

# Processus Stochastiques et Applications Financières (PSAF)

Pierre ÉTORÉ

ENSIMAG

Année 2022-2023



# Table des matières

<b>1</b>	<b>Rappels d'intégration et de probabilités</b>	<b>5</b>
1.1	Tribus, fonctions mesurables, mesures . . . . .	5
1.2	Intégration contre une mesure . . . . .	6
1.3	Exemples de mesure, mesure de Lebesgue . . . . .	10
1.4	Lien avec l'intégrale de Riemann . . . . .	11
1.5	Rappels de probabilités . . . . .	11
<b>2</b>	<b>Espérance conditionnellement à une tribu</b>	<b>15</b>
2.1	Introduction : notions d'espérance conditionnelle déjà rencontrées . . . . .	15
2.2	Espérance conditionnellement à une tribu . . . . .	17
<b>3</b>	<b>Processus stochastiques, exemple des chaînes de Markov</b>	<b>23</b>
3.1	Processus stochastiques : définitions . . . . .	23
3.2	Chaînes de Markov . . . . .	25
<b>4</b>	<b>Martingales à temps discret</b>	<b>29</b>
4.1	Définition et premières propriétés . . . . .	29
4.2	Intégrale stochastique discrète . . . . .	30
4.3	Théorèmes d'arrêt . . . . .	31
4.4	Résultats de convergence . . . . .	33
4.5	Décomposition de Doob . . . . .	35
<b>5</b>	<b>Modèles financiers à temps discret</b>	<b>37</b>
5.1	Introduction . . . . .	37
5.2	Absence d'opportunité d'arbitrage (AOA) . . . . .	38
5.3	Modèles financiers à temps discret, exemple du modèle Cox-Ross-Rubinstein (CRR), portefeuille autofinancé . . . . .	39
5.4	Mesure de probabilités risque-neutre . . . . .	42
5.5	Premier théorème fondamental . . . . .	43
5.6	Second théorème fondamental . . . . .	45
5.7	Conclusion : valorisation et couverture dans le cas d'un marché viable et complet . . . . .	47
	<b>Bibliographie</b>	<b>49</b>



# Chapitre 1

## Rappels d'intégration et de probabilités

Dans ce chapitre on cherche d'abord à rappeler, pour  $(E, \mathcal{E}, \mu)$  espace mesuré et  $f : E \rightarrow \mathbb{R}$  fonction mesurable (notions qui sont définies ci-après), la définition de l'intégrale de Lebesgue  $\int_E f d\mu$ . Puis on traduit certaines notions d'intégration en langage probabiliste, et on rappelle quelques notions et résultats propres aux probabilités.

### 1.1 Tribus, fonctions mesurables, mesures

Pour  $E$  ensemble on notera de façon standard  $\mathcal{P}(E)$  l'ensemble des parties (ou sous-ensembles) de  $E$ . Pour  $A \subset E$  (i.e.  $A \in \mathcal{P}(E)$ ) on note  $A^c = E \setminus A$  le complémentaire de  $A$  dans  $E$ .

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

- i) On a  $E \in \mathcal{E}$ .
- ii) Pour tout  $A \in \mathcal{E}$  on a  $A^c \in \mathcal{E}$ .
- iii) Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{E}$  on a  $\cup_{n \in \mathbb{N}} A_n \in \mathcal{E}$ .

**Exemple 1.1.1.** 1) La tribu grossière  $\mathcal{E} = \{E, \emptyset\}$ ; c'est la tribu la moins fine qu'on peut mettre sur  $E$ .

2) La tribu la plus fine qu'on peut mettre sur  $E$  c'est  $\mathcal{P}(E)$  (il est bien clair que c'est une tribu; attention au vocabulaire c'est la plus "grosse" au sens de l'inclusion).

3) Soit  $\mathcal{C} \subset \mathcal{P}(E)$ , l'ensemble  $\mathcal{C}$  n'étant pas forcément une tribu. On note  $\sigma(\mathcal{C})$  la tribu engendrée par  $\mathcal{C}$ , i.e. la plus petite tribu qui contient  $\mathcal{C}$  (en ce sens que si  $\mathcal{Y}$  est une tribu t.q.  $\mathcal{C} \subset \mathcal{Y}$  alors  $\sigma(\mathcal{C}) \subset \mathcal{Y}$ ). Sur l'existence de  $\sigma(\mathcal{C})$  cf l'Exercice 1 de la Feuille de TD1.

**Exemple 1.1.2.** Un cas particulier important de l'exemple 1.1.1 - 3) est la tribu des boréliens  $\mathcal{B}(\mathbb{R})$ . Il s'agit de la tribu engendrée par les ouverts de  $\mathbb{R}$ . Notons que  $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{P}(\mathbb{R})$ ; en effet  $\mathcal{B}(\mathbb{R})$  contient toutes sortes d'ensembles (ouverts, fermés, ...) mais pas toutes les parties possibles de  $\mathbb{R}$  (cf par exemple <http://www.lpsm.paris/pageperso/bolley/3M263-poly16-17.pdf>, p20).

**Remarque 1.1.1.** Soit  $\mathcal{E}$  une tribu sur  $E$ . Notons que par passage au complémentaire on a toujours  $\emptyset \in \mathcal{E}$  et que si  $(A_n)_n$  est une suite d'éléments de  $\mathcal{E}$  on a toujours  $\cap_n A_n \in \mathcal{E}$  (respectivement car  $E^c = \emptyset$  et  $\cap_n A_n = (\cup_n A_n^c)^c$ ).

**Remarque 1.1.2.** En fait dans la définition 1.1.1 on peut toujours remplacer de façon équivalente i) par

- i') On a  $\emptyset \in \mathcal{E}$

et iii) par

- iii') Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{E}$  on a  $\cap_{n \in \mathbb{N}} A_n \in \mathcal{E}$ .

Si bien que par exemple une définition équivalente à la définition 1.1.1 est la suivante :

- i') On a  $\emptyset \in \mathcal{E}$ .
- ii) Pour tout  $A \in \mathcal{E}$  on a  $A^c \in \mathcal{E}$ .
- iii') Pour toute suite  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{E}$  on a  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{E}$ .

**Définition 1.1.2.** Soient  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  deux espaces mesurables (i.e. munis chacun d'une tribu). Une fonction  $f : E_1 \rightarrow E_2$  est dite mesurable si  $\forall B \in \mathcal{E}_2$ , on a  $f^{-1}(B) \in \mathcal{E}_1$ .

Si  $(E_i, \mathcal{E}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $i = 1, 2$  on parle de fonction borélienne.

**Exemple 1.1.3.** Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est continue alors elle est borélienne.

En effet pour tout  $U$  ouvert de  $\mathbb{R}$  on a  $f^{-1}(U) \in \mathcal{B}(\mathbb{R})$ . Considérons alors

$$\mathcal{T} = \{ B \in \mathcal{P}(\mathbb{R}) : f^{-1}(B) \in \mathcal{B}(\mathbb{R}) \}.$$

Il est clair que  $\mathcal{T}$  contient les ouverts de  $\mathbb{R}$ . Le résultat est démontré si on montre que  $\mathcal{T}$  est une tribu sur  $\mathbb{R}$ . En effet on aura alors  $\mathcal{B}(\mathbb{R}) \subset \mathcal{T}$ , par définition de  $\mathcal{B}(\mathbb{R})$ . On aura donc  $\forall B \in \mathcal{B}(\mathbb{R})$ ,  $B \in \mathcal{T}$ , i.e.  $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ , ce qui montre bien que  $f$  est borélienne. Or on a

i)  $\emptyset \in \mathcal{T}$  car  $f^{-1}(\emptyset) = \emptyset \in \mathcal{B}(\mathbb{R})$  (par passage au complémentaire on peut vérifier le point i) avec l'ensemble vide au lieu de  $\mathbb{R}$ ).

ii) Pour tout  $B \in \mathcal{T}$  on a  $f^{-1}(B^c) = [f^{-1}(B)]^c \in \mathcal{B}(\mathbb{R})$  (car  $f^{-1}(B)$  appartient à la tribu  $\mathcal{B}(\mathbb{R})$ ). Donc  $B^c \in \mathcal{T}$ .

iii) Pour toute suite  $(B_n)$  d'éléments de  $\mathcal{T}$  on a  $f^{-1}(\bigcup_n B_n) = \bigcup_n f^{-1}(B_n) \in \mathcal{B}(\mathbb{R})$ . Donc  $\bigcup_n B_n \in \mathcal{T}$ .

**Propriété 1.1.1.** Soit  $(f_n)$  une suite de fonctions mesurables de  $(E, \mathcal{E})$  vers  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Alors les fonction  $\inf_n f_n$ ,  $\sup_n f_n$ ,  $\liminf_n f_n$  et  $\limsup_n f_n$  sont mesurables.

*Démonstration.* Cf proposition 2.6.9 dans [8]. □

**Définition 1.1.3.** Soit  $(E, \mathcal{E})$  un espace mesurable. Une mesure sur  $(E, \mathcal{E})$  est une application  $\mu : \mathcal{E} \rightarrow [0, +\infty]$  qui vérifie  $\mu(\emptyset) = 0$  et est  $\sigma$ -additive, i.e. si  $(A_n)$  est une suite d'éléments disjoints de  $\mathcal{E}$  alors  $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ .

**Remarque 1.1.3.** Toute mesure  $\mu$  est automatiquement  $\sigma$ -sous-additive, i.e. pour  $(A_n)$  suite quelconque d'éléments de  $\mathcal{E}$  on a  $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$ .

**Proposition 1.1.1** (Continuité séquentielle croissante de la mesure). Soit  $\mu$  une mesure sur  $(E, \mathcal{E})$ . On a pour toute suite croissante  $(A_n)$  d'éléments de  $\mathcal{E}$  (i.e.  $A_n \subset A_{n+1}$  pour tout  $n$ ),  $\lim_n \mu(A_n) = \mu(\bigcup_n A_n)$ .

*Démonstration.* Cf proposition 2.1.6 dans [8]. □

**Remarque 1.1.4.** Si  $\mu(E) < +\infty$  (mesure positive finie) on a la continuité séquentielle décroissante : pour  $(A_n)$  suite décroissante d'éléments de  $\mathcal{E}$ , on a  $\lim_n \mu(A_n) = \mu(\bigcap_n A_n)$ . La démonstration est laissée au lecteur. Une mesure de probabilités  $\mu$  (i.e.  $\mu(E) = 1$ ) vérifie cette propriété de continuité séquentielle décroissante.

## 1.2 Intégration contre une mesure

Dans ce qui suit  $(E, \mathcal{E})$  un espace mesurable est donné, muni d'une mesure  $\mu$  : on appelle le triplet  $(E, \mathcal{E}, \mu)$  un espace mesuré.

On considère d'abord les fonctions étagées  $f : E \rightarrow [0, \infty]$  i.e. les fonctions de la forme

$$f(x) = \sum_{i=1}^n r_i \mathbf{1}_{A_i}(x) \tag{1.2.1}$$

où  $r_i \in [0, \infty]$  et  $A_i \in \mathcal{E}$  pour tout  $i = 1, \dots, n$ .

Une fonction étagée peut toujours être mise sous la forme dite "canonique" i.e. avec les  $A_i$  disjoints et les valeurs de  $r_i$  distinctes. Dans cette situation il est clair que pour toutes les valeurs possibles  $r_i$  de  $f$  on a  $f^{-1}(r_i) = A_i$ , et la mesurabilité des fonctions étagées en découle naturellement (notons que  $[0, \infty]$  est muni de  $\mathcal{B}(\overline{\mathbb{R}}_+)$ ).

Il est parfois commode de travailler avec les versions canoniques des fonctions étagées, mais ce n'est parfois pas nécessaire, comme dans les définitions qui suivent.

**Définition 1.2.1.** Soit  $f$  fonction étagée (de la forme (1.2.1)). On appelle intégrale de  $f$  contre  $\mu$  et on note  $\int_E f d\mu$  ou  $\int_E f(x)\mu(dx)$ , ou plus rarement  $\int_E f(x)d\mu(x)$  la quantité

$$\sum_{i=1}^n r_i \mu(A_i) \in \overline{\mathbb{R}}_+.$$

Puis on passe aux fonctions mesurables positives.

**Définition 1.2.2.** Soit  $f : E \rightarrow [0, \infty]$  mesurable. On appelle intégrale de  $f$  contre  $\mu$  et on note  $\int_E f d\mu$  la quantité

$$\sup \left\{ \int_E \varphi d\mu, \text{ avec } \varphi \text{ étagée vérifiant } \varphi \leq f \right\} \in \overline{\mathbb{R}}_+.$$

La  $\sigma$ -additivité de  $\mu$  permet de montrer le lemme suivant, qui à son tour permet de montrer le théorème fondamental de convergence monotone de Beppo-Levi.

**Lemme 1.2.1.** Soient  $f, g : E \rightarrow [0, \infty]$  mesurables.

i) Pour  $a, b \in \mathbb{R}$ , on a

$$\int_E (af + bg) d\mu = a \int_E f d\mu + b \int_E g d\mu.$$

ii) Si  $f \leq g$  alors

$$\int_E f d\mu \leq \int_E g d\mu.$$

*Démonstration.* Cf proposition 2.7.2 dans [8]. □

**Théorème 1.2.1** (Théorème de convergence monotone de Beppo-Levi). Soit  $(f_n)$  une suite croissante de fonctions mesurables positives ( $f_n : E \rightarrow [0, \infty]$  pour tout  $n$ ). Alors

$$\lim_{n \rightarrow \infty} \int_E f_n d\mu = \int_E [\lim_{n \rightarrow \infty} f_n] d\mu.$$

*Démonstration.* Cf Théorème 2.9.1 dans [8]. □

De Beppo-Levi on tire immédiatement le théorème suivant.

**Théorème 1.2.2** (Lemme de Fatou). Soit  $(f_n)$  une suite de fonctions mesurables positives. Alors

$$\int_E [\liminf_n f_n] d\mu \leq \liminf_n \int_E f_n d\mu.$$

*Démonstration.* Considérons la suite  $(g_n)$  définie par  $g_n(x) = \inf_{k \geq n} f_k(x)$  pour tout  $n \geq 0$ . C'est bien une suite de fonctions mesurables positives, croissante en  $n$  (i.e.  $g_n(x) \leq g_{n+1}(x)$  pour tout  $n$ , pour tout  $x \in E$ ). Par Beppo-Levi et la définition de la  $\liminf$  on a donc

$$\lim_n \int_E g_n d\mu = \int_E [\liminf_n g_n] d\mu = \int_E [\liminf_n f_n] d\mu. \quad (1.2.2)$$

Mais pour tout  $n$  on a  $\int_E g_n d\mu \leq \int_E f_n d\mu$  (en utilisant le lemme 1.2.1). Comme la limite est égale en particulier à la  $\liminf$  quand elle existe et que le passage à la  $\liminf$  conserve les inégalités larges on a

$$\lim_n \int_E g_n d\mu = \liminf_n \int_E g_n d\mu \leq \liminf_n \int_E f_n d\mu.$$

On conclut par (1.2.2). □

**Définition 1.2.3.** Soit  $P(x)$  une propriété dépendant de  $x \in E$ . On dit que  $P(x)$  est vraie presque partout (p.p.), ou pour presque tout  $x$ , si l'ensemble  $\{x \in E : P(x) \text{ n'est pas vraie}\}$  est négligeable, i.e. inclus dans  $B \in \mathcal{E}$  avec  $\mu(B) = 0$ .

**Remarque 1.2.1.** Notons que dans la définition il est clair que  $\forall x \in B^c$  on a  $P(x)$  vraie.

**Remarque 1.2.2.** Parfois on dit "μ-p.p." si on veut insister sur la mesure utilisée dans la définition 1.2.3 (par exemple quand on travaille avec plusieurs mesures sur  $(E, \mathcal{E})$ ). Il y a plusieurs variantes de la définition du "presque partout". On en verra une dans la définition 1.5.2 à venir de la notion de "presque sûr". Le presque sûr est une version probabiliste du presque partout.

**Proposition 1.2.1.** i) Pour  $A \in \mathcal{E}$  on a

$$\int_E (+\infty \mathbf{1}_A) d\mu = \begin{cases} 0 & \text{si } \mu(A) = 0 \\ +\infty & \text{si } \mu(A) > 0 \end{cases}$$

ii) Pour  $f : E \rightarrow [0, \infty]$  on a  $\int_E f d\mu = 0$  si et seulement si  $f(x) = 0$  p.p. De plus si  $\int_E f d\mu < \infty$  alors  $f(x) < \infty$  p.p.

*Démonstration.* i) On voit  $(+\infty)\mathbf{1}_A$  comme limite de la suite croissante  $(n\mathbf{1}_A)_{n \in \mathbb{N}}$ . Notons qu'on a pour tout  $n$  que  $\int_E (n\mathbf{1}_A) d\mu = n\mu(A)$ . Or par Beppo-Levi il vient

$$\int_E (+\infty \mathbf{1}_A) d\mu = \lim_{n \rightarrow \infty} \int_E (n\mathbf{1}_A) d\mu = \lim_{n \rightarrow \infty} n\mu(A) = \begin{cases} 0 & \text{si } \mu(A) = 0 \\ +\infty & \text{si } \mu(A) > 0 \end{cases}$$

ii) Pour le premier point cf Feuille de TD 1 Exercice 3. Montrons que  $\int_E f d\mu < \infty$  implique  $f(x) < \infty$  p.p. Il revient au même de montrer que  $\mu(\{x \in E : f(x) = +\infty\}) > 0$  implique  $\int_E f d\mu = +\infty$ . Notons que par mesurabilité de  $f$  l'ensemble  $A := \{x \in E : f(x) = +\infty\}$  est dans  $\mathcal{E}$ . On a  $f = f(\mathbf{1}_A + \mathbf{1}_{A^c}) \geq +\infty \mathbf{1}_A$ . D'où  $\int_E f d\mu \geq \int_E (+\infty \mathbf{1}_A) d\mu = \infty$  en utilisant le i).  $\square$

Pour toute fonction à valeurs réelles  $f$  on rappelle qu'on note  $f_+$  et  $f_-$  les fonctions définies respectivement par  $x \mapsto \max(f(x), 0)$  et  $x \mapsto \max(-f(x), 0)$  (ce sont les parties positive et négative de la fonction  $f$ ).

**Définition 1.2.4.** Soit  $f : E \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$  mesurable ( $\overline{\mathbb{R}}$  est muni de  $\mathcal{B}(\overline{\mathbb{R}})$ ). On a la décomposition  $f = f_+ - f_-$ .

