - a)^3$  pour les temps ultérieurs est solution). De même, le contre-exemple  $f(x, t) = x^2$ , dont les solutions divergent pour un temps fini, montre la nécessité de la condition de croissance (9.2.13).

Sans entrer dans les détails, pour montrer l'existence d'une unique solution forte sous les conditions (9.2.12) et (9.2.13), on procède en deux étapes, tout à fait similaires à celles du cas déterministe :

- On montre que sous la condition de Lipschitz locale, toute trajectoire solution  $X_t(\omega)$  soit existe jusqu'au temps  $T$ , soit quitte tout compact  $\mathcal{K}$  en un temps  $\tau(\omega) < T$ . Par conséquent, il existe un temps d'arrêt  $\tau$ , appelé *temps d'explosion*, tel que soit  $\tau(\omega) = +\infty$  et alors  $X_t(\omega)$  existe jusqu'au temps  $T$ , soit  $\tau(\omega) \leq T$ , et alors  $X_t(\omega) \rightarrow \pm\infty$  lorsque  $t \rightarrow \tau(\omega)$ .
- On montre que sous la condition de croissance, les trajectoires  $X_t(\omega)$  ne peuvent pas exploser (car le terme de dérive ne croît pas assez vite, ou ramène les trajectoires vers l'origine si  $xf(x, t)$  est négatif).

### 9.3 Exercices

**Exercice 9.1.** On considère l'équation

$$dX_t = a(t)X_t dt + b(t) dt + c(t) dB_t,$$

où  $a(t)$ ,  $b(t)$  et  $c(t)$  sont des processus adaptés.

Résoudre cette équation par la méthode de la variation de la constante, c'est-à-dire

1. Soit  $\alpha(t) = \int_0^t a(s) ds$ . Vérifier que  $X_0 e^{\alpha(t)}$  est la solution de l'équation homogène, c-à-d avec  $b = c = 0$ .
2. Poser  $Y_t = e^{-\alpha(t)} X_t$  et calculer  $dY_t$  à l'aide de la formule d'Itô.
3. En déduire  $Y_t$  puis  $X_t$  sous forme intégrale.
4. Résoudre l'EDS

$$dX_t = -\frac{1}{1+t} X_t dt + \frac{1}{1+t} dB_t, \quad X_0 = 0.$$

**Exercice 9.2.** Résoudre l'EDS

$$dX_t = -\frac{1}{2} X_t dt + \sqrt{1 - X_t^2} dB_t, \quad X_0 = 0$$

à l'aide du changement de variable  $Y = \text{Arcsin}(X)$ .

**Exercice 9.3.**

1. En utilisant l'exercice 9.1, résoudre l'EDS:

$$dX_t = \frac{b - X_t}{1 - t} dt + dB_t, \quad 0 \leq t < 1, \quad X_0 = a.$$

2. Déterminer la variance de  $X_t$ . Calculer  $\lim_{t \rightarrow 1^-} X_t$  dans  $L^2$ .

3. Soit  $M_t = \int_0^t (1 - s)^{-1} dB_s$ . Montrer que

$$\mathbb{P} \left\{ \sup_{1-2^{-n} \leq t \leq 1-2^{-n-1}} (1-t)|M_t| > \varepsilon \right\} \leq \frac{2}{\varepsilon^2 2^n}.$$

A l'aide du lemme de Borel–Cantelli, en déduire  $\lim_{t \rightarrow 1^-} X_t$  au sens presque sûr.

4. Le processus  $X_t$  est appelé un pont Brownien — expliquer pourquoi.

**Exercice 9.4.** On se donne  $r, \alpha \in \mathbb{R}$ . Résoudre l'équation différentielle stochastique

$$dY_t = r dt + \alpha Y_t dB_t, \quad Y_0 = 1.$$

*Indication :* Soit le “facteur intégrant”

$$F_t = \exp \left\{ -\alpha B_t + \frac{1}{2} \alpha^2 t \right\}.$$

Considérer  $X_t = F_t Y_t$ .

# Chapitre 10

## Diffusions

On appelle *diffusion* un processus stochastique obéissant à une équation différentielle stochastique de la forme

$$dX_t = f(X_t) dt + g(X_t) dB_t . \quad (10.0.1)$$

Le terme  $f(x)$  peut s'interpréter comme la force déterministe agissant sur une particule dans un fluide au point  $x$ , et s'appelle donc le coefficient de dérive. Le terme  $g(x)$  mesure l'effet de l'agitation thermique des molécules du fluide en  $x$ , et s'appelle le coefficient de diffusion.

Dans l'étude des diffusions, il est particulièrement intéressant de considérer la dépendance des solutions dans la condition initiale  $X_0 = x$ . La propriété de Markov affirme que l'état  $X_t$  en un temps donné  $t$  détermine univoquement le comportement à tous les temps futurs. Ceci permet de démontrer la propriété de semi-groupe, qui généralise celle du flot d'une équation différentielle ordinaire. Un semi-groupe de Markov peut être caractérisé par son générateur, qui s'avère être un opérateur différentiel du second ordre dans le cas des diffusions. Il en résulte un ensemble de liens importants entre équations différentielles stochastiques et équations aux dérivées partielles.

### 10.1 La propriété de Markov

**Définition 10.1.1** (Diffusion d'Itô). *Une diffusion d'Itô homogène dans le temps est un processus stochastique  $\{X_t(\omega)\}_{t \geq 0}$  satisfaisant une équation différentielle stochastique de la forme*

$$dX_t = f(X_t) dt + g(X_t) dB_t , \quad t \geq s > 0 , \quad X_s = x , \quad (10.1.1)$$

où  $B_t$  est un mouvement Brownien standard de dimension  $m$ , et le coefficient de dérive  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  et le coefficient de diffusion  $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$  sont tels que l'EDS (10.1.1) admette une unique solution en tout temps.

Nous noterons la solution de (10.1.1)  $X_t^{s,x}$ . L'homogénéité en temps, c'est-à-dire le fait que  $f$  et  $g$  ne dépendent pas du temps, a la conséquence importante suivante.

**Lemme 10.1.2.** *Les processus  $\{X_{s+h}^{s,x}\}_{h \geq 0}$  et  $\{X_h^{0,x}\}_{h \geq 0}$  ont la même loi.*

DÉMONSTRATION. Par définition,  $X_h^{0,x}$  satisfait l'équation intégrale

$$X_h^{0,x} = x + \int_0^h f(X_v^{0,x}) dv + \int_0^h g(X_v^{0,x}) dB_v . \quad (10.1.2)$$

De même,  $X_{s+h}^{s,x}$  satisfait l'équation

$$\begin{aligned} X_{s+h}^{s,x} &= x + \int_s^{s+h} f(X_u^{s,x}) du + \int_s^{s+h} g(X_u^{s,x}) dB_u \\ &= x + \int_0^h f(X_{s+v}^{s,x}) dv + \int_0^h g(X_{s+v}^{s,x}) d\tilde{B}_v \end{aligned} \quad (10.1.3)$$

où nous avons utilisé le changement de variable  $u = s + v$ , et  $\tilde{B}_v = B_{s+v} - B_s$ . Par la propriété différentielle,  $\tilde{B}_v$  est un mouvement Brownien standard, donc par unicité des solutions de l'EDS (10.1.1), les intégrales (10.1.3) et (10.1.2) ont la même loi.  $\square$

Nous noterons  $\mathbb{P}^x$  la mesure de probabilité sur la tribu engendrée par toutes les variables aléatoires  $X_t^{0,x}$ ,  $t \geq 0$ ,  $x \in \mathbb{R}^n$ , définie par

$$\mathbb{P}^x \{X_{t_1} \in A_1, \dots, X_{t_k} \in A_k\} = \mathbb{P} \{X_{t_1}^{0,x} \in A_1, \dots, X_{t_k}^{0,x} \in A_k\} \quad (10.1.4)$$

pour tout choix de temps  $0 \leq t_1 < t_2 < \dots < t_k$  et de boréliens  $A_1, \dots, A_k \subset \mathbb{R}^n$ . Les espérances par rapport à  $\mathbb{P}^x$  seront notées  $\mathbb{E}^x$ .

**Théorème 10.1.3** (Propriété de Markov pour les diffusions d'Itô). *Pour toute fonction mesurable bornée  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ ,*

$$\mathbb{E}^x(\varphi(X_{t+h}) \mid \mathcal{F}_t)(\omega) = \mathbb{E}^{X_t(\omega)}(\varphi(X_h)) , \quad (10.1.5)$$

le membre de droite désignant la fonction  $\mathbb{E}^y(\varphi(X_h))$  évaluée en  $y = X_t(\omega)$ .

DÉMONSTRATION. Considérons pour  $y \in \mathbb{R}^n$  et  $s \geq t$  la fonction

$$F(y, t, s, \omega) = X_s^{t,y}(\omega) = y + \int_t^s f(X_u(\omega)) du + \int_t^s g(X_u(\omega)) dB_u(\omega) . \quad (10.1.6)$$

On notera que  $F$  est indépendante de  $\mathcal{F}_t$ . Par unicité des solutions de l'EDS (10.1.1), on a

$$X_s(\omega) = F(X_t(\omega), t, s, \omega) . \quad (10.1.7)$$

Posons  $g(y, \omega) = \varphi \circ F(y, t, t+h, \omega)$ . On vérifie que cette fonction est mesurable. La relation (10.1.5) est alors équivalente à

$$\mathbb{E}(g(X_t, \omega) \mid \mathcal{F}_t) = \mathbb{E}(\varphi \circ F(y, 0, h, \omega)) \Big|_{y=X_t(\omega)} . \quad (10.1.8)$$

On a

$$\mathbb{E}(g(X_t, \omega) \mid \mathcal{F}_t) = \mathbb{E}(g(y, \omega) \mid \mathcal{F}_t) \Big|_{y=X_t(\omega)} . \quad (10.1.9)$$

En effet, cette relation est vraie pour des fonctions de la forme  $g(y, \omega) = \phi(y)\psi(\omega)$ , puisque

$$\mathbb{E}(\phi(X_t)\psi(\omega) \mid \mathcal{F}_t) = \phi(X_t)\mathbb{E}(\psi(\omega) \mid \mathcal{F}_t) = \mathbb{E}(\phi(y)\psi(\omega) \mid \mathcal{F}_t) \Big|_{y=X_t(\omega)} . \quad (10.1.10)$$

Elle s'étend alors à toute fonction mesurable bornée en approximant celle-ci par une suite de combinaisons linéaires de fonctions comme ci-dessus. Or il suit de l'indépendance de  $F$  et de  $\mathcal{F}_t$  que

$$\begin{aligned} \mathbb{E}(g(y, \omega) \mid \mathcal{F}_t) &= \mathbb{E}(g(y, \omega)) \\ &= \mathbb{E}(\varphi \circ F(y, t, t+h, \omega)) \\ &= \mathbb{E}(\varphi \circ F(y, 0, h, \omega)) , \end{aligned} \quad (10.1.11)$$

la dernière égalité suivant du Lemme 10.1.2. Le résultat s'obtient alors en évaluant la dernière égalité en  $y = X_t$ .  $\square$

Comme pour le mouvement Brownien, la propriété de Markov se généralise à des temps d'arrêt.