On dit que  $f$  est intégrable si  $\int_E f_+ d\mu < +\infty$  et  $\int_E f_- d\mu < +\infty$  (ce qui équivaut à  $\int_E |f| d\mu < +\infty$ ). On note alors

$$\int_E f d\mu := \int_E f_+ d\mu - \int_E f_- d\mu.$$

**Proposition 1.2.2.** Soit  $f : E \rightarrow \overline{\mathbb{R}}$  mesurable. S'il existe  $g : E \rightarrow \overline{\mathbb{R}}_+$  mesurable avec  $\int_E g d\mu < \infty$  et  $|f| \leq g$  μ-p.p. alors  $f$  est intégrable.

*Démonstration.* Il existe  $N \in \mathcal{E}$  négligeable t.q.  $\forall x \in N^c$  on a  $|f(x)| \leq g(x)$ . Donc

$$|f| = |f|(\mathbf{1}_{N^c} + \mathbf{1}_N) \leq g\mathbf{1}_{N^c} + (+\infty)\mathbf{1}_N \leq g + (+\infty)\mathbf{1}_N.$$

Donc

$$\int_E |f| d\mu \leq \int_E (g + (+\infty)\mathbf{1}_N) d\mu = \int_E g d\mu + \int_E \infty \mathbf{1}_N d\mu,$$

ce qui permet de conclure car  $\int_E g d\mu < \infty$  par hypothèse et on a  $\int_E \infty \mathbf{1}_N d\mu = 0$  par la proposition 1.2.1, point 1).  $\square$

**Théorème 1.2.3.** 1) (Fatou pour les fonctions réelles). Soit  $(f_n)$  une suite de fonctions mesurables avec  $f_n(x) \geq m(x)$  pour μ-p.t.  $x \in E$ , pour tout  $n$ , avec  $m$  mesurable et intégrable. Alors

$$\int_E [\liminf_n f_n] d\mu \leq \liminf_n \int_E f_n d\mu.$$

1') Soit  $(f_n)$  une suite de fonctions mesurables avec  $f_n(x) \leq M(x)$  pour μ-p.t.  $x \in E$ , pour tout  $n$ , avec  $M$  mesurable et intégrable. Alors

$$\int_E [\limsup_n f_n] d\mu \geq \limsup_n \int_E f_n d\mu.$$



- 2) (Théorème de convergence dominée de Lebesgue). Soit  $(f_n)$  une suite de fonctions mesurables avec :
- i) Pour  $\mu$ -p.t.  $x \in E$ ,  $f_n(x) \rightarrow f(x)$  quand  $n \rightarrow \infty$
  - ii) Il existe  $h$  mesurable positive et intégrable telle que pour tout  $n$ , on a  $|f_n(x)| \leq h(x)$   $\mu$ -p.p.
- Alors  $f$  est intégrable et

$$\lim_{n \rightarrow \infty} \int_E f_n d\mu = \int_E f d\mu.$$

*Démonstration.* Pour alléger les écritures on note  $\int$  pour  $\int_E$  dans toute la preuve.

1) Comme  $f_n - m$  est positive pour tout  $n$ , on applique Fatou pour les fonctions positives et on obtient

$$\int \liminf f_n d\mu - \int m d\mu = \int \liminf [f_n - m] d\mu \leq \liminf \int [f_n - m] d\mu = \liminf \int f_n d\mu - \liminf \int m d\mu.$$

Comme  $\liminf \int m d\mu = \int m d\mu$  on a le résultat voulu.

1') La suite  $(-f_n)$  vérifie  $-f_n \geq -M$  p.p. pour tout  $n$  avec  $-M$  mesurable et intégrable (car  $\int |-M| d\mu = \int |M| d\mu < +\infty$ ). Donc en utilisant 1) et  $\liminf_n (-a_n) = -\limsup_n a_n$  on a

$$-\int \limsup f_n d\mu = \int \liminf (-f_n) d\mu \leq \liminf \int (-f_n) d\mu = -\limsup \int f_n d\mu,$$

d'où le résultat en multipliant l'inégalité par  $-1$ .

2) Pour tout  $n$  il existe  $N_n \in \mathcal{E}$  négligeable t.q.  $|f_n(x)| \leq h(x)$  pour tout  $x \in N_n^c$ . Par ailleurs il existe  $N \in \mathcal{E}$  négligeable t.q.  $f_n(x) \rightarrow f(x)$ , quand  $n \rightarrow \infty$ , pour tout  $x \in N^c$ . Considérons  $\tilde{N} = N \cup (\cup_n N_n) \in \mathcal{E}$ . Par  $\sigma$ -sous-additivité de  $\mu$  on peut voir que  $\mu(\tilde{N}) = 0$ . Pour tout  $x \in \tilde{N}^c = N^c \cap (\cap_n N_n^c)$  on a :

i)  $f_n(x) \rightarrow f(x)$  quand  $n \rightarrow \infty$  (car  $x \in N^c$ ).

ii)  $|f_n(x)| \leq h(x)$  pour tout  $n$  (car  $x \in N_n^c$  pour tout  $n$ ).

Donc, comme le passage à la limite conserve les inégalités larges on a  $\lim_{n \rightarrow \infty} |f_n(x)| = |f(x)| \leq h(x)$ .

Ceci montre que  $|f| \leq h$  p.p. puisque  $\tilde{N}$  est négligeable. Donc  $f$  est intégrable par la proposition 1.2.2.

Par ailleurs on a  $-h \leq f_n \leq h$  p.p. Donc par 1) on a

$$\int_E [\liminf_n f_n] d\mu \leq \liminf_n \int_E f_n d\mu$$

et par 1')

$$\int_E [\limsup_n f_n] d\mu \geq \limsup_n \int_E f_n d\mu.$$

Or pour tout  $x \in \tilde{N}^c$  on a  $f(x) = \lim_n f_n(x)$  donc  $f(x) = \liminf_n f_n(x) = \limsup_n f_n(x)$ . En remarquant que par exemple  $\int f d\mu = \int f \mathbf{1}_{\tilde{N}^c} d\mu$  on obtient

$$\limsup_n \int f_n d\mu \leq \int f d\mu \leq \liminf_n \int f_n d\mu,$$

et donc

$$\limsup_n \int f_n d\mu = \liminf_n \int f_n d\mu = \lim_n \int f_n d\mu = \int f d\mu.$$

Ceci achève la preuve. □

Le théorème de convergence dominée a par exemple pour conséquence le résultat suivant.

**Théorème 1.2.4** (Dérivation sous le signe somme). Soit  $I$  un intervalle de  $\mathbb{R}$  non réduit à un point et  $f : E \times I \rightarrow \mathbb{R}$  vérifiant :

- i) Il existe  $t_0 \in I$  t.q.  $f(\cdot, t_0) : x \mapsto f(x, t_0)$  est intégrable.
- ii) Pour tout  $t \in I$  la fonction  $f(\cdot, t)$  est mesurable.
- iii) Pour tout  $t \in I$ , la dérivée partielle  $\frac{\partial f}{\partial t}$  existe et  $|\frac{\partial f}{\partial t}(x, t)| \leq h(x)$   $\mu$ -p.p. avec  $h$  positive mesurable intégrable.

Alors pour tout  $t \in I$ , les fonctions  $f(\cdot, t)$  et  $\frac{\partial f}{\partial t}(\cdot, t)$  sont intégrables et la fonction

$$F(t) = \int_E f(x, t) \mu(dx)$$

est dérivable avec

$$\forall t \in I, \quad \frac{d}{dt} F(t) = \int_E \frac{\partial f}{\partial t}(x, t) \mu(dx).$$

*Démonstration.* Cf Feuille de TD 2, Exercice 2. □

On aura parfois besoin du résultat suivant.

**Théorème 1.2.5** (Théorème de Fubini). *Soit  $(F, \mathcal{F}, \nu)$  un autre espace mesuré. On note  $\mu \otimes \nu$  la mesure produit définie sur  $(E \times F, \mathcal{E} \otimes \mathcal{F})$  (ici  $\mathcal{E} \otimes \mathcal{F}$  désigne la plus petite tribu sur  $E \times F$  contenant toutes les parties du type  $A \times B$  avec  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ ). Cette mesure vérifie*

$$\forall A \in \mathcal{E}, \forall B \in \mathcal{F}, \quad \mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

Soit  $f : E \times F \rightarrow \overline{\mathbb{R}}$  mesurable qui est, soit à valeurs dans  $[0, \infty] = \overline{\mathbb{R}}_+$ , soit intégrable. Alors,

$$\int_{E \times F} f d(\mu \otimes \nu) = \int_E \left( \int_F f(x, y) \nu(dy) \right) \mu(dx) = \int_F \left( \int_E f(x, y) \mu(dx) \right) \nu(dy).$$

*Démonstration.* Cf théorème 2.24.2 dans [8]. □

### 1.3 Exemples de mesure, mesure de Lebesgue

On considère  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

C'est l'objet de l'Exercice 2 de la Feuille de TD 1 de voir que  $\delta_{x_0}$  définie par  $\delta_{x_0}(A) = \mathbf{1}_A(x_0)$ ,  $A \in \mathcal{B}(\mathbb{R})$  est une mesure sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  (pour  $x_0 \in \mathbb{R}$ ). Cela nous fournit un premier exemple.

Un deuxième exemple, en quelque sorte "orthogonal" au premier, est celui de la mesure de Lebesgue.

**Théorème 1.3.1.** *Il existe une unique mesure  $\lambda$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , appelée "mesure de Lebesgue", telle que  $\forall I \in \mathcal{B}(\mathbb{R})$  intervalle (i.e.  $I = [a, b]$ ,  $I = ]a, b[$ ,  $I = (a, b)$  ou  $I = ]a, b]$ ) on a  $\lambda(I) = |I|$  (où  $|I| = b - a$  est la "longueur" de  $I$ ).*

*Démonstration.* Pour une preuve complète cf par exemple <http://www.lpsm.paris/pageperso/bolley/3M263-poly16-17.pdf>, p32. La preuve est non triviale, il faut définir  $\lambda$  sur l'ensemble des intervalles de  $\mathbb{R}$ , avec  $\lambda(I) = |I|$  pour  $I$  intervalle. Puis il faut étendre  $\lambda$  en  $\lambda$  par le théorème de Carathéodory. Le théorème de classe monotone intervient pour l'unicité. □

Notons qu'en particulier

$$\lambda(\{a\}) = a - a = 0 \quad \forall a \in \mathbb{R}. \tag{1.3.1}$$

La mesure de Lebesgue attribue la masse nulle aux singletons.

**Définition 1.3.1.** Soient  $\mu$  et  $\nu$  deux mesures sur l'espace mesurable  $(E, \mathcal{E})$ . On dit que  $\nu$  est absolument continue par rapport à  $\mu$ , et on note  $\nu \ll \mu$ , si pour tout  $N \in \mathcal{E}$  t.q.  $\mu(N) = 0$  on a  $\nu(N) = 0$ .

**Théorème 1.3.2** (Théorème de Radon-Nikodym). *Soient  $\mu$  et  $\nu$  deux mesures finies sur  $(E, \mathcal{E})$ . Il y a équivalence entre*

i) On a  $\nu \ll \mu$

ii) Il existe  $f : E \rightarrow [0, \infty]$  mesurable et intégrable t.q.  $\nu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{E}$ .

*Démonstration.* Cf Théorème 2.35.3 dans [8]. □

## 1.4 Lien avec l'intégrale de Riemann

**Théorème 1.4.1.** *Si  $f : [a, b] \rightarrow \mathbb{R}$  est Riemann intégrable et mesurable alors elle est Lebesgue intégrable et on a*

$$\int_a^b f(x)dx = \int_a^b f d\lambda.$$

*Démonstration.* Cf Théorème 2.10.2 dans [8]. □

Cependant l'intégrale de Lebesgue est plus souple à manipuler et permet de manipuler des fonctions qui ne sont pas Riemann intégrables. On a le fameux exemple suivant.

**Exemple 1.4.1.** On considère le problème de calculer  $\int_0^1 \mathbf{1}_{\mathbb{Q}} d\lambda$  (noter que  $\mathbf{1}_{\mathbb{Q}}$  est mesurable et dominée par 1, qui est une fonction intégrable sur  $[0, 1]$ ; on sait déjà que  $\mathbf{1}_{\mathbb{Q}}$  est Lebesgue intégrable).

La fonction  $\mathbf{1}_{\mathbb{Q}}$  n'est pas continue, pas même continue par morceaux. En fait on peut montrer qu'elle n'est pas Riemann intégrable (en montrant que les sommes de Riemann inférieures et supérieures sont différentes).

Mais on peut calculer  $\int_0^1 \mathbf{1}_{\mathbb{Q}} d\lambda$  directement en utilisant les propriétés de l'intégrale de Lebesgue et de la mesure de Lebesgue. On a

$$\int_0^1 \mathbf{1}_{\mathbb{Q}} d\lambda = \lambda(\mathbb{Q} \cap [0, 1]) = \lambda\left[\bigcup_{q \in \mathbb{Q} \cap [0, 1]} \{q\}\right] = \sum_{q \in \mathbb{Q} \cap [0, 1]} \lambda(\{q\}) = 0,$$

on on a utilisé (1.3.1) et la  $\sigma$ -additivité de  $\lambda$ , grâce au fait que  $\mathbb{Q} \cap [0, 1]$  est dénombrable.

## 1.5 Rappels de probabilités

Un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  est donné. C'est un espace mesuré avec  $\mathbb{P}(\Omega) = 1$ .

**Définition 1.5.1.** Une variable aléatoire (v.a.) définie sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , à valeurs dans  $E$  (muni d'une tribu  $\mathcal{E}$ ), c'est une application  $X : \Omega \rightarrow E$ , mesurable de  $(\Omega, \mathcal{F})$  vers  $(E, \mathcal{E})$ .

Pour  $X$  v.a. à valeurs dans  $(E, \mathcal{E})$  et  $B \in \mathcal{E}$  on note

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

On note  $\mathbb{P}_X$  la mesure image de  $\mathbb{P}$  par  $X$ , définie sur  $(E, \mathcal{E})$  par

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

La mesure  $\mathbb{P}_X$  est appelée la loi de  $X$ . Notons que c'est à son tour une mesure de probabilités (sur  $(E, \mathcal{E})$ ).

Soit  $f : E \rightarrow \mathbb{R}$  t.q.  $\int_{\Omega} |f(X(\omega))| \mathbb{P}(d\omega) < +\infty$ . On note

$$\mathbb{E}(f(X)) = \int_{\Omega} f(X(\omega)) \mathbb{P}(d\omega) = \int_{\Omega} f(X) d\mathbb{P} = \int_E f(x) \mathbb{P}_X(dx) = \int_E f d\mathbb{P}_X.$$

Notons qu'on a utilisé ici le théorème de transfert.

On note  $\sigma(X)$  la plus petite tribu sur  $\Omega$  qui rend mesurable  $X$ , i.e.  $\sigma(X) \subset \mathcal{F}$ ,  $X$  est mesurable de  $(\Omega, \sigma(X))$  vers  $(E, \mathcal{E})$ , et  $\sigma(X)$  est le plus petite sous-tribu de  $\mathcal{F}$  qui réalise cela (cf Feuille de TD 2 Exercice 3).

**Définition 1.5.2.** On dit qu'une propriété  $P(\omega)$  est vraie presque sûrement (p.s.), ou pour presque tout  $\omega$ , si il existe  $A \in \mathcal{F}$  avec  $\mathbb{P}(A) = 0$  et  $P(\omega)$  vraie pour tout  $\omega \in A^c$ .

**Remarque 1.5.1.** Remarquer la légère différence entre cette définition et la définition 1.2.3 donnée précédemment. En fait on peut montrer que ce sont deux façons équivalentes de formuler les choses.

On dit par exemple que  $(X_n)$  suite de v.a. définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$  converge p.s. vers  $X$  définie sur le même espace (et à valeurs dans le même espace), et on note

$$X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X,$$

si et seulement si il existe  $A \in \mathcal{F}$  t.q.  $\mathbb{P}(A) = 0$  et

$$\forall \omega \in A^c, \quad X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega).$$

**Remarque 1.5.2.** En fait quand  $(X_n)$  converge p.s. vers  $X$  on peut montrer que l'évènement  $\{X_n \rightarrow X\}$  est dans  $\mathcal{F}$  (et est de probabilité 1). Sur ce point cf [2] Chap. 16 Section 4.

On a des version probabilistes des théorèmes de convergence vus à la section 1.2.

**Théorème 1.5.1.** 1) Soit  $(X_n)$  une suite de v.a. positives tendant en croissant vers  $X$ . Alors

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow \infty} X_n).$$

2) Soit  $(X_n)$  suite de v.a. positives on a

$$\mathbb{E}(\liminf_n X_n) \leq \liminf_n \mathbb{E}(X_n).$$

3) Si  $X_n \rightarrow X$  p.s. quand  $n \rightarrow \infty$ , et s'il existe  $Y$  v.a. positive intégrable (i.e. dans  $L^1(\mathbb{P})$ , i.e.  $\mathbb{E}|Y| < \infty$ ) t.q.  $|X_n| \leq Y$  p.s., alors

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n \rightarrow \infty} X_n).$$

On rappelle trois autres modes de convergence des v.a.

La convergence en loi c'est quand  $\mathbb{P}_{X_n} \rightarrow \mathbb{P}_X$  en un certain sens.

**Remarque 1.5.3.** Notons que les  $X_n$ 's et  $X$  sont à valeurs dans le même espace  $(E, \mathcal{E})$  mais que dans la convergence en loi ces v.a. pourraient être définies sur des espaces de probabilités différents; c'est à dire que  $X$  est définie sur  $(\Omega, \mathcal{F}, \mathbb{P})$  et chaque  $X_n$  est définie sur un certain  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ .

En fait on dit que  $(\mathbb{P}_{X_n})$  converge étroitement vers  $\mathbb{P}_X$ , et on note  $\mathbb{P}_{X_n} \rightarrow \mathbb{P}_X$ ,  $n \rightarrow \infty$ , si et seulement si pour toute fonction  $f$  continue bornée on a  $\int_E f d\mathbb{P}_{X_n} \rightarrow \int_E f d\mathbb{P}_X$  quand  $n \rightarrow \infty$ .

**Remarque 1.5.4.** Ceci s'écrit  $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ ,  $\forall f \in C_b(E)$  si toutes les v.a. sont définies sur le même espace de probabilités.

Cela donne lieu à une première définition.

**Définition 1.5.3.** Soit  $(X_n)$  suite de v.a. à valeurs dans  $(E, \mathcal{E})$  (chaque  $X_n$  est définie sur  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ ). Pour tout  $n$  on note  $\mathbb{P}_{X_n}$  la loi de  $X_n$  sur  $(E, \mathcal{E})$ . Soit  $X$  v.a. à valeurs dans  $(E, \mathcal{E})$ , définie sur  $(\Omega, \mathcal{F}, \mathbb{P})$ .