**Théorème 10.1.4** (Propriété de Markov forte pour les diffusions d'Itô). *Pour toute fonction mesurable bornée  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  et tout temps d'arrêt  $\tau$  fini presque sûrement,*

$$\mathbb{E}^x(\varphi(X_{\tau+h}) \mid \mathcal{F}_\tau)(\omega) = \mathbb{E}^{X_\tau(\omega)}(\varphi(X_h)) . \quad (10.1.12)$$

DÉMONSTRATION. La preuve est une adaptation relativement directe de la preuve précédente. Voir par exemple [Øks95, Theorem 7.2.4].  $\square$

## 10.2 Semigroupes et générateurs

**Définition 10.2.1** (Semi-groupe de Markov). *A toute fonction mesurable bornée  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ , on associe pour tout  $t \geq 0$  la fonction  $T_t\varphi$  définie par*

$$(T_t\varphi)(x) = \mathbb{E}^x(\varphi(X_t)) . \quad (10.2.1)$$

L'opérateur linéaire  $T_t$  est appelé le semi-groupe de Markov associé à la diffusion.

Par exemple, si  $\varphi(x) = 1_A(x)$  est la fonction indicatrice d'un Borélien  $A \subset \mathbb{R}^n$ , on a

$$(T_t 1_A)(x) = \mathbb{P}^x\{X_t \in A\} . \quad (10.2.2)$$

Le nom de semi-groupe est justifié par le résultat suivant.

**Lemme 10.2.2** (Propriété de semi-groupe). *Pour tous  $t, h \geq 0$ , on a*

$$T_h \circ T_t = T_{t+h} . \quad (10.2.3)$$

DÉMONSTRATION. On a

$$\begin{aligned} (T_h \circ T_t)(\varphi)(x) &= (T_h(T_t\varphi))(x) \\ &= \mathbb{E}^x((T_t\varphi)(X_h)) \\ &= \mathbb{E}^x(\mathbb{E}^{X_h}(\varphi(X_t))) \\ &= \mathbb{E}^x(\mathbb{E}^x(\varphi(X_{t+h}) \mid \mathcal{F}_t)) \\ &= \mathbb{E}^x(\varphi(X_{t+h})) \\ &= (T_{t+h}\varphi)(x) , \end{aligned} \quad (10.2.4)$$

où l'on a utilisé la propriété de Markov pour passer de la troisième à la quatrième ligne.  $\square$

De plus, on vérifie facilement les propriétés suivantes:

1.  $T_t$  préserve les fonctions constantes:  $T_t(c1_{\mathbb{R}^n}) = c1_{\mathbb{R}^n}$ ;
2.  $T_t$  préserve les fonctions non-négatives:  $\varphi(x) \geq 0 \forall x \Rightarrow (T_t\varphi)(x) \geq 0 \forall x$ ;
3.  $T_t$  est contractante par rapport à la norme  $L^\infty$ :

$$\sup_{x \in \mathbb{R}^n} |(T_t\varphi)(x)| = \sup_{x \in \mathbb{R}^n} |\mathbb{E}^x(\varphi(X_t))| \leq \sup_{y \in \mathbb{R}^n} |\varphi(y)| \sup_{x \in \mathbb{R}^n} \mathbb{E}^x(1) = \sup_{y \in \mathbb{R}^n} |\varphi(y)| . \quad (10.2.5)$$

Le semi-groupe de Markov est donc un opérateur linéaire positif, borné par rapport à la norme  $L^\infty$ . En fait, il est de norme opérateur 1.

La propriété de semi-groupe implique que le comportement de  $T_t$  sur tout intervalle  $[0, \varepsilon]$ , avec  $\varepsilon > 0$  arbitrairement petit, détermine son comportement pour tout  $t \geq 0$ . Il est donc naturel de considérer la dérivée de  $T_t$  en  $t = 0$ .

**Définition 10.2.3** (Générateur d'une diffusion d'Itô). *Le générateur infinitésimal  $L$  d'une diffusion d'Itô est défini par son action sur une fonction test  $\varphi$  via*

$$(L\varphi)(x) = \lim_{h \rightarrow 0_+} \frac{(T_h\varphi)(x) - \varphi(x)}{h}. \quad (10.2.6)$$

Le domaine de  $L$  est par définition l'ensemble des fonctions  $\varphi$  pour lesquelles la limite (10.2.6) existe pour tout  $x \in \mathbb{R}^n$ .