On dit que  $(X_n)$  converge en loi vers  $X$ , et on note  $X_n \xrightarrow{\mathcal{L}} X$ ,  $n \rightarrow \infty$  si  $\mathbb{P}_{X_n} \rightarrow \mathbb{P}_X$ ,  $n \rightarrow \infty$ .

Pour les variables aléatoires réelles (v.a.r.) on utilise souvent une définition équivalente (cf [2] Théorème 16.8.1) utilisant la notion de fonction de répartition (f.d.r.). On rappelle que pour  $X$  v.a.r. on note  $F_X$  sa f.d.r., définie par  $F_X(x) = \mathbb{P}(X \leq x)$ ,  $x \in \mathbb{R}$ .

**Définition 1.5.4.** Soient  $(X_n)$  suite de v.a.r et  $X$  v.a.r. On dit que  $X_n \xrightarrow{\mathcal{L}} X$ ,  $n \rightarrow \infty$ , si  $F_{X_n}(x) \rightarrow F_X(x)$  en tout point de continuité de  $F_X$ .

On définit maintenant deux modes de convergence supplémentaires.

**Définition 1.5.5.** Soient  $(X_n)$  suite de v.a. et  $X$  v.a. définies  $(\Omega, \mathcal{F}, \mathbb{P})$ .

i) On dit que  $(X_n)$  converge en probabilités vers  $X$  et on note

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X,$$

si

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

ii) On dit que  $(X_n)$  converge en moyenne  $L^p$  vers  $X$  et on note

$$X_n \xrightarrow[n \rightarrow \infty]{L^p} X$$

si  $\mathbb{E}|X_n - X|^p \rightarrow 0$ , quand  $n \rightarrow \infty$ .

Il y a des implications entre les modes de convergence. On a :

- i) La convergence  $L^p$  implique la convergence en probabilités.
- ii) La convergence p.s. implique la convergence en probabilités.
- iii) La convergence en probabilités implique la convergence en loi (cf Feuille de TD 2, Exercice 4).

On a des contre-exemples pour chacune des implications réciproques (cf [2] Chap. 16).

Il y a des petites subtilités comme par exemple : si  $(X_n)$  converge en probabilités vers  $X$ , il existe une sous-suite  $(X_{n_k})$  extraite de  $(X_n)$  qui converge p.s. vers  $X$  (cf [2] Théorème 16.4.5).

Ces modes de convergence interviennent dans la loi des grands nombres et le Théorème Central Limite (TCL) qu'on rappelle ci-dessous.

**Théorème 1.5.2** (Loi forte des grands nombres). *Soit  $(X_n)$  suite de v.a. dans  $L^1$ , indépendantes et identiquement distribuées (i.i.d.). Alors*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(X_1).$$

**Théorème 1.5.3** (TCL). *Soit  $(X_n)$  suite i.i.d. de v.a.r. dans  $L^2$ . Posons  $\sigma^2 = \text{Var}(X_1)$ . Alors*

$$\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

où  $\mathcal{N}(0, 1)$  désigne la loi normale centrée réduite.



## Chapitre 2

# Espérance conditionnellement à une tribu

Dans ce chapitre on rappelle les notions d'espérance conditionnelle déjà rencontrées (en 1A de l'ENSI-MAG). On introduit une notion d'espérance conditionnelle plus générale : l'espérance conditionnellement à une tribu. Cette nouvelle notion englobe les précédentes, en les plaçant dans un cadre commun.

### 2.1 Introduction : notions d'espérance conditionnelle déjà rencontrées

Dans tout ce chapitre un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  est donné et les variables aléatoires rencontrées sont définies sur cet espace de probabilités.

#### 2.1.1 Espérance conditionnellement à un évènement

Soit  $A \in \mathcal{F}$  avec  $\mathbb{P}(A) > 0$ . On note  $\mathbb{P}(\cdot | A)$  la mesure de probabilité définie sur  $(\Omega, \mathcal{F})$  par

$$\forall \Gamma \in \mathcal{F}, \quad \mathbb{P}(\Gamma | A) = \frac{\mathbb{P}(\Gamma \cap A)}{\mathbb{P}(A)}.$$

C'est une mesure de probabilités sur  $(\Omega, \mathcal{F})$  au même titre que  $\mathbb{P}$ . Elle a la particularité d'attribuer la masse 1 à l'évènement  $A$  et la masse 0 à tout  $\Gamma \in \mathcal{F}$  t.q.  $\Gamma \cap A = \emptyset$ . On l'appelle la mesure de probabilités sachant  $A$ .

Soit  $X$  v.a. On définit l'espérance de  $X$  sachant l'évènement  $A$  comme l'intégrale de  $X$  contre  $\mathbb{P}(\cdot | A)$ . On a

$$\mathbb{E}_A(X) = \mathbb{E}(X|A) := \int_{\Omega} X(\omega) \mathbb{P}(d\omega | A) = \frac{\mathbb{E}(\mathbf{1}_A X)}{\mathbb{P}(A)} = \frac{\int_A X(\omega) \mathbb{P}(d\omega)}{\mathbb{P}(A)}$$

(les deux premiers termes de l'égalité sont parmi les notations possibles pour l'espérance de  $X$  sachant  $A$ ).

#### 2.1.2 Espérance d'une variable conditionnellement à une autre : cas discret

Soient  $X$  et  $Y$  deux variables aléatoires, dont  $Y$  discrète. Pour faciliter l'exposé des choses on va aussi supposer que  $X$  est discrète (ça ne serait pas indispensable dans cette sous-section, mais sera utile en fin de chapitre).

Plus précisément :  $X$  et  $Y$  sont à valeurs respectivement dans  $E$  et  $E'$  dénombrables et on a  $\mathbb{P}(X = x) > 0, \forall x \in E$ , et  $\mathbb{P}(Y = y) > 0, \forall y \in E'$ .

On souhaite définir l'espérance de  $X$  sachant  $Y$ . Pour tout  $y \in E'$  on définit

$$\mathbb{E}(X|Y = y) := \mathbb{E}(X | \{Y = y\}) = \frac{\mathbb{E}[X \mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}.$$

Ici on a utilisé simplement la définition vue à la sous-section 2.1.1 de l'espérance de  $X$  conditionnellement à l'événement  $\{Y = y\}$ . Cela a un sens puisque  $\mathbb{P}(Y = y) > 0$  pour tout  $y \in E'$ .

En notant ensuite  $\varphi(y) = \mathbb{E}(X|Y = y)$ ,  $y \in E'$ , on définit la v.a.

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

Notons que comme  $\varphi$  est mesurable (c'est un exercice de le vérifier) la v.a.  $\mathbb{E}(X|Y)$  est  $\sigma(Y)$ -mesurable.

### 2.1.3 Espérance d'une variable aléatoire conditionnellement à une autre : cas à densité bivariée

Soit  $(X, Y)$  à valeurs dans  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  et de loi à densité bivariée  $f_{X,Y}(x, y)$ .

On rappelle que cela signifie que  $\mathbb{P}_{X,Y}(dx, dy) = f_{X,Y}(x, y) dx dy$ , en d'autres termes

$$\forall B \times C \in \mathcal{B}(\mathbb{R}^2), \quad \mathbb{P}(X \in B; Y \in C) = \int \int_{B \times C} f_{X,Y}(x, y) dx dy.$$

On rappelle en outre qu'on note  $f_Y(y) := \int_{\mathbb{R}} f_{X,Y}(x, y) dx$  la densité de la loi marginale de  $Y$ . Le fait que  $\mathbb{P}_Y(dy) = f_Y(y) dy$  peut se retrouver en appliquant le théorème de Fubini :

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}(Y \in B) = \mathbb{P}(X \in \mathbb{R}; Y \in B) = \int \int_{\mathbb{R} \times B} f_{X,Y}(x, y) dx dy = \int_B \left( \int_{\mathbb{R}} f_{X,Y}(x, y) dx \right) dy.$$

Ainsi  $Y$  à une loi à densité par rapport à la mesure de Lebesgue et donc  $\mathbb{P}(Y = y) = 0$  pour tout  $y \in \mathbb{R}$ . Donc la définition de l'espérance de  $X$  sachant  $Y$  donnée à la sous-section 2.1.2 n'a pas de sens. On pose donc

$$\mathbb{E}(X|Y = y) := \int_{\mathbb{R}} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx.$$

De façon analogue à la sous-section 2.1.2 on définit ensuite  $\varphi(y) = \mathbb{E}(X|Y = y)$ ,  $y \in \mathbb{R}$ , puis

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

A nouveau, comme  $\varphi$  est mesurable la v.a.  $\mathbb{E}(X|Y)$  est  $\sigma(Y)$ -mesurable.

**Remarque 2.1.1.** C'est un exercice de montrer que  $\varphi$  est mesurable. En fait si  $f_Y$  est continue et  $f_{X,Y}(x, \cdot)$  est continue pour tout  $x$  on peut voir que  $\varphi$  est continue donc borélienne, en utilisant des résultats du même type que le théorème 1.2.4 (mais pour la continuité). On peut gérer de façon analogue le cas continue par morceaux.

### 2.1.4 Vers une caractérisation commune des espérances conditionnelles vues précédemment

Considérons la v.a.  $\mathbb{E}(X|Y)$  définie à la sous-section 2.1.3. Soit  $A \in \sigma(Y)$  on a  $A = \{Y \in B\}$  pour un certain  $B \in \mathcal{B}(\mathbb{R})$  (cf Feuille de TD 2 Exercice 3). On a en utilisant Fubini

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A \mathbb{E}(X|Y)] &= \mathbb{E}[\mathbf{1}_{Y \in B} \varphi(Y)] \\ &= \int_B \mathbb{E}(X|Y = y) f_Y(y) dy \\ &= \int_B \left( \int_{\mathbb{R}} x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx \right) f_Y(y) dy \\ &= \int \int_{\mathbb{R} \times B} x f_{X,Y}(x, y) dx dy \\ &= \mathbb{E}[\mathbf{1}_{Y \in B} X] \\ &= \mathbb{E}[\mathbf{1}_A X]. \end{aligned}$$



En résumé la v.a.  $\mathbb{E}(X|Y)$  vérifie :

- i) Elle est  $\sigma(Y)$ -mesurable.
- ii) Pour tout  $A$  dans  $\sigma(Y)$  on a  $\mathbb{E}[\mathbf{1}_A \mathbb{E}(X|Y)] = \mathbb{E}[\mathbf{1}_A X]$ .

**Exercice 2.1.1.** Montrer qu'il en est de même dans le cas discret rappelé en sous-section 2.1.2 .

En fait c'est par la caractérisation par les points i) et ii) ci-dessus qu'on définira plus loin  $\mathbb{E}(X|Y)$  (cf sous-section 2.2.3), en utilisant une nouvelle notion : l'espérance conditionnellement à une tribu.

## 2.2 Espérance conditionnellement à une tribu

A partir de maintenant et jusqu'à la fin du chapitre on a  $\mathcal{H} \subset \mathcal{F}$  sous-tribu de  $\mathcal{F}$ . Pour  $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  on cherche à définir une v.a.  $\mathbb{E}(X|\mathcal{H})$  qui vérifie :

- i)  $\mathbb{E}(X|\mathcal{H})$  est  $\mathcal{H}$ -mesurable
- ii) Pour tout  $A \in \mathcal{H}$ , on a  $\mathbb{E}[\mathbf{1}_A \mathbb{E}(X|\mathcal{H})] = \mathbb{E}[\mathbf{1}_A X]$ .

### 2.2.1 Première approche : le cadre $L^2$

On suppose pour commencer que  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$ , i.e.  $\mathbb{E}|X|^2 < +\infty$  (ici  $|\cdot|$  désigne la norme euclidienne dans  $\mathbb{R}^d$  si  $X$  est à valeurs dans  $\mathbb{R}^d$ ).

On peut alors utiliser des éléments de théorie hilbertienne pour définir  $\mathbb{E}(X|\mathcal{H})$ .

On note  $L^2(\Omega, \mathcal{H}, \mathbb{P})$  l'ensemble des variables aléatoires  $\mathcal{H}$ -mesurables et de carré intégrable. On rappelle les résultats suivants :

**Lemme 2.2.1.** *i) L'espace  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  muni du produit scalaire  $(X, Y) \mapsto \mathbb{E}(XY)$  est un espace de Hilbert.*

*ii) Le sous-espace  $L^2(\Omega, \mathcal{H}, \mathbb{P}) \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$  est une sous-espace fermé de  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ .*

*Démonstration.* Cf [5], en particulier le théorème 22.3. □

Pour un élément  $X$  de  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  on peut donc considérer son projeté orthogonal sur  $L^2(\Omega, \mathcal{H}, \mathbb{P})$ , cf par exemple les théorème 22.6 et 22.7 dans [5].

On décide de noter  $\mathbb{E}(X|\mathcal{H})$  ce projeté orthogonal. Par nature  $\mathbb{E}(X|\mathcal{H})$  est  $\mathcal{H}$ -mesurable. De plus par orthogonalité on a

$$\forall A \in \mathcal{H}, \quad \mathbb{E}[\{X - \mathbb{E}(X|\mathcal{H})\} \mathbf{1}_A] = 0$$

c'est à dire

$$\forall A \in \mathcal{H}, \quad \mathbb{E}[\mathbf{1}_A \mathbb{E}(X|\mathcal{H})] = \mathbb{E}[\mathbf{1}_A X]. \tag{2.2.1}$$

La v.a.  $\mathbb{E}(X|\mathcal{H})$  vérifie donc les points i) et ii) évoqués en début de section.

Pour le moment  $\mathbb{E}(X|\mathcal{H})$  peut être pensée comme la meilleure approximation de  $X$  au sens des moindres carrés. De plus sa norme  $L^2$  est toujours inférieure ou égale à celle de  $X$ . En effet

$$\|X\|_{L^2}^2 = \|X - \mathbb{E}(X|\mathcal{H}) + \mathbb{E}(X|\mathcal{H})\|_{L^2}^2 = \|X - \mathbb{E}(X|\mathcal{H})\|_{L^2}^2 + \|\mathbb{E}(X|\mathcal{H})\|_{L^2}^2$$

donc  $\|\mathbb{E}(X|\mathcal{H})\|_{L^2}^2 \leq \|X\|_{L^2}^2$ .

**Remarque 2.2.1.** Ainsi défini l'objet  $\mathbb{E}(X|\mathcal{H})$  est unique à l' "égalité presque sûre près" (cf sous-section suivante pour une définition précise). C'est parce qu'en toute rigueur l'espace  $L^2(\Omega, \mathcal{H}, \mathbb{P})$  par exemple est l'ensemble des classes d'équivalence pour la relation "égalité p.s. près" de v.a.  $\mathcal{H}$ -mesurables et de carré intégrable (cf dans [5] la discussion pp 57-58). C'est par un abus de langage généralisé et admis qu'on le considère comme l'ensemble des variables aléatoires  $\mathcal{H}$ -mesurables et de carré intégrable.

## 2.2.2 Extension au cadre $L^1$ et premières propriétés

Mais on peut aussi définir  $\mathbb{E}(X|\mathcal{H})$  pour  $X \in L^1(\Omega, \mathcal{F}, \mathbb{P}) = L^1$ . On stocke l'information dans le théorème-définition suivant.

**Théorème-Définition 2.2.1.** *Soit  $X$  un vecteur aléatoire, supposé intégrable, i.e.  $\mathbb{E}|X| < +\infty$ .*

*Alors il existe une v.a.  $Y$  vérifiant*

*1)  $Y$  est  $\mathcal{H}$ -mesurable*

*2) Pour tout  $A \in \mathcal{H}$ , on a  $\mathbb{E}(\mathbf{1}_A Y) = \mathbb{E}(\mathbf{1}_A X)$ .*

*Cette v.a.  $Y$  est unique à l'égalité p.s. près et se note  $\mathbb{E}(X|\mathcal{H})$ .*

*Elle est appelée espérance de  $X$  sachant  $\mathcal{H}$  ou espérance de  $X$  conditionnellement à  $\mathcal{H}$ .*

*Démonstration.* L'unicité sera traitée à part dans la proposition 2.2.2 ci-après, où on donnera un sens précis à l'égalité p.s. près. On se concentre donc sur l'existence.

On traite le cas  $X$  v.a. réelle, le cas vectoriel se traitant ensuite composante par composante.

En utilisant la décomposition  $X = X_+ - X_-$  on voit qu'il suffit de traiter le cas  $X \geq 0$ ,  $X \in L^1$ . Les parties positives et négatives ayant été traitées à part on pose

$$\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X_+|\mathcal{H}) - \mathbb{E}(X_-|\mathcal{H})$$

et cet objet a bien les propriétés voulues par linéarité.

Supposons donc  $X \geq 0$ ,  $X \in L^1$ . On pose  $X_n = X \wedge n$ ,  $n \in \mathbb{N}$ . Pour tout  $n$  la v.a.  $X_n$  est bornée donc dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ . On définit donc  $Y_n = \mathbb{E}(X_n|\mathcal{H})$  au sens  $L^2$ .

Le fait que  $X_n \geq 0$  pour tout  $n$  entraîne que  $Y_n \geq 0$  p.s. pour tout  $n$ . En effet : soit  $n$ , la v.a.  $Y_n$  est  $\mathcal{H}$ -mesurable donc  $\{Y < 0\} \in \mathcal{H}$ , donc en utilisant (2.2.1)

$$0 \geq \mathbb{E}[\mathbf{1}_{Y_n < 0} Y_n] = \mathbb{E}[\mathbf{1}_{Y_n < 0} X_n].$$

Mais  $X_n \geq 0$  p.s. donc  $\mathbf{1}_{Y_n < 0} X_n \geq 0$  p.s. et donc  $\mathbb{E}[\mathbf{1}_{Y_n < 0} X_n] \geq 0$  (cela découle du lemme 1.2.1). C'est donc que  $\mathbb{E}[\mathbf{1}_{Y_n < 0} Y_n] = 0$ .

Supposons que  $\mathbb{P}(Y_n < 0) > 0$  alors il existe  $q_0 \in \mathbb{Q}_+^*$  t.q.  $\mathbb{P}(Y_n \leq q_0) > 0$ . En effet, comme  $\{Y_n < 0\} = \cup_{q \in \mathbb{Q}_+^*} \{Y_n \leq q\}$ , si tel n'était pas le cas on aurait  $\mathbb{P}(Y_n < 0) \leq \sum_{q \in \mathbb{Q}_+^*} \mathbb{P}(Y_n \leq q) = 0$ .

On a donc

$$0 = \mathbb{E}[\mathbf{1}_{Y_n \leq q_0} Y_n] + \mathbb{E}[\mathbf{1}_{q_0 < Y_n < 0} Y_n].$$

Mais  $\mathbb{E}[\mathbf{1}_{q_0 < Y_n < 0} Y_n] \leq 0$  et  $\mathbb{E}[\mathbf{1}_{Y_n \leq q_0} Y_n] \leq q_0 \mathbb{P}(Y_n \leq q_0) < 0$ . On arrive à une absurdité, c'est donc que  $\mathbb{P}(Y_n < 0) = 0$ , i.e.  $Y_n \geq 0$  p.s.

Par ailleurs  $X_{n+1} \geq X_n$  donc  $X_{n+1} - X_n \geq 0$  pour tout  $n$ , donc

$$\forall n \in \mathbb{N}, \quad Y_{n+1} - Y_n = \mathbb{E}(X_{n+1} - X_n | \mathcal{H}) \geq 0 \text{ p.s.}$$

par les mêmes arguments que ci-dessus.

La suite  $(Y_n)$  est donc p.s. positive et croissante. Posons donc  $Y = \lim_n Y_n$  (cette limite existe p.s.). Notons que la v.a.  $Y$  est  $\mathcal{H}$ -mesurable par la propriété 1.1.1.

La suite  $(X_n)$  tend en croissant vers  $X$ . Donc par convergence monotone et en utilisant à nouveau (2.2.1) on a pour  $A \in \mathcal{H}$ ,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A X] &= \int_A \lim_n X_n d\mathbb{P} = \lim_n \int_A X_n d\mathbb{P} = \lim_n \mathbb{E}[\mathbf{1}_A X_n] \\ &= \lim_n \mathbb{E}[\mathbf{1}_A Y_n] = \int_A \lim_n Y_n d\mathbb{P} = \mathbb{E}[\mathbf{1}_A Y]. \end{aligned}$$

En résumé :

1)  $Y$  est  $\mathcal{H}$ -mesurable

2) Pour tout  $A \in \mathcal{H}$ , on a  $\mathbb{E}(\mathbf{1}_A Y) = \mathbb{E}(\mathbf{1}_A X)$ . □

**Proposition 2.2.1.** *Dans le théorème-définition 2.2.1 la condition 2) peut être remplacée de façon équivalente par*

*2') Pour toute v.a.r.  $Z$  bornée et  $\mathcal{H}$ -mesurable on a  $\mathbb{E}(ZY) = \mathbb{E}(ZX)$ .*

*En conséquence l'espérance conditionnelle  $\mathbb{E}(X|\mathcal{H})$  vérifie  $\mathbb{E}[Z\mathbb{E}(X|\mathcal{H})] = \mathbb{E}[ZX]$  pour toute v.a.r.  $Z$  bornée et  $\mathcal{H}$ -mesurable.*

*Démonstration.* Pour l'une des implication il suffit de prendre  $Z = \mathbf{1}_A$  pour  $A \in \mathcal{H}$ . Pour l'autre implication il s'agit d'approcher  $Z$  par des v.a. étagées.  $\square$

**Proposition 2.2.2.** *Si  $Y$  et  $Y'$  vérifient les points 1) et 2) du théorème-définition 2.2.1 on a  $Y = Y'$  p.s., d'où l'égalité p.s. près.*

*Démonstration.* Pour tout  $A \in \mathcal{H}$  on a  $\mathbb{E}[\mathbf{1}_A Y] = \mathbb{E}[\mathbf{1}_A Y'] = \mathbb{E}[\mathbf{1}_A X]$  et donc

$$\mathbb{E}[\mathbf{1}_A(Y - Y')] = 0 = \mathbb{E}[\mathbf{1}_A(Y' - Y)], \quad \forall A \in \mathcal{H}. \quad (2.2.2)$$

Pour  $\varepsilon > 0$  définissons  $A_\varepsilon = \{|Y - Y'| \geq \varepsilon\}$ . On a  $A_\varepsilon \in \mathcal{H}$  car  $Y$  et  $Y'$  sont  $\mathcal{H}$ -mesurables et  $(x, y) \mapsto |x - y|$  est continue donc borélienne. On a

$$\mathbb{E}[\mathbf{1}_{A_\varepsilon} |Y - Y'|] = \mathbb{E}[\mathbf{1}_{A_\varepsilon \cap \{Y - Y' > 0\}}(Y - Y') + \mathbf{1}_{A_\varepsilon \cap \{Y - Y' < 0\}}(Y' - Y)].$$

Mais  $A_\varepsilon \cap \{Y - Y' > 0\}$  et  $A_\varepsilon \cap \{Y - Y' < 0\}$  sont à nouveau dans  $\mathcal{H}$  donc le terme de droite dans l'égalité ci-dessus vaut zéro par (2.2.2). Par ailleurs  $\mathbb{E}[\mathbf{1}_{A_\varepsilon} |Y - Y'|] \geq \varepsilon \mathbb{P}(A_\varepsilon)$ , d'où on déduit que  $\mathbb{P}(A_\varepsilon) = 0$  pour tout  $\varepsilon > 0$ .

Mais

$$\begin{aligned} \{Y \neq Y'\} &= \{\omega \in \Omega : \exists q \in \mathbb{Q}_+^* \text{ avec } |Y(\omega) - Y'(\omega)| > q\} = \bigcup_{q \in \mathbb{Q}_+^*} \{\omega \in \Omega : |Y(\omega) - Y'(\omega)| > q\} \\ &= \bigcup_{q \in \mathbb{Q}_+^*} \{|Y - Y'| > q\} \in \mathcal{F}, \end{aligned}$$

et  $\mathbb{P}(\bigcup_{q \in \mathbb{Q}_+^*} \{|Y - Y'| > q\}) \leq \sum_{q \in \mathbb{Q}_+^*} \mathbb{P}(A_q) = 0$ . Donc  $\mathbb{P}(Y \neq Y') = 0$ , i.e.  $Y = Y'$  p.s.  $\square$

**Lemme 2.2.2.** *Pour  $X \in L^1$  on a  $\mathbb{E}(X|\mathcal{H}) \in L^1$ .*

*Démonstration.* Cf Feuille de TD 3 Exercice 2, point 1).  $\square$

On liste maintenant une série de propriétés de l'espérance conditionnellement à une tribu. Dans les propriétés ci-dessous on considère que  $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ .

**Propriété 2.2.1.** *On a (les égalités ou inégalités sont au sens p.s.) :*

- i) Pour  $a, b \in \mathbb{R}$ ,  $\mathbb{E}(aX + bY|\mathcal{H}) = a\mathbb{E}(X|\mathcal{H}) + b\mathbb{E}(Y|\mathcal{H})$  (linéarité de l'espérance conditionnelle).
- ii) Si  $X \geq 0$  alors  $\mathbb{E}(X|\mathcal{H}) \geq 0$ .
- iii)  $\mathbb{E}[\mathbb{E}(X|\mathcal{H})] = \mathbb{E}(X)$ .
- iv) Si  $X$  est  $\mathcal{H}$ -mesurable alors  $\mathbb{E}(X|\mathcal{H}) = X$ .
- v) Si  $Y$  est  $\mathcal{H}$ -mesurable (mais  $X$  v.a. intégrable quelconque) alors  $\mathbb{E}(YX|\mathcal{H}) = Y\mathbb{E}(X|\mathcal{H})$ .
- vi) Si  $X$  est indépendante de  $\mathcal{H}$  (i.e.  $X$  ind. de  $\mathbf{1}_A$ ,  $\forall A \in \mathcal{H}$ ), alors  $\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(X)$ .
- vii) Si  $\mathcal{G}$  est une sous-tribu avec  $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ , alors

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

*Démonstration.* i) Cf [5], théorème 23.4.

ii) En fait nous avons déjà montré cette propriété, à l'intérieur de la preuve du Théorème-Définition 2.2.1, quand nous avons montré que  $X_n \geq 0$  entraîne  $Y_n = \mathbb{E}(X_n|\mathcal{H}) \geq 0$  p.s. (le fait que  $Y_n$  était définie au sens  $L^2$  ne jouait aucun rôle particulier). Il suffit de reproduire la démarche.

iii) Il suffit de prendre  $A = \Omega$  dans le point 2) du théorème-définition 2.2.1.

iv) Cf Feuille de TD 3 Exercice 1, point 1).

v) Cf Feuille de TD 3 Exercice 1, point 2).

vi) Cf Feuille de TD 3 Exercice 1, point 3).

vii) La deuxième partie de l'égalité découle simplement du point iii) car  $\mathbb{E}(X|\mathcal{H})$  est  $\mathcal{H}$ -mesurable, donc  $\mathcal{G}$ -mesurable.

Montrons la première partie. La v.a.  $\mathbb{E}(X|\mathcal{H})$  est bien évidemment  $\mathcal{H}$ -mesurable. Soit  $A \in \mathcal{H}$ , on a  $A \in \mathcal{G}$  puisque  $\mathcal{H} \subset \mathcal{G}$ . Donc on a

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

En résumé :

- 1)  $\mathbb{E}(X|\mathcal{H})$  est  $\mathcal{H}$ -mesurable.  
 2) Pour tout  $A \in \mathcal{H}$  on a  $\mathbb{E}[\mathbf{1}_A \mathbb{E}(X|\mathcal{H})] = \mathbb{E}[\mathbf{1}_A \mathbb{E}(X|\mathcal{G})]$ .  
 Par unicité ceci montre

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G}) | \mathcal{H}] = \mathbb{E}(X|\mathcal{H}) \quad \text{p.s.}$$

□

**Définition 2.2.1.** Pour  $A \in \mathcal{F}$  on note  $\mathbb{P}(A|\mathcal{H})$  la variable aléatoire  $\mathbb{E}(\mathbf{1}_A|\mathcal{H})$ . C'est la probabilité de l'évènement  $A$  sachant la sous-tribu  $\mathcal{H}$ .

### 2.2.3 Espérance d'une variable aléatoire conditionnellement à une autre

**Définition 2.2.2.** Soient  $X$  et  $Y$  définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , avec  $X$  intégrable. On note

$$\mathbb{E}(X|Y) := \mathbb{E}[X | \sigma(Y)]$$

où on rappelle que  $\sigma(Y)$  est la tribu engendrée par  $Y$ .

On a besoin du lemme suivant.

**Lemme 2.2.3** (Lemme de Doob). *Soit  $Y$  v.a. Une variable aléatoire  $Z$  est  $\sigma(Y)$ -mesurable si et seulement si il existe  $\varphi$  borélienne t.q.  $Z = \varphi(Y)$ .*

*Démonstration.* Cf Lemme VI.3.1 dans [1].

□

Du lemme de Doob découle naturellement la proposition suivante.

**Proposition 2.2.3.** *Dans le contexte de la définition 2.2.2 il existe une fonction  $\varphi$  borélienne telle que*

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

Cela nous amène à poser la notation

$$\mathbb{E}(X | Y = y) := \varphi(y), \tag{2.2.3}$$

avec la fonction  $\varphi$  en jeu dans la proposition.

Quel est le lien avec les objets  $\mathbb{E}(X|Y)$  et  $\mathbb{E}(X | Y = y)$  rappelés dans la section 2.1 ?

**Cas à densité bivariée.** Considérons  $\mathbb{E}(X|Y)$  définie dans la sous-section 2.1.3. On a déjà vu que cette v.a. vérifie :

- i) Elle est  $\sigma(Y)$ -mesurable.
- ii) Pour tout  $A$  dans  $\sigma(Y)$  on a  $\mathbb{E}[\mathbf{1}_A \mathbb{E}(X|Y)] = \mathbb{E}[\mathbf{1}_A X]$ .

Donc par unicité c'est  $\mathbb{E}[X|\sigma(Y)]$ , c'est à dire  $\mathbb{E}(X|Y)$  au sens de la nouvelle définition. C'est donc que  $\mathbb{E}(X|Y)$  au sens de la nouvelle définition vaut  $\varphi(Y)$  avec la fonction  $\varphi(y)$  définie dans la sous-section 2.1.3. C'est donc que pour tout  $y \in \mathbb{R}$  la quantité  $\mathbb{E}(X | Y = y)$  au sens de la nouvelle définition vaut

$$\varphi(y) = \int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)} dx.$$

La nouvelle définition aboutit donc à la même valeur pour  $\mathbb{E}(X | Y = y)$  que l'ancienne.

**Cas discret.** Soit  $y \in E'$  (on reprend les notations de la sous-section 2.1.2). Bien sûr  $\{Y = y\} \in \sigma(Y)$  donc on a, en utilisant les nouvelles définitions de  $\mathbb{E}(X|Y)$  et  $\mathbb{E}(X | Y = y)$ ,

$$\mathbb{E}[X \mathbf{1}_{\{Y=y\}}] = \mathbb{E}[\mathbf{1}_{\{Y=y\}} \mathbb{E}(X|Y)] = \sum_{z \in E'} \mathbf{1}_{z=y} \mathbb{E}(X | Y = z) \mathbb{P}(Y = z) = \mathbb{E}(X | Y = y) \mathbb{P}(Y = y)$$

et donc

$$\mathbb{E}(X | Y = y) = \frac{\mathbb{E}[X \mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y = y)}.$$

A nouveau la nouvelle définition aboutit à la même valeur pour  $\mathbb{E}(X | Y = y)$ .

## 2.2.4 Autres propriétés

**Proposition 2.2.4.** Soient  $X$  et  $Y$  définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . Supposons que  $X$  est  $\mathcal{H}$ -mesurable et que  $Y$  est indépendante de  $\mathcal{H}$  (donc de  $X$ ).

Alors pour tout fonction  $f$  borélienne telle que  $f(X, Y) \in L^1$  on a

$$\mathbb{E}[f(X, Y) | \mathcal{H}] = F(X) \quad \text{où} \quad F(x) = \mathbb{E}[f(x, Y)], \quad \forall x \in \mathbb{R}.$$

*Démonstration.* Cf [5], Exercice 23.7, ou [6] proposition 8.2.5. □

**Remarque 2.2.2.** Remarquer le lien avec les points v) et vi) de la propriété 2.2.1 : dans le cas produit ( $f(x, y) = xy$ ) on a simplement  $F(x) = x\mathbb{E}(Y)$ .

**Théorème 2.2.1.** 1) (Convergence monotone) : Si  $(X_n)$  est une suite de v.a. positives qui tend en croissant vers  $X \in L^1$  alors

$$\mathbb{E}(X_n | \mathcal{H}) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(X | \mathcal{H}).$$

2) (Fatou) : Si  $X_n \geq Z$  pour tout  $n$  avec  $Z \in L^1$  et  $\mathbb{E}|X_n| < \infty$  pour tout  $n$  et  $\mathbb{E}|\liminf_n X_n| < \infty$  alors

$$\mathbb{E}(\liminf_n X_n | \mathcal{H}) \leq \liminf_n \mathbb{E}(X_n | \mathcal{H}) \quad \text{p.s.}$$

3) (Convergence dominée) : Si

$$X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X \in L^1$$

et s'il existe  $Z \in L^1$  t.q.  $|X_n| \leq Z$  p.s. pour tout  $n$  alors,

$$\mathbb{E}(X_n | \mathcal{H}) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(X | \mathcal{H}).$$

4) (Jensen) : Si  $X \in L^1$  et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  est mesurable convexe avec  $\varphi(X) \in L^1$  alors

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

*Démonstration.* Cf [5], théorèmes 23.8 et 23.9. □



## Chapitre 3

# Processus stochastiques, exemple des chaînes de Markov

Dans ce chapitre on introduit la notion de processus stochastique, et on donne une série de définitions générales liées à cette notion. On explore ensuite la notion de chaîne de Markov, un exemple particulier de processus stochastique.

### 3.1 Processus stochastiques : définitions

Dans les définitions qui suivent  $(E, \mathcal{E})$  désigne un espace mesurable.

**Définition 3.1.1.** On appelle processus aléatoire (ou stochastique) une famille  $X = (X_\theta)_{\theta \in \Theta}$  de variables aléatoires  $X_\theta$  à valeurs dans  $(E, \mathcal{E})$  et définies sur un même espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  (c'est à dire que  $X_\theta : \Omega \rightarrow E$  est une application mesurable pour tout  $\theta \in \Theta$ ), indexée par  $\theta \in \Theta$  où  $\Theta$  est un ensemble.

Si  $\Theta$  est fini ou dénombrable (ex :  $\Theta = \mathbb{N}, \mathbb{Z}, \dots$ ) on parle de processus à temps discret.

Si  $\Theta$  a la puissance du continu (ex :  $\Theta = \mathbb{R}_+, \mathbb{R}, \mathbb{R}^2, \dots$ ) on parle de processus à temps continu.

L'espace  $E$  est appelé l'espace d'état du processus  $X$  (on dit aussi que le processus  $X$  est à valeurs dans  $E$ ).

**Remarque 3.1.1.** Il y a deux autres façons de voir un processus  $X = (X_\theta)_{\theta \in \Theta}$  défini sur  $(\Omega, \mathcal{F}, \mathbb{P})$  et à valeurs dans  $E$ .

1) On peut voir  $X$  comme une application bivariable :

$$\begin{aligned} X : \Theta \times \Omega &\rightarrow E \\ (\theta, \omega) &\mapsto X_\theta(\omega) \end{aligned}$$

Notons que sans plus d'hypothèse sa mesurabilité de  $\Theta \times \Omega$  vers  $E$  n'est pas claire. Il faudrait déjà mettre une tribu sur  $\Theta$ ... À ce propos voir - par anticipation - la définition des processus progressivement mesurables dans le cours de 3A IF ENSIMAG "Fondements Mathématiques du Calcul Stochastique" (FMCS).

2) On peut voir  $X$  comme une application de  $\Omega$  vers l'espace fonctionnel  $E^\Theta$  (l'ensemble des fonctions de  $\Theta$  vers  $E$ ) :

$$\begin{aligned} X : \Omega &\rightarrow E^\Theta \\ \omega &\mapsto X(\omega) : \Theta \rightarrow E \\ &\quad \theta \mapsto X_\theta(\omega) \end{aligned}$$

Pour  $\omega \in \Omega$  la fonction  $X(\omega)$ , élément de  $E^\Theta$ , est appelée une trajectoire du processus  $X$  (c'est la trajectoire associée à l'aléa  $\omega$ ).

Pour voir  $X$  comme une variable aléatoire à valeurs dans l'espace des trajectoires  $E^\Theta$  il faut mettre une tribu sur  $E^\Theta$  (c'est chose possible, cf à nouveau FMCS en 3A). Ce point de vue est très fécond, c'est une des façons de construire le mouvement brownien (cf FMCS).

Dans ce qui suit, sauf précision contraire, on considère un processus à valeurs dans  $E$  et défini sur un certain espace de probabilités. On suppose que l'addition fait sens pour les éléments de l'ensemble  $\Theta$ , mais dans les deux définitions suivantes on ne suppose pas que c'est un groupe additif, ni qu'il est muni d'une structure d'ordre.