**Remarque 10.2.4.** Formellement, la relation (10.2.6) peut s'écrire

$$L = \left. \frac{dT_t}{dt} \right|_{t=0}. \quad (10.2.7)$$

Par la propriété de Markov, cette relation se généralise en

$$\frac{d}{dt}T_t = \lim_{h \rightarrow 0_+} \frac{T_{t+h} - T_t}{h} = \lim_{h \rightarrow 0_+} \frac{T_h - \text{id}}{h}T_t = LT_t, \quad (10.2.8)$$

et on peut donc écrire formellement

$$T_t = e^{tL}. \quad (10.2.9)$$

Nous précisons ce point dans la Section 10.4.

**Proposition 10.2.5.** *Le générateur de la diffusion d'Itô (10.1.1) est l'opérateur différentiel*

$$L = \sum_{i=1}^n f_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^n (gg^T)_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}. \quad (10.2.10)$$

Le domaine de  $L$  contient l'ensemble des fonctions deux fois continûment différentiables à support compact.

DÉMONSTRATION. Considérons le cas  $n = m = 1$ . Soit  $\varphi$  une fonction deux fois continûment différentiable à support compact, et soit  $Y_t = \varphi(X_t)$ . Par la formule d'Itô,

$$Y_h = \varphi(X_0) + \int_0^h \varphi'(X_s) f(X_s) ds + \int_0^h \varphi'(X_s) g(X_s) dB_s + \frac{1}{2} \int_0^h \varphi''(X_s) g(X_s)^2 ds. \quad (10.2.11)$$

En prenant l'espérance, comme l'espérance de l'intégrale d'Itô est nulle, on trouve

$$\mathbb{E}^x(Y_h) = \varphi(x) + \mathbb{E}^x \left( \int_0^h \varphi'(X_s) f(X_s) ds + \frac{1}{2} \int_0^h \varphi''(X_s) g(X_s)^2 ds \right), \quad (10.2.12)$$

d'où

$$\frac{\mathbb{E}^x(\varphi(X_h)) - \varphi(x)}{h} = \frac{1}{h} \int_0^h \mathbb{E}^x(\varphi'(X_s) f(X_s)) ds + \frac{1}{2h} \int_0^h \mathbb{E}^x(\varphi''(X_s) g(X_s)^2) ds. \quad (10.2.13)$$

En prenant la limite  $h \rightarrow 0_+$ , on obtient

$$(L\varphi)(x) = \varphi'(x)f(x) + \frac{1}{2}\varphi''(x)g(x)^2. \quad (10.2.14)$$

Les cas où  $n \geq 2$  ou  $m \geq 2$  se traitent de manière similaire, en utilisant la formule d'Itô multidimensionnelle.  $\square$

**Exemple 10.2.6** (Générateur du mouvement Brownien). Soit  $B_t$  le mouvement Brownien de dimension  $m$ . C'est un cas particulier de diffusion, avec  $f = 0$  et  $g = \mathbb{1}$ . Son générateur est donné par

$$L = \frac{1}{2} \sum_{i=1}^m \frac{\partial^2}{\partial x_i^2} = \frac{1}{2} \Delta. \quad (10.2.15)$$

C'est donc le Laplacien à un facteur  $1/2$  près.

### 10.3 La formule de Dynkin

La formule de Dynkin est essentiellement une généralisation de l'expression (10.2.12) à des temps d'arrêt. Elle fournit une première classe de liens entre diffusions et équations aux dérivées partielles.

**Proposition 10.3.1** (Formule de Dynkin). *Soit  $\{X_t\}_{t \geq 0}$  une diffusion de générateur  $L$ ,  $x \in \mathbb{R}^n$ ,  $\tau$  un temps d'arrêt tel que  $\mathbb{E}^x(\tau) < \infty$ , et  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction deux fois continûment différentiable à support compact. Alors*

$$\mathbb{E}^x(\varphi(X_\tau)) = \varphi(x) + \mathbb{E}^x\left(\int_0^\tau (L\varphi)(X_s) ds\right). \quad (10.3.1)$$

DÉMONSTRATION. Considérons le cas  $n = m = 1$ ,  $m$  étant la dimension du mouvement Brownien. En procédant comme dans la preuve de la Proposition 10.2.5, on obtient

$$\mathbb{E}^x(\varphi(X_\tau)) = \varphi(x) + \mathbb{E}^x\left(\int_0^\tau (L\varphi)(X_s) ds\right) + \mathbb{E}^x\left(\int_0^\tau g(X_s)\varphi'(X_s) dB_s\right). \quad (10.3.2)$$

Il suffit donc de montrer que l'espérance de l'intégrale stochastique est nulle. Or pour toute fonction  $h$  bornée par  $M$  et tout  $N \in \mathbb{N}$ , on a

$$\mathbb{E}^x\left(\int_0^{\tau \wedge N} h(X_s) dB_s\right) = \mathbb{E}^x\left(\int_0^N 1_{\{s < \tau\}} h(X_s) dB_s\right) = 0, \quad (10.3.3)$$

en vertu de la  $\mathcal{F}_s$ -mesurabilité de  $1_{\{s < \tau\}}$  et  $h(X_s)$ . De plus,

$$\begin{aligned} \mathbb{E}^x\left(\left[\int_0^\tau h(X_s) dB_s - \int_0^{\tau \wedge N} h(X_s) dB_s\right]^2\right) &= \mathbb{E}^x\left(\int_{\tau \wedge N}^\tau h(X_s)^2 ds\right) \\ &\leq M^2 \mathbb{E}^x(\tau - \tau \wedge N), \end{aligned} \quad (10.3.4)$$

qui tend vers 0 lorsque  $N \rightarrow \infty$ , en vertu de l'hypothèse  $\mathbb{E}^x(\tau) < \infty$ , par convergence dominée. On peut donc écrire

$$0 = \lim_{N \rightarrow \infty} \mathbb{E}^x\left(\int_0^{\tau \wedge N} h(X_s) dB_s\right) = \mathbb{E}^x\left(\int_0^\tau h(X_s) dB_s\right), \quad (10.3.5)$$

ce qui conclut la preuve, en substituant dans (10.3.2). La preuve du cas général est analogue.  $\square$