**Définition 3.1.2.** Un processus  $X = (X_\theta)_{\theta \in \Theta}$  est dit stationnaire si pour tout  $n \in \mathbb{N}^*$ , pour tous  $\theta_1, \dots, \theta_n \in \Theta$ , et pour tout  $h \in \Theta$  avec  $\theta_i + h \in \Theta$ ,  $\forall 1 \leq i \leq n$ , les vecteurs  $(X_{\theta_1+h}, \dots, X_{\theta_n+h})'$  et  $(X_{\theta_1}, \dots, X_{\theta_n})'$  ont même loi.

**Définition 3.1.3.** Un processus  $X = (X_\theta)_{\theta \in \Theta}$  est à accroissements stationnaires si la loi de  $X_{t+\theta} - X_t$  ne dépend pas de  $t$ , pour tous  $\theta, t \in \Theta$  t.q.  $t + \theta \in \Theta$ .

Dorénavant on suppose qu'il existe une relation d'ordre sur  $\Theta$  : l'indice  $\theta \in \Theta$  représente le temps (typiquement on a  $\Theta = \mathbb{N}$  ou  $\Theta = \mathbb{R}_+$ ).

**Définition 3.1.4.** Un processus  $X = (X_\theta)_{\theta \in \Theta}$  est dit à accroissements indépendants si pour tous  $\theta_1 < \dots < \theta_k$  dans  $\Theta$  les v.a.  $X_{\theta_2} - X_{\theta_1}, \dots, X_{\theta_k} - X_{\theta_{k-1}}$  sont mutuellement indépendantes.

**Définition 3.1.5.** Une filtration  $(\mathcal{F}_\theta)_{\theta \in \Theta}$  sur un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  c'est un ensemble croissant de sous-tribus  $\mathcal{F}_\theta$  de  $\mathcal{F}$ , i.e. pour tous  $s \leq t$  dans  $\Theta$ , on a  $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ .

**Définition 3.1.6.** On dit qu'un processus  $X = (X_\theta)_{\theta \in \Theta}$  est adapté à une filtration  $(\mathcal{F}_\theta)_{\theta \in \Theta}$  si pour tout  $\theta \in \Theta$  la v.a.  $X_\theta$  est  $\mathcal{F}_\theta$ -mesurable.

**Exemple 3.1.1** (Filtration naturelle). Soit  $X = (X_\theta)_{\theta \in \Theta}$  un processus aléatoire. Définissons pour tout  $\theta \in \Theta$  la tribu  $\mathcal{F}_\theta^X$  par  $\mathcal{F}_\theta^X = \sigma(X_s : s \leq \theta)$ , où  $\sigma(X_s : s \leq \theta)$  est la plus petite tribu qui rend mesurable tous les  $X_s$  pour  $s \leq \theta$  (cette tribu existe car on peut la définir comme étant la plus petite qui contient  $\mathcal{C}_\theta = \{X_s^{-1}(A), A \in \mathcal{E}, s \leq \theta\}$ ).

Notons que, pour  $\theta \leq t$ , tout  $X_s$  avec  $s \leq \theta$  est mesurable par rapport à  $\mathcal{F}_t^X$  puisque  $s \leq t$ . Par suite  $\mathcal{F}_\theta^X \subset \mathcal{F}_t^X$ , ce qui prouve que  $(\mathcal{F}_\theta^X)_{\theta \in \Theta}$  est une filtration.

De plus, pour tout  $\theta \in \Theta$  la v.a.  $X_\theta$  est bien  $\mathcal{F}_\theta^X$ -mesurable (notons qu'elle est  $\sigma(X_\theta)$ -mesurable et que  $\sigma(X_\theta) \subset \mathcal{F}_\theta^X$ ; attention le plus souvent on a  $\sigma(X_\theta) \subsetneq \mathcal{F}_\theta^X$ ).

Donc  $X$  est bien adapté à  $(\mathcal{F}_\theta^X)$ .

Le filtration  $(\mathcal{F}_\theta^X)$  est appelée filtration naturelle du processus  $X$ . L'élément  $\mathcal{F}_\theta^X$  de la filtration représente l'ensemble des évènements qu'on peut décrire en connaissant les trajectoires du processus au plus jusqu'à l'instant  $\theta$ .

On se tourne maintenant vers la notion de processus de Markov. On rappelle la notation introduite dans la définition 2.2.1.

**Définition 3.1.7.** Soit  $(\mathcal{F}_\theta)_{\theta \in \Theta}$  une filtration et  $X = (X_\theta)_{\theta \in \Theta}$  un processus qui lui est adapté. On dit que  $X$  est  $(\mathcal{F}_\theta)$ -Markov si pour tous  $s < t$  dans  $\Theta$ , et tout  $\Gamma$  dans  $\mathcal{E}$  on a

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s) = \mathbb{P}(X_t \in \Gamma | X_s).$$

**Exercice 3.1.1.** Montrer que si  $X = (X_\theta)_{\theta \in \Theta}$  est  $(\mathcal{F}_\theta)$ -Markov pour une certaine filtration  $(\mathcal{F}_\theta)_{\theta \in \Theta}$ , alors il est  $(\mathcal{F}_\theta^X)$ -Markov.

Compte tenu de l'exercice 3.1.1 un processus de Markov  $X$  vérifie, pour tous  $s < t$  dans  $\Theta$ , et tout  $\Gamma$  dans  $\mathcal{E}$ ,

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s).$$

Cela signifie que, pour un processus de Markov  $X$  et  $s < t$ , la loi de  $X_t$  connaissant le comportement du processus jusqu'à l'instant  $s$  compris, ne dépend que de sa position à l'instant  $s$  : on oublie tout ce qui s'est passé juste avant l'instant  $s$ .

**Remarque 3.1.2.** Dans le contexte où  $E$  est fini ou dénombrable le caractère  $(\mathcal{F}_\theta^X)$ -Markov de  $X$  peut s'écrire de façon équivalente : pour tout  $n \in \mathbb{N}$ , tous  $0 \leq t_1 \leq \dots \leq t_n \leq t_{n+1}$ , et tous  $x_1, \dots, x_n, x_{n+1} \in E$  on a

$$\mathbb{P}(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \dots, X_{t_1} = x_1) = \mathbb{P}(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n)$$

dès que  $\mathbb{P}(X_{t_n} = x_n, \dots, X_{t_1} = x_1) > 0$ .



**Définition 3.1.8.** Un processus de Markov  $X = (X_\theta)_{\theta \in \Theta}$  est dit de Markov homogène si pour tous  $s < t$  dans  $\Theta$ , et tout  $\Gamma \in \mathcal{E}$  on a

$$\mathbb{P}(X_t \in \Gamma | X_s) = g_\Gamma(t - s, X_s) \quad (3.1.1)$$

où la fonction  $g_\Gamma(t - s, \cdot)$  dépend de l'écart  $t - s$  mais pas de  $s$ .

Dans ce cas on a pour tous  $s < t$  dans  $\Theta$  et tout  $x \in E$ ,

$$\mathbb{P}(X_t \in \Gamma | X_s = x) = g_\Gamma(t - s, x) = \mathbb{P}(X_{t-s} \in \Gamma | X_0 = x). \quad (3.1.2)$$

**Remarque 3.1.3.** Dans (3.1.2) notons que par exemple  $\mathbb{P}(X_t \in \Gamma | X_s = x)$  est à comprendre au sens des définitions 2.2.1 et 2.2.2 et de l'équation (2.2.3) : la fonction  $x \mapsto \mathbb{P}(X_t \in \Gamma | X_s = x)$  c'est celle qui, appliquée à la variable aléatoire  $X_s$ , nous donne la variable aléatoire  $\mathbb{P}(X_t \in \Gamma | X_s)$  qu'on rencontre dans (3.1.1). De même  $x \mapsto \mathbb{P}(X_{t-s} \in \Gamma | X_0 = x)$  appliquée à  $X_0$  nous donnerait  $\mathbb{P}(X_{t-s} \in \Gamma | X_0)$ .

Notons que pour  $u \geq 0$  il y a donc un sens à considérer  $\mathbb{P}(X_u \in \Gamma | X_0 = x)$  même si  $\mathbb{P}(X_0 = x) = 0$ , ce qui serait le cas si par exemple la loi de  $X_0$  est à densité, ou si  $X_0 = y \neq x$  p.s. (i.e.  $\mathbb{P}(X_0 = y) = 1$ ). En effet la quantité  $\mathbb{P}(X_u \in \Gamma | X_0 = x)$  nous renseigne sur la "tendance" du processus  $X$  à être dans  $\Gamma$  au bout d'un temps  $u$ , s'il était parti de  $x$ . Et le processus  $X$  aura exactement le même tendance à être dans  $\Gamma$  au temps  $u + s$ , s'il était en  $x$  au temps  $s$ , pour tout  $s \geq 0$ , c'est bien l'essence de la propriété de Markov homogène. La quantité  $\mathbb{P}(X_u \in \Gamma | X_0 = x)$  a donc un sens et un intérêt, même si  $\mathbb{P}(X_0 = x) = 0$ .

**Remarque 3.1.4.** A nouveau dans le cas où  $E$  est discret la définition de Markov homogène admet une variante : pour tous  $s < t$  dans  $\Theta$  et tous  $x, y$  dans  $E$

$$\mathbb{P}(X_t = y | X_s = x) = p_{t-s}(x, y)$$

où  $p_{t-s}(x, y)$  dépend de  $x$ , de  $y$  et de l'écart  $t - s$  mais pas de l'instant  $s$ .

## 3.2 Chaînes de Markov

Une chaîne de Markov c'est un processus de Markov homogène à valeurs dans  $E$  au plus dénombrable et indexé par  $n \in \mathbb{N}$  ( i.e.  $X = (X_n)_{n \in \mathbb{N}}$  ; on notera aussi souvent  $X = (X_n)_{n \geq 0}$ , le  $n$  indiquant la nature discrète du temps).

Certains ouvrages ne respectent pas cette terminologie et parlent de chaînes de Markov pour des processus à temps continu (cf par exemple [4]). Disons que la définition de chaîne de Markov donnée ci-dessus est celle communément adoptée dans le monde francophone.

**Proposition 3.2.1.** *Pour une chaîne de Markov  $X = (X_n)_{n \geq 0}$  la propriété de Markov homogène s'écrit de façon équivalente : pour tout  $n \in \mathbb{N}^*$ , et tous  $x_0, x_1, \dots, x_n, x_{n+1} \in E$*

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) = \mathbb{P}(X_1 = x_{n+1} | X_0 = x_n)$$

dès que  $\mathbb{P}(X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) > 0$ .

**Exercice 3.2.1.** Montrer la Proposition 3.2.1.

**Exemple 3.2.1.** Soit  $(\xi_i)_{i \geq 1}$  une suite i.i.d. de v.a. à valeurs dans  $\mathbb{Z}$ . Le processus  $(S_n)_{n \geq 0}$  défini par

$$S_0 = 0 \quad \text{et} \quad S_n = \sum_{i=1}^n \xi_i \quad \forall n \geq 1$$

est une chaîne de Markov (cf Exercice 2 de la feuille de TD 4). C'est une marche aléatoire sur  $\mathbb{Z}$ .

### 3.2.1 Notion de matrice de transition

Ainsi le comportement d'une chaîne de Markov va être complètement défini par la donnée de la loi de sa position initiale  $X_0$  et de ses probabilités de transition, c'est à dire les quantités du type  $\mathbb{P}(X_1 = y | X_0 = x)$ .

Cela nous amène à la notion de matrice de transition (ou matrice stochastique) d'une chaîne de Markov : soit  $X = (X_n)_{n \geq 0}$  une chaîne de Markov à valeurs dans  $E$ , on pose pour tous  $x, y \in E$

$$Q(x, y) = \mathbb{P}(X_1 = y | X_0 = x).$$

On appelle  $Q = \{Q(x, y)\}_{(x, y) \in E^2}$  la matrice de transition de la chaîne  $X$ . Notons que  $Q$  peut être "infinie" (par exemple si  $E = \mathbb{Z}$ ), c'est une famille de probabilités, on ne peut pas toujours la représenter comme une matrice carré à coefficients réels et de dimension finie.

Notons qu'on a  $\sum_{y \in E} Q(x, y) = 1$  pour tout  $x \in E$  (les lignes d'une matrice de transition somment toujours à 1).

On a le résultat suivant.

**Proposition 3.2.2.** *Soit  $X = (X_n)_{n \geq 0}$  une chaîne de Markov de matrice de transition  $Q$ . Alors pour tous  $n \geq 1$  et  $x, y \in E$  on a*

$$\mathbb{P}(X_n = y | X_0 = x) = Q^n(x, y)$$

(ici  $Q^n$  désigne simplement le produit matriciel constitué par  $Q$  multipliée  $n - 1$  fois par elle-même).

En particulier on a les équations de Chapman-Kolmogorov :

$$\forall x, y \in E, \quad \forall m, n \in \mathbb{N}, \quad \mathbb{P}(X_{m+n} = y | X_0 = x) = \sum_{z \in E} \mathbb{P}(X_m = z | X_0 = x) \mathbb{P}(X_n = y | X_0 = z)$$

ou, écrit matriciellement,

$$\forall x, y \in E, \quad \forall m, n \in \mathbb{N}, \quad Q^{m+n}(x, y) = \sum_{z \in E} Q^m(x, z) Q^n(z, y).$$

*Démonstration.* Notons que pour fixer les idées, et pour justifier notamment le recours à la formule (3.2.1) ci-après on va supposer que  $\mathbb{P}(X_0 = x) \neq 0, \forall x \in E$ . Mais ce n'est pas limitatif, et le résultat reste vrai pour une chaîne qui partirait par exemple p.s. d'un certain point  $y \in E$  (cf remarque 3.1.3). Les calculs se mènent alors d'une façon légèrement différente (cf par exemple [3] Lemme 1.2).

La démonstration du premier point se fait par récurrence. Le résultat est vrai à  $n = 1$  par définition de la matrice de transition. Supposons qu'il est vrai au rang  $n$  et montrons qu'il est vrai à  $n + 1$ . On a pour tous  $x, y \in E$

$$\begin{aligned} \mathbb{P}(X_{n+1} = y | X_0 = x) &= \sum_{z \in E} \mathbb{P}(X_{n+1} = y; X_n = z | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_{n+1} = y | X_n = z | X_0 = x) \mathbb{P}(X_n = z | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_{n+1} = y | X_n = z; X_0 = x) \mathbb{P}(X_n = z | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_{n+1} = y | X_n = z) \mathbb{P}(X_n = z | X_0 = x) \\ &= \sum_{z \in E} Q(z, y) Q^n(x, z) \\ &= (Q^n \times Q)(x, y) \\ &= Q^{n+1}(x, y). \end{aligned}$$

Dans ce calcul on a utilisé à la troisième ligne le fait que

$$\mathbb{P}(A|B|C) = \frac{\mathbb{P}(A \cap B|C)}{\mathbb{P}(B|C)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} \times \frac{\mathbb{P}(C)}{\mathbb{P}(B \cap C)} = \mathbb{P}(A | B \cap C). \quad (3.2.1)$$

Puis à la quatrième ligne on a utilisé le fait que  $X$  est Markov. A la cinquième ligne on a utilisé la définition de la matrice de transition et l'hypothèse de récurrence. Enfin à la sixième ligne on a utilisé la formule d'algèbre linéaire (pour  $M$  et  $N$  deux matrices carrées de même taille)

$$(M \times N)_{ij} = \sum_k M_{ik} N_{kj}. \quad (3.2.2)$$

Le premier point étant démontré, les équations de Chapman-Kolmogorov en découlent simplement car - en utilisant à nouveau (3.2.2) -

$$\begin{aligned} \mathbb{P}(X_{m+n} = y | X_0 = x) &= Q^{m+n}(x, y) = (Q^m \times Q^n)(x, y) \\ &= \sum_{z \in E} Q^m(x, z) Q^n(z, y) = \sum_{z \in E} \mathbb{P}(X_m = z | X_0 = x) \mathbb{P}(X_n = y | X_0 = z). \end{aligned}$$

□

### 3.2.2 Phénomène d'équilibre

La quantité  $Q^n(x, y)$  représente donc la probabilité que la chaîne soit à l'état  $y$  à l'instant  $n$ , sachant que son point de départ est  $x$ .

On verra en TD (Exercice 1 de la Feuille 5 ; il s'agira d'une chaîne de Markov à deux états) que, pour  $|E| = k < \infty$  et sous certaines conditions supplémentaires, on a le phénomène de convergence

$$Q^n \xrightarrow{n \rightarrow \infty} \begin{pmatrix} p_1 & \dots & p_k \\ \vdots & & \vdots \\ p_1 & \dots & p_k \end{pmatrix}$$

où  $\sum_{i=1}^k p_i = 1$ .

On insiste sur le fait que les lignes de la matrice limite sont toutes les mêmes. Cela signifie que, en temps long, le système est distribué selon  $\mu$  définie par  $\mu(i) = p_i$ ,  $1 \leq i \leq k$ , et ce, quelle que soit sa position de départ.

Le mesure  $\mu$  apparait comme la "mesure d'équilibre" ou "distribution à l'équilibre" de la chaîne (cf [4] pour plus de détails ; on rentre là dans la théorie des mesures invariantes et la théorie ergodique).



# Chapitre 4

## Martingales à temps discret

Dans ce chapitre on introduit les martingales à temps discret et on étudie certaines de leurs propriétés. On construit notamment une intégrale stochastique discrète qui sera d'une grande importance pour le chapitre 5 à venir.

Dans tout le chapitre un espace de probabilités  $(\Omega, \mathcal{F}, \mathbb{P})$  est donné. On a également une filtration  $(\mathcal{F}_n)_{n \geq 0}$  sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . Le cas échéant on travaille avec d'autres filtrations, qui seront précisées. On suppose que les processus et variables aléatoires rencontrés sont à valeurs dans  $\mathbb{R}$ , sauf précision contraire.

### 4.1 Définition et premières propriétés

**Définition 4.1.1.** Un processus  $X = (X_n)_{n \geq 0}$  défini sur  $(\Omega, \mathcal{F}, \mathbb{P})$  est une  $(\mathcal{F}_n)$ -martingale (ou une martingale par rapport à la filtration  $(\mathcal{F}_n)$ , on dit simplement une martingale s'il n'y a pas d'ambiguïté quant à la filtration) si

- i) Pour tout  $n \geq 0$  on a  $\mathbb{E}|X_n| < +\infty$  (on dit que  $X$  est intégrable).
- ii) Le processus  $X$  est  $(\mathcal{F}_n)$ -adapté.
- iii) Pour tout  $n \geq 0$  on a  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$  (propriété de martingale).

Si la condition iii) est remplacée par  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$  on dit que  $X$  est une sur-martingale.

Si la condition iii) est remplacée par  $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$  on dit que  $X$  est une sous-martingale.

**Propriété 4.1.1.** On a les propriétés immédiates suivantes.

- 1) Pour tout  $m \geq n \geq 0$ , on a  $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$ .
- 2) Pour tout  $n \geq 0$  on a  $\mathbb{E}(X_n) = \mathbb{E}(X_0)$  (une martingale est constante en espérance).

*Démonstration.* 1) Soit  $n$  fixé, montrons par récurrence que la propriété est vraie pour  $m \geq n$ . Elle l'est déjà pour  $m = n$  et  $m = n + 1$ . Supposons qu'elle est vraie au rang  $m \geq n$  et montrons qu'elle est vraie au rang  $m + 1$ . Cela vient du fait que

$$\mathbb{E}(X_{m+1}|\mathcal{F}_n) = \mathbb{E}[\mathbb{E}(X_{m+1}|\mathcal{F}_m) | \mathcal{F}_n] = \mathbb{E}(X_m|\mathcal{F}_n) = X_n$$

(on a utilisé en particulier le point vii) de la Propriété 2.2.1).

- 2) Il suffit de remarquer que grâce à 1) on a  $\mathbb{E}(X_n) = \mathbb{E}[\mathbb{E}(X_n|\mathcal{F}_0)] = \mathbb{E}(X_0)$ . □

**Exemple 4.1.1.** On considère le processus  $S = (S_n)_{n \geq 0}$  défini par

$$S_0 = x \quad \text{et} \quad S_n = x + \sum_{i=1}^n X_i, \quad \forall n \geq 1,$$

avec  $(X_i)$  suite i.i.d. vérifiant  $\mathbb{E}|X_1| < \infty$  et  $\mathbb{E}(X_1) = 0$ . Un tel processus est appelé marche aléatoire symétrique, issue de  $x$  (le processus rencontré dans l'exemple 3.2.1 était une marche aléatoire sur  $\mathbb{Z}$  issue de zéro; elle n'était pas forcément symétrique car on n'avait pas fait l'hypothèse que la loi des  $\xi_i$ 's était centrée).

On définit la filtration  $(\mathcal{G}_n)$  par  $\mathcal{G}_0 = \{\emptyset, \Omega\}$  puis  $\mathcal{G}_n = \sigma(X_k, 1 \leq k \leq n)$  pour  $n \geq 1$ .

On peut alors voir que  $S$  est une  $(\mathcal{G}_n)$ -martingale.

En effet pour tout  $n \geq 0$  on a  $\mathbb{E}|S_n| \leq |x| + \sum_{k=1}^n \mathbb{E}|X_k| < +\infty$ .

Par ailleurs pour  $S_0$  est constante donc  $\mathcal{G}_0$ -mesurable (Exercice 3 Feuille 2). Puis pour  $n \geq 1$  la v.a.  $S_n$  s'obtient en appliquant à  $X_1, \dots, X_n$  une application continue (la somme), donc mesurable; donc  $S_n$  est  $\mathcal{G}_n$ -mesurable. On a donc que  $S$  est  $(\mathcal{G}_n)$ -adapté (en fait on a mieux, on pourrait voir que  $(\mathcal{G}_n)$  est la filtration naturelle de  $S$ ).

Reste à vérifier la propriété de martingale : on a pour tout  $n \geq 0$ ,

$$\mathbb{E}(S_{n+1}|\mathcal{G}_n) = \mathbb{E}(S_n + X_{n+1}|\mathcal{G}_n) = \mathbb{E}(S_n|\mathcal{G}_n) + \mathbb{E}(X_{n+1}|\mathcal{G}_n) = S_n + \mathbb{E}(X_{n+1}) = S_n$$

en utilisant successivement la linéarité de l'espérance conditionnelle, le fait que  $S_n$  est  $\mathcal{G}_n$ -mesurable et le fait que  $X_{n+1}$  est indépendant de  $\mathcal{G}_n$ .

**Théorème 4.1.1.** 1) Soit  $(X_n)_{n \geq 0}$  une martingale et soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction convexe telle que  $\mathbb{E}|\varphi(X_n)| < +\infty$  pour tout  $n \geq 0$ . Alors  $(\varphi(X_n))_{n \geq 0}$  est une sous-martingale.

2) Soit  $(X_n)$  une sous-martingale et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction convexe croissante avec  $\mathbb{E}|\varphi(X_n)| < +\infty$  pour tout  $n \geq 0$ . Alors  $(\varphi(X_n))_{n \geq 0}$  est une sous-martingale.

*Démonstration.* On prouve le 1), le 2) se démontrant de façon analogue.

Pour tout  $n$  la v.a.  $\varphi(X_n)$  est intégrable par hypothèse. Le fait que  $\varphi$  est convexe entraîne en fait qu'elle est borélienne, donc pour tout  $n$  la v.a.  $\varphi(X_n)$  est  $\mathcal{F}_n$ -mesurable, car  $X_n$  l'est (argument de composition). Enfin on a la propriété de sous-martingale car en utilisant Jensen conditionnel (théorème 2.2.1, point 4) on a

$$\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n)$$

pour tout  $n$ . □

## 4.2 Intégrale stochastique discrète

**Définition 4.2.1.** Un processus  $(H_n)_{n \geq 0}$  est dit prévisible si  $H_n$  est  $\mathcal{F}_{n-1}$  mesurable pour tout  $n \geq 1$  (et si  $H_0$  est  $\mathcal{F}_0$ -mesurable).

**Remarque 4.2.1.** Un processus prévisible est adapté. La réciproque n'est pas vraie en général.

**Théorème 4.2.1.** Soit  $(X_n)_{n \geq 0}$  une sur-martingale et soit  $(H_n)_{n \geq 0}$  un processus prévisible, à valeurs positives et bornées.

Alors le processus  $H \cdot X = ((H \cdot X)_n)_{n \geq 0}$  défini par  $(H \cdot X)_0 = 0$  et

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

est à son tour une sur-martingale.

*Démonstration.* Le processus  $H \cdot X$  est clairement  $(\mathcal{F}_n)$ -adapté. Vérifions l'intégrabilité. Par hypothèse il existe  $0 < c < \infty$  t.q.  $0 \leq H_m \leq c$  pour tout  $m$  donc on a pour tout  $n$

$$|(H \cdot X)_n| \leq \sum_{m=1}^n c|X_m - X_{m-1}| \leq 2c \sum_{m=0}^n |X_m| \in L^1,$$

ce qui entraîne  $(H \cdot X)_n \in L^1$ .

Enfin vérifions la propriété de sur-martingale. Pour tout  $n$  on a

$$\mathbb{E}[(H \cdot X)_{n+1}|\mathcal{F}_n] = \mathbb{E}\left[\sum_{m=1}^{n+1} H_m(X_m - X_{m-1})|\mathcal{F}_n\right] = \sum_{m=1}^n H_m(X_m - X_{m-1}) + \mathbb{E}[H_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n].$$

Or, en utilisant le fait que  $H_{n+1}$  est  $\mathcal{F}_n$ -mesurable, on a

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

et cette dernière quantité est négative ou nulle. En effet  $H_{n+1} \geq 0$  par hypothèse et  $\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = \mathbb{E}(X_{n+1}|\mathcal{F}_n) - X_n \leq 0$  car  $X$  est une sur-martingale.

Donc on a bien  $\mathbb{E}[(H \cdot X)_{n+1}|\mathcal{F}_n] \leq (H \cdot X)_n$  pour tout  $n$ . □

**Remarque 4.2.2.** On a des versions sous-martingale et martingale du théorème 4.2.1.

1) Si  $X$  est une sous-martingale et avec les mêmes hypothèses sur  $H$  on a que  $H \cdot X$  est à son tour une sous-martingale.

2) Si  $X$  est une martingale et  $H$  est un processus prévisible à valeurs bornées (mais de signe quelconque) alors  $H \cdot X$  est à son tour une martingale.

### 4.3 Théorèmes d'arrêt

**Définition 4.3.1.** Un temps d'arrêt (t.a.) pour la filtration  $(\mathcal{F}_n)$  (on parle aussi de  $(\mathcal{F}_n)$ -t.a.) est une v.a.  $T$  à valeurs dans  $\mathbb{N} \cup \{+\infty\}$  telle que

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

**Remarque 4.3.1.** Comme on est en temps discret on peut remplacer de façon équivalente  $\{T \leq n\}$  par  $\{T = n\}$  dans la définition (ATTENTION : dans les modèles à temps continu ce n'est pas le cas).

En effet si on a la définition avec  $\{T = n\}$  : pour tout  $n$  on a  $\{T \leq n\} = \bigcup_{i=0}^n \{T = i\}$ , mais  $\{T = i\} \in \mathcal{F}_i \subset \mathcal{F}_n$  pour tout  $i \leq n$ . Donc par stabilité par réunion dénombrable a on bien  $\{T \leq n\} \in \mathcal{F}_n$ .

Réciproquement si on a la définition avec  $\{T \leq n\}$  : pour tout  $n$  on remarque que

$$\{T = n\} = \{T \leq n\} \cap \{T > n - 1\} = \{T \leq n\} \cap \{T \leq n - 1\}^c.$$

Or ceci est dans  $\mathcal{F}_n$  car  $\{T \leq n\} \in \mathcal{F}_n$  et  $\{T \leq n - 1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$  (on utilise alors la stabilité par passage au complémentaire et par intersection dénombrable).

**Exemple 4.3.1.** Soit  $(X_n)$  un processus à valeurs dans  $\mathbb{R}$  et  $(\mathcal{G}_n)$  sa filtration naturelle (on suppose en outre que  $(\mathcal{G}_n)$  est strictement croissante ; c'est souvent le cas dans les modèles étudiés, cf exemple 4.1.1).

Soit  $a \in \mathbb{R}$ , on considère le temps aléatoire  $T_a = \inf\{n \geq 0 : X_n = a\}$ .

(NB : on pose la convention  $\inf \emptyset = +\infty$ , si bien que l'évènement  $\{T_a = \infty\}$  correspond à "le processus  $X$  ne touche jamais le point  $a$ ").

On a que  $T_a$  est un  $(\mathcal{G}_n)$ -t.a.

En effet  $\{T_a \leq n\} = \{\text{Entre les instants } 0 \text{ et } n, X \text{ a touché } a\}$  et cet évènement est dans  $\mathcal{G}_n$  car il peut se décrire en connaissant les trajectoires de  $X$  sur l'intervalle de temps  $\llbracket 0, n \rrbracket$ .

En revanche  $S_a = \sup\{n \geq 0 : X_n = a\}$  n'est pas un temps d'arrêt.

En effet  $\{S_a \leq n\} = \{\text{Le dernier instant où } X \text{ a touché } a \text{ est avant } n\}$  n'est pas dans  $\mathcal{G}_n$ , on peut juste dire qu'il est dans  $\mathcal{F}$ . En effet il faut savoir ce qui se passe après l'instant  $n$  (le processus ne touche plus  $a$ ) pour le décrire.

On va maintenant énoncer quelques "théorèmes d'arrêt".

**Théorème 4.3.1.** Soit  $X = (X_n)_{n \geq 0}$  une  $(\mathcal{F}_n)$ -sur-martingale et  $T$  un  $(\mathcal{F}_n)$ -t.a. Alors le processus  $X^T = (X_{n \wedge T})_{n \geq 0}$  est encore une sur-martingale (on parle de martingale arrêtée en  $T$ ).

Le résultat reste vrai en remplaçant partout "sur-martingale" par "sous-martingale" ou "martingale".

*Démonstration.* On traite le cas sur-martingale en utilisant le théorème 4.2.1.

On pose  $H_0 = 0$  puis  $H_k = \mathbf{1}_{k \leq T}$  pour tout  $k \geq 1$ . Le processus  $(H_n)$  est bien évidemment à valeurs positives et bornées.

En outre il est prévisible. En effet  $H_0$  est bien  $\mathcal{F}_0$ -mesurable (c'est une constante donc mesurable par rapport à la tribu grossière, donc par rapport à n'importe quelle tribu), et pour tout  $k \geq 1$  on a

$$\{k \leq T\} = \{T < k\}^c = \{T \leq k - 1\}^c \in \mathcal{F}_{k-1},$$

ce qui entraîne que  $H_k$  est  $\mathcal{F}_{k-1}$ -mesurable.

Donc par le théorème 4.2.1 le processus  $H \cdot X$  est une sur-martingale.

On va maintenant voir que pour tout  $n$ ,  $X_{n \wedge T} = (H \cdot X)_n + X_0$ , ce qui amènera le résultat.

Soit  $n$  fixé. On a  $(H \cdot X)_n = \sum_{k=1}^n \mathbf{1}_{k \leq T} (X_k - X_{k-1})$ . Pour tout  $j \in \mathbb{N} \cup \{+\infty\}$  sur l'évènement  $\{T = j\}$  on a

$$(H \cdot X)_n = \begin{cases} \sum_{k=1}^n (X_k - X_{k-1}) = X_n - X_0 & \text{si } j \geq n \\ \sum_{k=1}^j (X_k - X_{k-1}) = X_j - X_0 & \text{si } j < n \end{cases}$$

donc

$$(H \cdot X)_n + X_0 = \begin{cases} X_n & \text{si } j \geq n \\ X_j & \text{si } j < n \end{cases} = X_{n \wedge j}$$

Donc

$$(H \cdot X)_n + X_0 = \left( \sum_{j=0}^{+\infty} \mathbf{1}_{T=j} \right) [(H \cdot X)_n + X_0] = \sum_{j=0}^{+\infty} \mathbf{1}_{T=j} X_{n \wedge j} = \sum_{j=0}^{+\infty} \mathbf{1}_{T=j} X_{n \wedge T} = X_{n \wedge T},$$

ce qui achève la preuve.  $\square$

**Définition 4.3.2** (Tribu des évènements antérieurs à un temps d'arrêt). Soit  $T$  un  $(\mathcal{F}_n)$ -t.a. On note

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq n\} \in \mathcal{F}_n, \forall n \geq 0\}$$

la "tribu des évènements antérieurs au t.a.  $T$ ".

**Remarque 4.3.2.** C'est un exercice de montrer que  $\mathcal{F}_T \subset \mathcal{F}$  est effectivement une tribu (cf Exercice 1 de la Feuille de TD 7).

Il existe de nombreuses variantes des théorèmes d'arrêt mettant en jeu la tribu des évènements antérieurs. On a par exemple le résultat suivant.

**Théorème 4.3.2.** Soit  $X = (X_n)_{n \geq 0}$  une  $(\mathcal{F}_n)$ -martingale. Soient  $S$  et  $T$  deux  $(\mathcal{F}_n)$ -t.a. bornés vérifiant  $S \leq T \leq k$  p.s. (pour  $k < \infty$ ).

Alors on a

$$\mathbb{E}(X_T | \mathcal{F}_S) = X_S \quad \text{p.s.}$$

(noter que par exemple  $X_T(\omega) = X_{T(\omega)}(\omega)$  pour tout  $\omega \in \Omega$ ).

En particulier on a  $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ .

*Démonstration. Étape préliminaire.* Dans la définition 4.3.2 on peut remplacer de façon équivalente le signe  $\leq$  par le signe  $=$ .

C'est à dire que si on note

$$\bar{\mathcal{F}}_T = \{A \in \mathcal{F} : A \cap \{T = n\} \in \mathcal{F}_n, \forall n \geq 0\}$$

on a  $\mathcal{F}_T = \bar{\mathcal{F}}_T$  (et donc  $\bar{\mathcal{F}}_T$  est une tribu, au même titre que  $\mathcal{F}_T$ ).

En effet si  $A \in \bar{\mathcal{F}}_T$  on a pour tout  $n$

$$A \cap \{T \leq n\} = \bigcup_{i \leq n} A \cap \{T = i\} \in \mathcal{F}_n$$

car chaque  $A \cap \{T = i\}$  est dans  $\mathcal{F}_i$ , donc dans  $\mathcal{F}_n$  pour  $i \leq n$ . Donc  $A \in \mathcal{F}_T$ , et ceci montre  $\bar{\mathcal{F}}_T \subset \mathcal{F}_T$ .

Réciproquement, si  $A \in \mathcal{F}_T$ , on a pour tout  $n$ ,

$$A \cap \{T = n\} = A \cap \left( \{T \leq n\} \cap \{T > n-1\} \right) = A \cap \{T \leq n\} \cap \{T \leq n-1\}^c \in \mathcal{F}_n$$

et ceci est dans  $\mathcal{F}_n$  car comme  $A \in \mathcal{F}_T$  on a  $A \cap \{T \leq n\} \in \mathcal{F}_n$ , et comme  $T$  est un t.a. on a  $\{T \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$ . Donc  $A \in \bar{\mathcal{F}}_T$  et ceci montre  $\mathcal{F}_T \subset \bar{\mathcal{F}}_T$ , d'où le résultat.

**Étape 1.** On m.q.  $X_S$  est  $\mathcal{F}_S$ -mesurable. Le processus  $X$  prend ses valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , soit donc  $B \in \mathcal{B}(\mathbb{R})$ . On a pour tout  $n$ ,

$$\{X_S \in B\} \cap \{S = n\} = \{X_n \in B\} \cap \{S = n\} \in \mathcal{F}_n$$

car  $\{X_n \in B\} \in \mathcal{F}_n$  car  $X_n$  est  $\mathcal{F}_n$ -mesurable (puisque  $X$  est adapté), et  $\{S = n\} \in \mathcal{F}_n$  puisque  $S$  est un t.a. (cf remarque 4.3.1). Donc  $\{X_S \in B\} \in \mathcal{F}_S$  par l'étape préliminaire.

**Étape 2.** Pour tout  $A \in \mathcal{F}_S$ , on montre qu'on a  $\mathbb{E}[\mathbf{1}_A X_S] = \mathbb{E}[\mathbf{1}_A X_T]$ .



**Sous-étape 1.** On suppose  $S \leq T = k$ . Soit  $A \in \mathcal{F}_S$ . Pour tout  $0 \leq j \leq k$  on a  $A \cap \{S = j\} \in \mathcal{F}_j$ , donc

$$\begin{aligned}\mathbb{E}[\mathbf{1}_A X_S] &= \sum_{j=0}^k \mathbb{E}[\mathbf{1}_A \mathbf{1}_{\{S=j\}} X_S] = \sum_{j=0}^k \mathbb{E}[\mathbf{1}_{A \cap \{S=j\}} X_j] \\ &= \sum_{j=0}^k \mathbb{E}[\mathbf{1}_{A \cap \{S=j\}} \mathbb{E}(X_k | \mathcal{F}_j)] = \sum_{j=0}^k \mathbb{E}[\mathbf{1}_{A \cap \{S=j\}} X_k] \\ &= \mathbb{E}[\mathbf{1}_A X_k].\end{aligned}$$

Notons au passage que ceci montre que  $\mathbb{E}(X_k | \mathcal{F}_S) = X_S$ .