Considérons le cas où le temps d'arrêt  $\tau$  est le temps de première sortie d'un ouvert borné  $D \subset \mathbb{R}^n$ . Supposons que le problème avec conditions au bord

$$\begin{aligned} (Lu)(x) &= \theta(x) & x \in D \\ u(x) &= \psi(x) & x \in \partial D \end{aligned} \quad (10.3.6)$$

admet une unique solution. C'est le cas si  $D$ ,  $\theta$  et  $\psi$  sont suffisamment réguliers. Substituant  $\varphi$  par  $u$  dans la formule de Dynkin, on obtient la relation

$$u(x) = \mathbb{E}^x \left( \psi(X_\tau) - \int_0^\tau \theta(X_s) ds \right). \quad (10.3.7)$$

Pour  $\psi = 0$  et  $\theta = -1$ ,  $u(x)$  est égal à l'espérance de  $\tau$ , partant de  $x$ . Pour  $\theta = 0$  et  $\psi$  l'indicatrice d'une partie  $A$  du bord  $\partial D$ ,  $u(x)$  est la probabilité de quitter  $D$  par  $A$ . Ainsi, si l'on sait résoudre le problème (10.3.6), on obtient des informations sur le temps et le lieu de sortie de  $D$ . Inversément, en simulant l'expression (10.3.7) par une méthode de Monte-Carlo, on obtient une approximation numérique de la solution du problème (10.3.6).

**Exemple 10.3.2** (Temps de sortie moyen du mouvement Brownien d'une boule). Soit  $K = \{x \in \mathbb{R}^n : \|x\| < R\}$  la boule de rayon  $R$  centrée à l'origine. Soit

$$\tau_K = \inf\{t > 0 : x + B_t \notin K\} \quad (10.3.8)$$

et soit

$$\tau(N) = \tau_K \wedge N. \quad (10.3.9)$$

La fonction  $\varphi(x) = \|x\|^2 1_{\{\|x\| \leq R\}}$  est à support compact et satisfait  $\Delta\varphi(x) = 2n$  pour  $x \in K$ . On peut par ailleurs la prolonger en dehors de  $K$  de manière qu'elle soit lisse et à support compact. En substituant dans la formule de Dynkin, on obtient

$$\begin{aligned} \mathbb{E}^x(\|B_{\tau(N)}\|^2) &= \|x\|^2 + \mathbb{E}^x \left( \int_0^{\tau(N)} \frac{1}{2} \Delta\varphi(B_s) ds \right) \\ &= \|x\|^2 + n\mathbb{E}^x(\tau(N)). \end{aligned} \quad (10.3.10)$$

Comme  $\|B_{\tau(N)}\| \leq R$ , faisant tendre  $N$  vers l'infini, on obtient par convergence dominée

$$\mathbb{E}^x(\tau_K) = \frac{R^2 - \|x\|^2}{n}. \quad (10.3.11)$$

**Exemple 10.3.3** (Récurrence/transience du mouvement Brownien). Soit à nouveau  $K = \{x \in \mathbb{R}^n : \|x\| < R\}$ . Nous considérons maintenant le cas où  $x \notin K$ , et nous voulons déterminer si le mouvement Brownien partant de  $x$  touche  $K$  presque sûrement, on dit alors qu'il est *récurrent*, ou s'il touche  $K$  avec une probabilité strictement inférieure à 1, on dit alors qu'il est *transient*. Comme dans le cas des marches aléatoires, la réponse dépend de la dimension  $n$  de l'espace.

Nous définissons

$$\tau_K = \inf\{t > 0 : x + B_t \in K\}. \quad (10.3.12)$$

Pour  $N \in \mathbb{N}$ , soit  $A_N$  l'anneau

$$A_N = \{x \in \mathbb{R}^n : R < \|x\| < 2^N R\}, \quad (10.3.13)$$

et soit  $\tau$  le temps de première sortie de  $x + B_t$  de  $A_N$ . On a donc

$$\tau = \tau_K \wedge \tau', \quad \tau' = \inf\{t > 0: \|x + B_t\| = 2^N R\}. \quad (10.3.14)$$

Soit enfin

$$p = \mathbb{P}^x\{\tau_K < \tau'\} = \mathbb{P}^x\{\|B_\tau\| = R\} = 1 - \mathbb{P}^x\{\|B_\tau\| = 2^N R\}. \quad (10.3.15)$$

Les solutions à symétrie sphérique de  $\Delta\varphi = 0$  sont de la forme

$$\varphi(x) = \begin{cases} |x| & \text{si } n = 1, \\ -\log\|x\| & \text{si } n = 2, \\ \|x\|^{2-n} & \text{si } n > 2. \end{cases} \quad (10.3.16)$$