**Sous-étape 2.** On traite le cas  $S \leq T \leq k$  p.s. Pour tout  $A \in \mathcal{F}_S$  on a

$$\mathbb{E}[\mathbf{1}_A X_S] = \mathbb{E}[\mathbf{1}_A X_S^T] = \mathbb{E}[\mathbf{1}_A X_k^T] = \mathbb{E}[\mathbf{1}_A X_T].$$

Ici on a appliqué à la deuxième égalité la sous-étape 1 à la martingale arrêtée  $(X_n^T) = (X_{n \wedge T})$  (cf théorème 4.3.1).

Comme on a vu à l'étape 1 que  $X_S$  est  $\mathcal{F}_S$ -mesurable on conclut que  $\mathbb{E}(X_T | \mathcal{F}_S) = X_S$ .  $\square$

## 4.4 Résultats de convergence

On a le type de résultat suivant.

**Théorème 4.4.1.** *Soit  $(X_n)$  une sous-martingale telle que  $\sup_n \mathbb{E}[(X_n)_+] < +\infty$ . Alors il existe  $X \in L^1$  telle que*

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad \text{p.s.}$$

Pour la preuve du théorème 4.4.1 il nous faut introduire quelques notations.

Soit  $X = (X_n)_{n \geq 0}$  une sous-martingale et  $a < b$  donnés. On définit  $N_0 = 0$  puis, pour tout  $k \geq 1$ ,

$$N_{2k-1} = \inf\{m > N_{2k-2} : X_m \leq a\}$$

$$\text{et } N_{2k} = \inf\{m > N_{2k-1} : X_m \geq b\}$$

(c'est un exercice de vérifier que ce sont des temps d'arrêt). Il peut être utile de dessiner une trajectoire de  $X$  (erratique, qui passe au-dessus et en-dessous de  $a$  et  $b$  à plusieurs reprises) pour comprendre ce que signifient les t.a.  $N_p$  : entre les instants  $N_1$  et  $N_2$  le processus  $X$  effectue une "ascension" de  $a$  vers  $b$ ; entre  $N_2$  et  $N_3$  il effectue une descente de  $b$  vers  $a$  etc...

On définit ensuite le processus  $H = (H_m)_{m \geq 0}$  par  $H_0 = 0$  et

$$H_m = \mathbf{1}_{\cup_{k=1}^{\infty} \{N_{2k-1} < m \leq N_{2k}\}}, \quad m \geq 1.$$

C'est à dire que  $H_m$  vaut 1 si l'instant  $m$  fait partie d'une phase d'ascension de  $a$  vers  $b$  et 0 sinon.

Le processus  $H$  est prévisible. En effet pour tout  $m \geq 1$  on a

$$\bigcup_{k=1}^{\infty} \{N_{2k-1} < m \leq N_{2k}\} = \bigcup_{k=1}^{\infty} \{N_{2k-1} \leq m-1\} \cap \{N_{2k} \leq m-1\}^c.$$

Or les  $N_p$ 's sont des t.a. donc  $\{N_{2k-1} \leq m-1\}$  et  $\{N_{2k} \leq m-1\}$  sont dans  $\mathcal{F}_{m-1}$ . Par stabilité de  $\mathcal{F}_{m-1}$  par passage au complémentaire et réunion dénombrable l'évènement  $\cup_{k=1}^{\infty} \{N_{2k-1} < m \leq N_{2k}\}$  est donc dans  $\mathcal{F}_{m-1}$ .

Enfin on pose  $U_n = \sup\{k : N_{2k} \leq n\}$ , pour  $n \geq 0$ . C'est le nombre d'ascensions complètes de  $a$  vers  $b$  effectuées par  $X$  jusqu'au temps  $n$ .

Le théorème suivant nous servira en fait de lemme dans la démonstration du théorème 4.4.1.

**Théorème 4.4.2** (Inégalités des cascades). *Sous les hypothèses du théorème 4.4.1, et avec les notations qui le suivent on a*

$$(b-a)\mathbb{E}(U_n) \leq \mathbb{E}[(X_n - a)_+] - \mathbb{E}[(X_0 - a)_+].$$

*Démonstration.* Tout d'abord remarquons que le processus  $(Y_n)$  défini par  $Y_n = a + (X_n - a)_+$ ,  $n \geq 0$ , est une sous-martingale par le théorème 4.1.1-2).

Le processus  $Y$  fait le même nombre d'ascensions de  $a$  vers  $b$  jusqu'au temps  $n$  que  $X$  : c'est à dire  $U_n$ .

Considérons donc pour  $n \geq 1$  la quantité  $\sum_{k=1}^n H_k(Y_k - Y_{k-1})$ , c'est la somme des hauteurs atteintes au cours des ascensions complètes de  $a$  vers  $b$  intervenues jusqu'au temps  $n$ , et par l'observation qui précède on a donc

$$\sum_{k=1}^n H_k(Y_k - Y_{k-1}) \geq (b - a)U_n.$$

Mais par ailleurs  $\sum_{k=1}^n H_k(Y_k - Y_{k-1}) = (H \cdot Y)_n$  où  $H \cdot Y$  est l'intégrale stochastique considérée dans le théorème 4.2.1. On a donc

$$(b - a)U_n \leq (H \cdot Y)_n, \quad \forall n \geq 0. \quad (4.4.1)$$

Posons  $K_m = 1 - H_m$  on a  $H_m + K_m = 1$  pour tout  $m$ . Le processus  $(K_m)$  est prévisible à valeurs positives et bornées (au même titre que  $H$ ). Donc  $K \cdot X$  est une sous-martingale d'après la remarque 4.2.2, point 1). Donc

$$\mathbb{E}[(K \cdot Y)_n] \geq \mathbb{E}[(K \cdot Y)_0] = 0, \quad \forall n \geq 0. \quad (4.4.2)$$

Or pour tout  $n$ ,

$$Y_n - Y_0 = (1 \cdot Y)_n = ((H + K) \cdot Y)_n = (H \cdot Y)_n + (K \cdot Y)_n.$$

En combinant (4.4.1) et (4.4.2) il vient

$$\mathbb{E}(Y_n - Y_0) \geq \mathbb{E}[(H \cdot Y)_n] \geq (b - a)\mathbb{E}(U_n).$$

Or  $\mathbb{E}(Y_n - Y_0) = \mathbb{E}(X_n - a)_+ - \mathbb{E}(X_0 - a)_+$  d'où le résultat.  $\square$

*Démonstration du théorème 4.4.1.* On fixe  $a < b$  dans un premier temps. On rappelle que  $(x - a) = x - a$  si  $x \geq a$ , que  $x \leq x_+$  et que  $-a \leq |a|$ . Ainsi on a  $(x - a)_+ \leq x_+ + |a|$  et par le théorème 4.4.2 il vient

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

On a donc

$$\mathbb{E}(U_n) \leq \frac{\sup_n \mathbb{E}[(X_n)_+] + |a|}{b - a} < +\infty \quad (4.4.3)$$

par hypothèse. Mais  $(U_n)$  tend en croissant vers  $U$ , le nombre total d'ascensions de  $a$  vers  $b$  sur toute le trajectoire (à horizon temporel infini) de  $X$ . Notons qu'éventuellement on a  $U = \infty$ . Par convergence monotone (théorème 1.5.1, point 1)) on a donc que  $\mathbb{E}(U_n) \uparrow \mathbb{E}(U)$  quand  $n \rightarrow \infty$ . En tenant compte de (4.4.3) on a donc

$$\mathbb{E}(U) \leq \frac{\sup_n \mathbb{E}[(X_n)_+] + |a|}{b - a} < +\infty$$

(le passage à la limite préserve les inégalités larges).

Comme  $U \geq 0$  ceci implique que  $U < \infty$  p.s. (Proposition 1.2.1, point ii)).

En conséquence  $\mathbb{P}(\liminf_n X_n < a < b < \limsup_n X_n) = 0$ .

(En effet : L'évènement  $\{\liminf_n X_n < a < b < \limsup_n X_n\}$  est inclus dans l'évènement  $\{U = \infty\}$  qui est de probabilité nulle.

Comment voir cette inclusion ? Avoir  $\liminf_n X_n < a$  signifie que  $\inf_{k \geq n} X_k$  converge vers un certain  $-\infty \leq l < a$  quand  $n \rightarrow \infty$ . Qu'on soit dans la situation  $l = -\infty$  ou  $-\infty < l$  cela implique que pour  $n$  assez grand on a  $\inf_{k \geq n} X_k < a$ . De même pour  $n$  assez grand on a  $\sup_{k \geq n} X_k > b$ .

Finalement, pour tout  $N \geq 0$  il existe  $k_1 > N$  t.q.  $X_{k_1} < a$  et  $k_2 > k_1$  t.q.  $X_{k_2} > b$ . Ce qui signifie que  $X$  fait un nombre infini d'ascensions de  $a$  vers  $b$ .)

On fait maintenant varier  $a$  et  $b$  dans  $\mathbb{Q}$  dénombrable. On a par  $\sigma$ -sous-additivité

$$\mathbb{P}\left(\bigcup_{a,b \in \mathbb{Q}, a < b} \{\liminf_n X_n < a < b < \limsup_n X_n\}\right) = 0$$

ce qui équivaut à

$$\mathbb{P}\left(\bigcap_{a,b \in \mathbb{Q}, a < b} \{a \leq \liminf_n X_n \leq \limsup_n X_n \leq b\}\right) = 1.$$

Mais  $\bigcap_{a,b \in \mathbb{Q}, a < b} \{a \leq \liminf_n X_n \leq \limsup_n X_n \leq b\} = \{\liminf_n X_n = \limsup_n X_n\}$ , c'est l'évènement " $(X_n)$  converge". On a donc  $X_n \rightarrow X$  p.s., pour une certaine v.a.r.  $X$ . Reste à voir que  $X \in L^1$ .

Par Fatou et l'hypothèse d'intégrabilité (et la continuité de la fonction positive) il vient

$$\mathbb{E}(X_+) = \mathbb{E}[\liminf_n (X_n)_+] \leq \liminf_n \mathbb{E}[(X_n)_+] \leq \limsup_n \mathbb{E}[(X_n)_+] < +\infty.$$

Mais  $\mathbb{E}(X_n) = \mathbb{E}[(X_n)_+] - \mathbb{E}[(X_n)_-]$  et donc

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

en utilisant le fait que  $(X_n)$  est une sous-martingale. Par Fatou il vient

$$\mathbb{E}(X_-) \leq \liminf_n \mathbb{E}[(X_n)_-] \leq \limsup_n \mathbb{E}[(X_n)_-] \leq \limsup_n (\mathbb{E}[(X_n)_+] - \mathbb{E}(X_0)) < +\infty.$$

Donc  $X_-, X_+ \in L^1$ , ceci montre  $X \in L^1$ . □

**Corollaire 4.4.1.** Soit  $(X_n)_{n \geq 0}$  une sur-martingale positive alors

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad \text{p.s.}$$

où  $\mathbb{E}(X) \leq \mathbb{E}(X_0)$ .

*Démonstration.* Laissez en exercice. □

On a aussi des résultats de convergence  $L^1$ . Dans le théorème suivant on en énonce certains et on en profite pour introduire la notion de martingale uniformément intégrable (U.I.)

**Théorème 4.4.3.** Soit  $(X_n)$  une martingale. Il y a équivalence entre :

i) Elle est U.I. (i.e.  $\sup_n \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] \xrightarrow[a \uparrow \infty]{} 0$ ).

ii) Elle converge p.s. et dans  $L^1$  vers  $X \in L^1$ .

iii) Il existe  $X \in L^1$  t.q.  $X_n = \mathbb{E}(X | \mathcal{F}_n)$ .

Dans les points ii) et iii) c'est la même variable aléatoire  $X$  qui est en jeu.

*Démonstration.* Cf Feuille de TD 7 Exercice 4. □

**Remarque 4.4.1.** Si  $|X_n| \leq Y \in L^1$  pour tout  $n$  alors  $(X_n)$  est U.I.

En effet on a

$$\sup_n \mathbb{E}[|X_n| \mathbf{1}_{|X_n| > a}] \leq \mathbb{E}[|Y| \mathbf{1}_{|Y| > a}] \xrightarrow[a \uparrow \infty]{} 0$$

(en utilisant la convergence dominée pour justifier la convergence).

## 4.5 Décomposition de Doob

**Théorème 4.5.1.** Toute sous-martingale  $(X_n)_{n \geq 0}$  admet la décomposition

$$X_n = M_n + A_n, \quad \forall n \geq 0,$$

où  $(M_n)_{n \geq 0}$  est une martingale et  $(A_n)_{n \geq 0}$  un processus prévisible croissant, avec  $A_0 = 0$ . Cette décomposition est unique.

De plus on a les expressions explicites :

$$M_0 = X_0 \quad \text{et} \quad M_n = X_n - \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}), \quad \forall n \geq 1$$

et

$$A_0 = 0 \quad \text{et} \quad A_n = \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}), \quad \forall n \geq 1.$$

*Démonstration. Existence.* Le processus  $(A_n)$  ainsi défini est bien prévisible (en particulier on a bien  $A_n$  qui est  $\mathcal{F}_{n-1}$ -mesurable pour tout  $n \geq 1$ , puisque  $A_n$  est constitué d'espérances sachant  $\mathcal{F}_{k-1}$  avec  $k \leq n$ ).

De plus il est bien croissant car

$$A_{n+1} - A_n = \mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) \geq 0, \quad \forall n \geq 0,$$

car  $(X_n)$  est une sous-martingale.

Le processus  $(M_n)$  est adapté et intégrable (découle du caractère adapté et intégrable de la martingale  $(X_n)$ ). Vérifions la propriété de martingale. On a pour  $n \geq 1$ ,

$$\begin{aligned} \mathbb{E}(M_n | \mathcal{F}_{n-1}) &= \mathbb{E}(X_n | \mathcal{F}_{n-1}) - \sum_{k=1}^n \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}) \\ &= \mathbb{E}(X_n | \mathcal{F}_{n-1}) - \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) - \sum_{k=1}^{n-1} \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}) \\ &= X_{n-1} - \sum_{k=1}^{n-1} \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1}) \\ &= M_{n-1}. \end{aligned}$$

**Unicité.** Supposons que  $X = \bar{M} + \bar{A}$  avec une autre martingale  $\bar{M}$  et un autre processus prévisible croissant  $\bar{A}$  (avec  $\bar{A}_0 = 0$ ).

On a pour tout  $n \geq 1$

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] = \mathbb{E}[\bar{M}_n + \bar{A}_n | \mathcal{F}_{n-1}] = \bar{M}_{n-1} + \bar{A}_n = X_{n-1} - \bar{A}_{n-1} + \bar{A}_n$$

ce qui implique

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

(en utilisant l'expression explicite du processus  $A$  mis en lumière dans la partie existence).

Or  $A_0 = \bar{A}_0 = 0$ , on a donc  $\bar{A}_n = A_n$  pour tout  $n$ . Finalement  $\bar{M}_n = X_n - \bar{A}_n = X_n - A_n = M_n$  pour tout  $n$ , ce qui achève la preuve de l'unicité.  $\square$

# Chapitre 5

## Modèles financiers à temps discret

Dans ce chapitre on présente des notions financières comme les options ou l'absence d'opportunité d'arbitrage. Puis on introduit les modèles financiers à temps discret type Cox-Ross-Rubinstein, qui utilisent les martingales à temps discret introduites dans le chapitre 4. Le contenu de ce chapitre est en partie inspiré de [6] et [7].

### 5.1 Introduction

**Exemple des options.** Pour introduire les problématiques auxquelles va faire face un ingénieur financier on va partir d'un exemple assez précis : les options.

Une option c'est un contrat passé entre un acheteur et un vendeur (de l'option). En  $t = 0$  l'acheteur verse une prime  $p_0$  au vendeur en échange du droit d'effectuer en  $t = T$  ( $T$  est la maturité ou échéance de l'option) une opération sur un sous-jacent, ou actif. Le sous-jacent peut être une action, un taux de change, une matière première (ex : un baril de pétrole), de l'énergie (ex : de l'électricité)...

Pour concrétiser les choses on va préciser encore l'exemple : considérons que le sous-jacent est une action cotée sur le marché, dont le prix à l'instant  $t$  est noté  $S_t$ . Par la suite on considère son processus prix  $S = (S_t)_{t \geq 0}$ . Par abus de langage on parlera parfois du sous-jacent  $S$  (c'est un raccourci). Considérons alors l'option d'achat (option Call, ou Call) de maturité  $T$  et de "strike"  $K$  : c'est le droit d'acheter à l'instant  $T$  l'action au prix  $K$ . Les quantités  $T$  et  $K$  sont convenues à l'avance dans le contrat qui constitue l'option Call.

Si on résume ce qui se passe :

En  $t = 0$  l'acheteur verse  $p_0$  euros au vendeur du Call.

En  $t = T$  : si  $S_T > K$  l'acheteur exerce son droit, il peut ainsi acheter l'action au prix  $K$  et la revendre aussitôt au prix  $S_T$ . Il réalise ainsi un bénéfice de  $S_T - K$  euros.

Si  $S_T \leq K$  l'acheteur n'a aucun intérêt à exercer son droit. Il n'exerce pas (on dit que le Call est mort), et réalise un bénéfice de 0 euros.

**Point de vue de l'acheteur :** Il reçoit en  $t = T$  la richesse  $(S_T - K)_+$  euros (cette quantité est appelée le "pay-off" de l'option). Rappelons qu'il a versé  $p_0$  euros en  $t = 0$ .

Quel est l'intérêt d'une telle opération ? Pourquoi l'acheteur est-il prêt à payer  $p_0$  euros en  $t = 0$  dans la perspective de recevoir  $(S_T - K)_+$  euros en  $t = T$  ?

Intérêt spéculatif : Si  $(S_T - K)_+ > p_0$  l'acheteur aura au final gagné de l'argent. Il peut se lancer dans une telle opération s'il anticipe que en  $t = T$  les cours de l'action  $S$  seront élevés (on parle de scénario haussier).

Couverture d'un risque : Si  $S$  est le cours d'un baril de pétrole par exemple. Une compagnie aérienne qui a besoin constamment de kérosène peut acheter des options Call pour se garantir un prix d'achat à  $K$  euros le baril. Elle se prémunit ainsi contre le risque de voir soudain le prix du pétrole monter en flèche en  $t = T$ , si elle anticipe un problème sur le marché ou un besoin pour son fonctionnement à cette date.