Pour un tel  $\varphi$ , la formule de Dynkin donne

$$\mathbb{E}^x(\varphi(B_\tau)) = \varphi(x). \quad (10.3.17)$$

Par ailleurs, on a

$$\mathbb{E}^x(\varphi(B_\tau)) = \varphi(R)p + \varphi(2^N R)(1-p). \quad (10.3.18)$$

En résolvant par rapport à  $p$ , on obtient

$$p = \frac{\varphi(x) - \varphi(2^N R)}{\varphi(R) - \varphi(2^N R)}. \quad (10.3.19)$$

Lorsque  $N \rightarrow \infty$ , on a  $\tau' \rightarrow \infty$ , d'où

$$\mathbb{P}^x\{\tau_K < \infty\} = \lim_{N \rightarrow \infty} \frac{\varphi(x) - \varphi(2^N R)}{\varphi(R) - \varphi(2^N R)}. \quad (10.3.20)$$

Considérons alors séparément les cas  $n = 1$ ,  $n = 2$  et  $n > 2$ .

1. Pour  $n = 1$ , on a

$$\mathbb{P}^x\{\tau_K < \infty\} = \lim_{N \rightarrow \infty} \frac{2^N R - |x|}{2^N R - R} = 1, \quad (10.3.21)$$

donc le mouvement Brownien est récurrent en dimension 1.

2. Pour  $n = 2$ , on a

$$\mathbb{P}^x\{\tau_K < \infty\} = \lim_{N \rightarrow \infty} \frac{\log\|x\| + N \log 2 - \log R}{N \log 2} = 1, \quad (10.3.22)$$

donc le mouvement Brownien est également récurrent en dimension 2.

3. Pour  $n > 2$ , on a

$$\mathbb{P}^x\{\tau_K < \infty\} = \lim_{N \rightarrow \infty} \frac{(2^N R)^{2-n} + \|x\|^{2-n}}{(2^N R)^{2-n} + R^{2-n}} = \left(\frac{R}{\|x\|}\right)^{n-2} < 1. \quad (10.3.23)$$

Le mouvement Brownien est donc transient en dimension  $n > 2$ .

## 10.4 Les équations de Kolmogorov

La seconde classe de liens entre équations différentielles stochastiques et équations aux dérivées partielles est constituée par les équations de Kolmogorov, qui sont des problèmes aux valeurs initiales.

On remarque qu'en dérivant par rapport à  $t$  la formule de Dynkin, dans le cas particulier  $\tau = t$ , on obtient

$$\frac{\partial}{\partial t}(T_t\varphi)(x) = \frac{\partial}{\partial t}\mathbb{E}^x(\varphi(X_t)) = \mathbb{E}^x((L\varphi)(X_t)) = (T_tL\varphi)(x), \quad (10.4.1)$$

que l'on peut abréger sous la forme

$$\frac{d}{dt}T_t = T_tL. \quad (10.4.2)$$

Or nous avons vu dans la remarque 10.2.4 que l'on pouvait aussi écrire formellement  $\frac{d}{dt}T_t = LT_t$ . Par conséquent, les opérateurs  $L$  et  $T_t$  commutent, du moins formellement. Le théorème suivant rend ce point rigoureux.

**Théorème 10.4.1** (Equation de Kolmogorov rétrograde). *Soit  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction deux fois continûment différentiable à support compact.*

1. La fonction

$$u(t, x) = (T_t\varphi)(x) = \mathbb{E}^x(\varphi(X_t)) \quad (10.4.3)$$

satisfait le problème aux valeurs initiales

$$\begin{aligned} \frac{\partial u}{\partial t}(t, x) &= (Lu)(t, x), & t > 0, \quad x \in \mathbb{R}^n, \\ u(0, x) &= \varphi(x), & x \in \mathbb{R}^n. \end{aligned} \quad (10.4.4)$$

2. Si  $w(t, x)$  est une fonction bornée, continûment différentiable en  $t$  et deux fois continûment différentiable en  $x$ , satisfaisant le problème (10.4.4), alors  $w(t, x) = (T_t\varphi)(x)$ .

DÉMONSTRATION.

1. On a  $u(0, x) = (T_0\varphi)(x) = \varphi(x)$  et

$$\begin{aligned} (Lu)(t, x) &= \lim_{h \rightarrow 0^+} \frac{(T_h \circ T_t\varphi)(x) - (T_t\varphi)(x)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{(T_{t+h}\varphi)(x) - (T_t\varphi)(x)}{h} \\ &= \frac{\partial}{\partial t}(T_t\varphi)(x) = \frac{\partial}{\partial t}u(t, x). \end{aligned} \quad (10.4.5)$$

2. Si  $w(t, x)$  satisfait le problème (10.4.4), alors on a

$$\tilde{L}w = 0 \quad \text{où} \quad \tilde{L}w = -\frac{\partial w}{\partial t} + Lw. \quad (10.4.6)$$

Fixons  $(s, x) \in \mathbb{R}_+ \times \mathbb{R}^n$ . Le processus  $Y_t = (s - t, X_t^{0,x})$  admet  $\tilde{L}$  comme générateur. Soit

$$\tau_R = \inf\{t > 0: \|X_t\| \geq R\}. \quad (10.4.7)$$

La formule de Dynkin montre que

$$\mathbb{E}^{s,x}(w(Y_{t \wedge \tau_R})) = w(s, x) + \mathbb{E}^{s,x} \left( \int_0^{t \wedge \tau_R} (Lw)(Y_u) du \right) = w(s, x). \quad (10.4.8)$$

Faisant tendre  $R$  vers l'infini, on obtient

$$w(s, x) = \mathbb{E}^{s,x}(w(Y_t)) \quad \forall t \geq 0. \quad (10.4.9)$$

En particulier, prenant  $t = s$ , on a

$$w(s, x) = \mathbb{E}^{s,x}(w(Y_s)) = \mathbb{E}(w(0, X_s^{0,x})) = \mathbb{E}(\varphi(X_s^{0,x})) = \mathbb{E}^x(\varphi(X_s)). \quad (10.4.10)$$

□

On remarquera que dans le cas du mouvement Brownien, dont le générateur est  $L = \frac{1}{2}\Delta$ , l'équation de Kolmogorov rétrograde (10.4.4) est l'équation de la chaleur.

La linéarité de l'équation de Kolmogorov rétrograde implique qu'il suffit de la résoudre pour une famille complète de conditions initiales  $\varphi$ , pour connaître la solution pour toute condition initiale.

Un premier cas important est celui où l'on connaît toutes les fonctions propres et valeurs propres de  $L$ . Dans ce cas, la solution générale se décompose sur les fonctions propres, avec des coefficients dépendant exponentiellement du temps.

**Exemple 10.4.2** (Mouvement Brownien). Les fonctions propres du générateur  $L = \frac{1}{2}\frac{d^2}{dx^2}$  du mouvement Brownien unidimensionnel sont de la forme  $e^{ikx}$ . Décomposer la solution sur la base de ces fonctions propres revient à résoudre l'équation de la chaleur par transformation de Fourier. On sait que la solution s'écrit

$$u(t, x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-k^2 t/2} \hat{\varphi}(k) e^{ikx} dk, \quad (10.4.11)$$

où  $\hat{\varphi}(k)$  est la transformée de Fourier de la condition initiale.

Un second cas important revient à décomposer formellement la condition initiale sur une "base" de distributions de Dirac. En pratique, cela revient à utiliser la notion de densité de transition.

**Définition 10.4.3** (Densité de transition). *On dit que la diffusion  $\{X_t\}_t$  admet la densité de transition  $p_t(x, y)$ , aussi notée  $p(y, t|x, 0)$ , si*

$$\mathbb{E}^x(\varphi(X_t)) = \int_{\mathbb{R}^n} \varphi(y) p_t(x, y) dy \quad (10.4.12)$$

pour toute fonction mesurable bornée  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ .

Par linéarité, la densité de transition, si elle existe et est lisse, satisfait l'équation de Kolmogorov rétrograde (le générateur  $L$  agissant sur la variable  $x$ ), avec la condition initiale  $p_0(x, y) = \delta(x - y)$ .

**Exemple 10.4.4** (Mouvement Brownien et noyau de la chaleur). Dans le cas du mouvement Brownien unidimensionnel, nous avons vu (c.f. (7.3.2)) que la densité de transition était donnée par

$$p(y, t|x, 0) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}, \quad (10.4.13)$$

qui est appelé le *noyau de la chaleur*. C'est également la valeur de l'intégrale (10.4.11) avec  $\hat{\varphi}(k) = e^{-iky} / \sqrt{2\pi}$ , qui est bien la transformée de Fourier de  $\varphi(x) = \delta(x - y)$ .

L'adjoint du générateur  $L$  est par définition l'opérateur linéaire  $L^*$  tel que

$$\langle L\phi|\psi\rangle = \langle\phi|L^*\psi\rangle \quad (10.4.14)$$

pour tout choix de fonctions  $\phi, \psi : \mathbb{R}^n \rightarrow \mathbb{R}$  deux fois continûment différentiables, avec  $\phi$  à support compact, où  $\langle\cdot|\cdot\rangle$  désigne le produit scalaire usuel de  $L^2$ . En intégrant  $\langle L\phi|\psi\rangle$  deux fois par parties, on obtient

$$(L^*\psi)(y) = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2}{\partial y_i \partial y_j} ((gg^T)_{ij}\psi)(y) - \sum_{i=1}^n \frac{\partial}{\partial y_i} (f_i\psi)(y). \quad (10.4.15)$$

**Théorème 10.4.5** (Equation de Kolmogorov progressive). *Si  $X_t$  possède une densité de transition lisse  $p_t(x, y)$ , alors celle-ci satisfait l'équation*

$$\frac{\partial}{\partial t} p_t(x, y) = L_y^* p_t(x, y), \quad (10.4.16)$$

la notation  $L_y^*$  signifiant que  $L^*$  agit sur la variable  $y$ .

DÉMONSTRATION. La formule de Dynkin, avec  $\tau = t$ , implique

$$\begin{aligned} \int_{\mathbb{R}^n} \varphi(y) p_t(x, y) dy &= \mathbb{E}^x(\varphi(X_t)) \\ &= \varphi(x) + \int_0^t \mathbb{E}^x((L\varphi)(X_s)) ds \\ &= \varphi(x) + \int_0^t \int_{\mathbb{R}^n} (L\varphi)(y) p_s(x, y) dy. \end{aligned} \quad (10.4.17)$$

En dérivant par rapport au temps, et en utilisant (10.4.14), il vient

$$\frac{\partial}{\partial t} \int_{\mathbb{R}^n} \varphi(y) p_t(x, y) dy = \int_{\mathbb{R}^n} (L\varphi)(y) p_t(x, y) dy = \int_{\mathbb{R}^n} \varphi(y) (L_y^* p_t)(x, y) dy, \quad (10.4.18)$$

d'où le résultat.  $\square$

Supposons que la loi  $X_0$  admette une densité  $\rho$  par rapport à la mesure de Lebesgue. Alors  $X_t$  aura une densité donnée par

$$\rho(t, y) = (S_t \rho)(y) := \int_{\mathbb{R}^n} p_t(x, y) \rho(x) dx. \quad (10.4.19)$$

En appliquant l'équation de Kolmogorov progressive (10.4.16), on obtient l'équation de Fokker-Planck

$$\frac{\partial}{\partial t} \rho(t, y) = L_y^* \rho(t, y), \quad (10.4.20)$$

que l'on peut aussi écrire formellement

$$\frac{d}{dt} S_t = L^* S_t. \quad (10.4.21)$$

Le générateur adjoint  $L^*$  est donc le générateur du semi-groupe adjoint  $S_t$ .

**Corollaire 10.4.6.** *Si  $\rho_0(y)$  est la densité d'une mesure de probabilité satisfaisant  $L^* \rho_0 = 0$ , alors  $\rho_0$  est une mesure stationnaire de la diffusion. En d'autres termes, si la loi de  $X_0$  admet la densité  $\rho_0$ , alors  $X_t$  admettra la densité  $\rho_0$  pour tout  $t \geq 0$ .*

## 10.5 Exercices

**Exercice 10.1.** On considère la diffusion définie par l'équation

$$dX_t = -X_t dt + dB_t$$

(processus d'Ornstein–Uhlenbeck).

1. Donner le générateur  $L$  associé et son adjoint  $L^*$ .
2. Soit  $\rho(x) = \pi^{-1/2} e^{-x^2}$ . Calculer  $L^*\rho(x)$ . Que peut-on en conclure?

**Exercice 10.2.** On considère la diffusion définie par l'équation

$$dX_t = X_t dB_t .$$

1. Donner le générateur  $L$  associé.
2. Trouver la solution générale de l'équation  $Lu = 0$ .
3. En déduire  $\mathbb{P}^x\{\tau_a < \tau_b\}$ , où  $\tau_a$  dénote le temps de premier passage de  $X_t$  en  $a$ .  
*Indication:* Il s'agit de calculer  $\mathbb{E}^x(\psi(X_\tau))$ , où  $\tau$  est le temps de première sortie de  $[a, b]$ , et  $\psi(a) = 1$ ,  $\psi(b) = 0$ .

**Exercice 10.3.** On considère plus généralement la diffusion définie par l'équation

$$dX_t = rX_t dt + X_t dB_t , \quad r \in \mathbb{R}$$

(mouvement Brownien géométrique).

1. Calculer son générateur  $L$ .
2. Montrer que si  $r \neq 1/2$ , la solution générale de l'équation  $Lu = 0$  s'écrit

$$u(x) = c_1 x^\gamma + c_2 ,$$

où  $\gamma$  est une fonction de  $r$  qu'on déterminera.

3. On suppose  $r < 1/2$ . Calculer  $\mathbb{P}^x\{\tau_b < \tau_a\}$  pour  $0 < a < x < b$ , puis  $\mathbb{P}^x\{\tau_b < \tau_0\}$  en faisant tendre  $a$  vers 0. On remarquera que si  $X_{t_0} = 0$  alors  $X_t = 0$  pour tout  $t \geq t_0$ . Par conséquent si  $\tau_0 < \tau_b$ , alors  $X_t$  n'atteindra jamais  $b$ . Quelle est la probabilité que cela arrive?
4. On suppose maintenant  $r > 1/2$ .
  - (a) Calculer  $\mathbb{P}^x\{\tau_a < \tau_b\}$  pour  $0 < a < x < b$ , et montrer que cette probabilité tend vers 0 pour tout  $x \in ]a, b[$  lorsque  $a \rightarrow 0_+$ . En conclure que presque sûrement,  $X_t$  n'atteindra jamais 0 dans cette situation.
  - (b) Trouver  $\alpha$  et  $\beta$  tels que  $u(x) = \alpha \log x + \beta$  satisfasse le problème

$$\begin{cases} (Lu)(x) = -1 & \text{si } 0 < x < b , \\ u(x) = 0 & \text{si } x = b . \end{cases}$$

- (c) En déduire  $\mathbb{E}^x(\tau_b)$ .



# Bibliographie et sources d'inspiration

- [Bol07] E. Bolthausen, *Einführung in die stochastik*, notes de cours, Université de Zurich, 2007.
- [Dur96] R. Durrett, *Probability: Theory and examples*, Duxbury Press, Belmont, CA, 1996.
- [Fri75] Avner Friedman, *Stochastic differential equations and applications. Vol. 1*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975, Probability and Mathematical Statistics, Vol. 28.
- [GS72] Ī. Ī. Gihman and A. V. Skorohod, *Stochastic differential equations*, Springer, New York, 1972.
- [McK69] H. P. McKean, Jr., *Stochastic integrals*, Probability and Mathematical Statistics, No. 5, Academic Press, New York, 1969.
- [Øks95] Bernt Øksendal, *Stochastic differential equations*, 4th ed., Universitext, Springer, Berlin, 1995.
- [Res92] Sidney Resnick, *Adventures in stochastic processes*, Birkhäuser Boston Inc., Boston, MA, 1992.