Dans la suite du chapitre le point de vue de l'acheteur va être évacué. Un ingénieur financier a en effet principalement pour mission de proposer des solutions pour le vendeur de l'option.

**Point de vue du vendeur.** Il reçoit  $p_0$  euros en  $t = 0$ .

En  $t = T$  tout se passe comme s'il devait fournir la richesse  $(S_T - K)_+$  euros à l'acheteur (en effet par exemple si  $S_T > K$ , le Call s'exerce ; si le vendeur donne  $S_T - K$  euros à l'acheteur, ça permet bien à ce dernier d'acheter à  $K$  euros l'actif qui en vaut  $S_T$  sur le marché à cet instant).

Cela soulève deux problèmes :

1) Le problème du prix du Call (problème de la valorisation ou du pricing) : Quel est le montant  $p_0$  à faire payer à l'acheteur en  $t = 0$  pour accepter de lui fournir  $(S_T - K)_+$  en  $t = T$  ?

2) Le problème de la couverture (ou du hedging) : partant des  $p_0$  euros que l'acheteur lui a donnés en  $t = 0$  comment le vendeur va-t-il gérer au mieux cette somme d'argent initiale pour avoir à sa disposition  $(S_T - K)_+$  euros en  $t = T$  ?

En fait on va voir dans la suite du chapitre que ces deux problèmes sont intrinsèquement liés : en fait le prix d'une option c'est en quelque sorte le prix de sa couverture.

On peut se risquer à une première définition du prix d'une option payant  $h$  en  $t = T$  ( $h$  est en toute généralité une fonction de la trajectoire  $\{S_t, 0 \leq t \leq T\}$ ) :

le prix à l'instant  $t = 0$  d'une option payant la richesse  $h$  en  $t = T$  c'est la quantité d'argent  $p_0$  telle que, partant de la richesse  $p_0$  en  $t = 0$ , et sans apport d'argent extérieur entre  $t = 0$  et  $t = T$ , on est presque sûr (i.e. avec probabilité 1) d'avoir à sa disposition la richesse  $h$  en  $t = T$ .

**Remarque 5.1.1.** Le notion de "sans apport d'argent extérieur" sera formalisée très précisément dans la section 5.3 ci-après, en introduisant le concept de "portefeuille autofinancé", comportant une part investie dans l'actif  $S$  sur lequel porte l'option, et une part investie dans un "actif sans risque" (en fait des euros qu'on place ou emprunte à la banque au taux sans risque).

Notons que bien sûr, pour vivre, le vendeur de l'option la vend à son juste prix  $p_0$  plus une marge. Mais on ne va pas rentrer dans ces considérations ici...

Le but du chapitre est de développer une théorie mathématique, basée sur des martingales à temps discret, pour calculer très précisément  $p_0$ .

Avant de rentrer dans les modèles à partir de la section 5.3, on va dans la prochaine section évoquer la notion d'absence d'opportunité d'arbitrage, qui reste assez indépendante du modèle qu'on pourrait utiliser.

## 5.2 Absence d'opportunité d'arbitrage (AOA)

**Définition 5.2.1.** On dit qu'il y a possibilité d'arbitrage si, partant d'une richesse initiale  $V(0) = 0$  et sans apport d'argent entre  $t = 0$  et  $t = T$ , on peut disposer en  $t = T$  d'une richesse  $V(T) \geq 0$  p.s. avec  $\mathbb{P}(V(T) > 0) > 0$ .

Dans la plupart des modèles on travaille sous l'hypothèse d'AOA. Cela assure une cohérence mathématique (cf section 5.5) en plus d'être financièrement satisfaisant. En effet s'il y avait de telles opportunités d'arbitrage tout le monde se précipiterait dessus et ferait de l'argent à partir de rien. Le marché est supposé éliminer de lui-même de telles opportunités (par exemple s'il y a la possibilité d'acheter une première entité très bon marché et de l'échanger tout de suite contre une seconde plus chère, ça va faire monter le prix de la première ; il va y avoir retour à l'équilibre).

On a le résultat suivant.

**Théorème 5.2.1.** *En AOA, si les valeurs de deux portefeuilles coïncident à une date donnée ( $V_1(T) = V_2(T)$  pour  $0 < T < \infty$ ) alors elles coïncident à toute date intermédiaire antérieure ( $V_1(t) = V_2(t)$  pour tout  $t \leq T$ ).*

Dans cette section on fera intervenir le facteur de capitalisation au taux sans risque  $r$  garanti par la banque : 1 euros placé en  $t$  en vaut  $e^{r(T-t)}$  en  $T$  (noter que la fonction  $f(s) = e^{r(s-t)}$  satisfait  $f(t) = 1$  et  $df(s) = rf(s) ds$ ). On notera aussi  $B(t, T) = e^{-r(T-t)}$  le prix en  $t$  du zéro-coupon d'échéance  $T$  : c'est un titre émis par la banque et qui vaut 1 euros en  $T$  de façon certaine.

*Démonstration du théorème 5.2.1.* On a  $V_1(T) = V_2(T)$  pour un certain  $0 < T < \infty$ . Soit  $t < T$ , supposons que  $V_1(t) > V_2(t)$  et montrons que cela conduit à une opportunité d'arbitrage.

En  $t$  on se constitue un portefeuille global comportant :  $-1$  part de  $V_1$ ,  $1$  part de  $V_2$  et  $V_1(t) - V_2(t)$  euros. Notons que la valeur globale de ce portefeuille en  $t$  est bien de  $0$  euros.

Que signifie en pratique se constituer ce portefeuille global ? Eh bien : on emprunte une part de  $V_1$  (à une tierce partie ; on a une dette, il faudra rendre cette part empruntée quand on dénouera notre position en  $T$ ), on la vend aussitôt et on empoche  $V_1(t)$  euros. Avec cet argent on achète une part de  $V_2$  à  $V_2(t)$  euros. Il nous reste  $V_1(t) - V_2(t)$  euros qu'on place à la banque au taux sans risque  $r$ .

En  $T$  notre portefeuille global vaut :  $V_2(T) - V_1(T) = 0$ , qui correspond à la valeur de  $-1$  part de  $V_1$  et  $1$  part de  $V_2$ , auquel il faut rajouter  $(V_1(t) - V_2(t))e^{r(T-t)} > 0$  (c'est l'argent placé en  $t$  qui a capitalisé au taux  $r$  entre  $t$  et  $T$ ).

On a donc finalement  $(V_1(t) - V_2(t))e^{r(T-t)} > 0$  euros en  $T$  (c'est ce le "cash" qui nous reste ; en  $T$  on a vendu notre part de  $V_2$  pour acheter aussitôt une part de  $V_1$  et la restituer pour solder notre dette).

Or on était parti de rien : c'est typiquement une opportunité d'arbitrage. Comme on est en AOA c'est que l'hypothèse  $V_1(t) > V_2(t)$  était absurde (on si on veut on raisonne par contraposition). C'est donc que  $V_1(t) \leq V_2(t)$ , pour tout  $t \leq T$ .

En intervertissant les rôles de  $V_1$  et  $V_2$  on obtient  $V_1(t) \geq V_2(t)$  pour tout  $t \leq T$ .

On a donc montré  $V_1(t) = V_2(t)$  pour tout  $t \leq T$ . □

La proposition suivante va s'obtenir à partir du théorème et établir la "relation de parité Call-Put". Un option Put est une option de vente : elle donne le droit à son détenteur de vendre le sous-jacent au prix  $K$  à l'échéance  $T$ . Son pay-off est donc  $(K - S_T)_+$  (pour le voir on peut s'inspirer du raisonnement fait sur le Call en section 5.1).

**Proposition 5.2.1** (Relation de parité Call-Put). *Soit  $(S_t)$  le processus donnant le cours d'un sous-jacent.*

*Soit  $0 < T < \infty$  une maturité et  $K > 0$  un strike.*

*On note  $C_t$  le prix à l'instant  $t \leq T$  du Call (de strike  $K$  et de maturité  $T$ ) portant sur  $S$ .*

*On note  $P_t$  le prix à l'instant  $t \leq T$  du Put (de strike  $K$  et de maturité  $T$ ) portant sur  $S$ .*

*Alors on a pour tout  $t \leq T$*

$$C_t - P_t = S_t - Ke^{-r(T-t)}.$$

*Démonstration.* Pour utiliser le théorème 5.2.1 on définit deux portefeuilles :

$V_1$  est constitué d'1 part de Call (position longue, on le possède) et de  $-1$  part de Put (position courte, on l'a vendu, on va devoir en honorer le pay-off).

$V_2$  est constitué de 1 part d'actif  $S$  et de  $-K$  parts de zéro-coupon (cela signifie qu'on a emprunté  $KB(t, T) = Ke^{-r(T-t)}$  euros à la banque au taux  $r$ ).

On a clairement  $V_2(T) = S_T - K$ . Pour  $V_1$  on a, en utilisant  $(-a)_+ = (a)_-$  pour tout  $a \in \mathbb{R}$ ,

$$V_1(T) = C_T - P_T = (S_T - K)_+ - (K - S_T)_+ = (S_T - K)_+ - (S_T - K)_- = S_T - K$$

(on a aussi utilisé le fait qu'à l'échéance  $T$  la valeur des options Call et Put est celle de leur pay-off). Donc  $V_1(T) = V_2(T)$  et par le théorème on a pour tout  $t \leq T$ ,

$$C_t - P_t = V_1(t) = V_2(t) = S_t - Ke^{-r(T-t)}.$$

Le résultat est prouvé. □

## 5.3 Modèles financiers à temps discret, exemple du modèle Cox-Ross-Rubinstein (CRR), portefeuille autofinancé

### 5.3.1 Modèle financier à temps discret, exemple du modèle CRR

En toute généralité les modèles considérés dans ce chapitre et en TD vérifierons les hypothèses communes suivantes. Pour simplifier on considère des modèles à un seul actif risqué.

Un horizon temporel  $N \in \mathbb{N}$  est fixé. On a un actif sans risque de valeur  $S^0 = (S_n^0)_{0 \leq n \leq N}$ , donné par  $S_n^0 = (1 + r)^n$  pour tout  $0 \leq n \leq N$ . Noter que la dynamique de  $S^0$  est déterministe (les modèles avec une dynamique de l'actif sans risque aléatoire existent mais on ne les rencontrera pas dans ce chapitre).

Cet actif sans risque va nous permettre de représenter nos emprunts et placements à la banque (au taux sans risque  $r$ ).

Puis on a un actif risqué, ou sous-jacent, dont le cours est donné par le processus  $S = (S_n)_{0 \leq n \leq N}$ , à valeurs dans  $\mathbb{R}_+^*$ , et défini sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . Ici  $|\Omega| < \infty$ , on a  $\mathcal{F} = \mathcal{P}(\Omega)$  et la mesure de probabilités  $\mathbb{P}$ , appelée "mesure de probabilités historique", vérifie  $\mathbb{P}(\omega) > 0$  pour tout  $\omega \in \Omega$ . La loi de  $S$  sous  $\mathbb{P}$  est censée se rapprocher au mieux de ce qu'on observe sur le marché. Par la suite on cherchera à mettre en lumière une mesure probabilités risque-neutre  $\mathbb{P}^*$ , qui sera a priori différente de  $\mathbb{P}$ , mais sera celle qui intervient dans notre formule de valorisation.

On a une filtration  $(\mathcal{F}_n)_{0 \leq n \leq N}$  sur  $(\Omega, \mathcal{F}, \mathbb{P})$  vérifiant  $\mathcal{F}_0 = \{\emptyset, \Omega\}$  et  $\mathcal{F}_N = \mathcal{F}$ . Le processus  $(S_n)$  est  $(\mathcal{F}_n)$ -adapté. Noter que ça implique que  $S_0$  est déterministe et constante.

**Exemple du modèle CRR.** Le modèle CRR nous servira très souvent d'exemple. Il rentre dans le cadre général défini ci-dessus et est construit de la façon suivante.

On a  $-1 < a < b$ . On définit  $\Omega = \{1+a, 1+b\}^N$  et  $\mathcal{F} = \mathcal{P}(\Omega)$ .

On a une suite  $(T_n)_{n=1}^N$  définie sur  $(\Omega, \mathcal{F})$  et à valeurs dans  $\{1+a, 1+b\}$ , définie par  $T_n(\omega) = \omega_n$  pour tout  $\omega = (\omega_1, \dots, \omega_N) \in \Omega$ .

Sur  $(\Omega, \mathcal{F})$  on a une probabilité historique  $\mathbb{P}$  telle que la suite  $(T_n)_{n=1}^N$  est i.i.d. sous  $\mathbb{P}$  avec

$$\begin{aligned} \mathbb{P}(T_1 = 1+a) &= \mathbb{P}(\{\omega \in \Omega : \omega_1 = 1+a\}) \\ &= \mathbb{P}(\{(1+a, \omega_2, \dots, \omega_N), \omega_i \in \{1+a, 1+b\}, \forall 2 \leq i \leq N\}) \\ &= p \end{aligned}$$

avec  $0 < p < 1$ . Sur l'existence d'une telle mesure de probabilités  $\mathbb{P}$  cf l'Exercice 1 de la Feuille de TD 9. Dans l'exercice ce n'est pas la mesure de probabilités historique qu'on cherche à construire mais la mesure de probabilités risque-neutre. Mais on peut adapter la démarche (en remplaçant en particulier le  $p(1+a) = (b-r)/(b-a)$  de l'exercice par  $p$ ).

Le fait que  $0 < p < 1$  entraîne que  $\mathbb{P}(\omega) > 0$  pour tout  $\omega \in \Omega$ .

On note  $(\mathcal{F}_n)_{0 \leq n \leq N}$  la filtration définie par  $\mathcal{F}_0 = \{\emptyset, \Omega\}$  et

$$\forall 1 \leq n \leq N, \quad \mathcal{F}_n = \sigma(T_i, 1 \leq i \leq n).$$

Noter qu'on a bien  $\mathcal{F}_N = \mathcal{P}(\Omega)$ .

On peut alors définir la dynamique du processus  $(S_n)$ , défini sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . La valeur initiale  $S_0$  de l'actif risqué est déterministe. Puis on a

$$S_{n+1} = T_{n+1}S_n, \quad \forall 0 \leq n \leq N-1.$$

On a bien que  $(S_n)$  est  $(\mathcal{F}_n)$ -adapté (en fait on a mieux : pour tout  $n$  on a  $\mathcal{F}_n = \sigma(S_k, 1 \leq k \leq n)$ , la filtration  $(\mathcal{F}_n)$  est la filtration naturelle de  $S$ ).

### 5.3.2 Portefeuille autofinancé

On considère un portefeuille (ou stratégie) constitué à l'instant  $n$  d'une quantité  $H_n^0$  d'actif sans risque et d'une quantité  $H_n^1$  d'actif risqué. On notera par la suite  $H = (H_n)_{0 \leq n \leq N}$  le processus à valeurs vectorielles (ici à valeurs dans  $\mathbb{R}^2$ ) défini par

$$H_n = \begin{pmatrix} H_n^0 \\ H_n^1 \end{pmatrix}, \quad \forall 0 \leq n \leq N.$$

La valeur du portefeuille à l'instant  $0 \leq n \leq N$  est donnée par

$$V_n = V_n(H) = H_n^0 S_n^0 + H_n^1 S_n = H_n \cdot S_n.$$

Ici le  $\cdot$  désigne le produit scalaire dans  $\mathbb{R}^2$  et on a noté  $S_n = (S_n^0, S_n^1)^T$  (noter qu'on emploie la même lettre pour le vecteur  $S_n$  et sa deuxième composante, mais il n'y a pas d'ambiguïté possible car on a pour la quantité scalaire une multiplication par  $H_n^1$  et pour la quantité vectorielle un produit scalaire par  $H_n$ ).



Plus précisément  $H_n^0$  (resp.  $H_n^1$ ) est la quantité d'actif  $S^0$  (resp.  $S$ ) détenue dans le portefeuille sur l'intervalle de temps  $(n-1, n]$  (pour  $0 < n \leq N$ ;  $H_0$  donne la composition initiale du portefeuille, mais qu'on va changer aussitôt). C'est à dire qu'à l'instant  $n-1$  on décide de la composition  $H_n$  du portefeuille sur  $(n-1, n]$  : en particulier  $(H_n^0)$  et  $(H_n^1)$  sont donc des processus  $(\mathcal{F}_n)$ -prévisibles.

On dit du portefeuille qu'il est autofinancé si les recompositions de portefeuille se font à valeur constante (du portefeuille); c'est à dire si pour tout  $0 \leq n \leq N-1$

$$H_n \cdot S_n = H_n^0 S_n^0 + H_n^1 S_n = H_{n+1}^0 S_n^0 + H_{n+1}^1 S_n = H_{n+1} \cdot S_n \quad (5.3.1)$$

Dans cette expression  $H_n \cdot S_n$  représente la valeur du portefeuille juste avant le rebalancement (ou recomposition) à l'instant  $n$ . Et  $H_{n+1} \cdot S_n$  représente cette valeur juste après le rebalancement (la valeur des actifs présents dans le portefeuille n'a pas encore changé, mais sa composition si).

Notons que par (5.3.1) on a

$$V_{n+1} - V_n = H_{n+1} \cdot S_{n+1} - H_n \cdot S_n = H_{n+1} \cdot S_{n+1} - H_{n+1} \cdot S_n = H_{n+1} \cdot \Delta S_{n+1},$$

où on a noté  $\Delta S_{n+1} = (S_{n+1}^0 - S_n^0, S_{n+1} - S_n)^T$ . En d'autres termes on a,

$$\Delta V_n = H_n \cdot \Delta S_n, \quad \forall 1 \leq n \leq N. \quad (5.3.2)$$

**Remarque 5.3.1.** L'équation (5.3.2) est à mettre en regard de la définition de l'autofinancement qui sera faite pour les modèles à temps *continu*, et qui sera du type  $dV_t = H_t dS_t$  (cf Introduction aux Produits Dérivés en 2A, Gestion Dynamique des Risques Financiers en 3A,...).

Actualisation : On note

$$\tilde{S}_n^0 = \frac{S_n^0}{S_n^0} \equiv 1, \quad \tilde{S}_n = \frac{S_n}{S_n^0}, \quad \tilde{V}_n = \frac{V_n}{S_n^0} \quad \text{et} \quad \tilde{S} = (\tilde{S}^0, \tilde{S})^T$$

(toutes ces quantités sont exprimées en actif sans risque, au lieu de l'être en euro). La quantité  $\tilde{S}_n$  est la "valeur de l'actif risqué actualisée" (au temps  $n$ ) et  $\tilde{V}_n$  est la "valeur du portefeuille actualisée" (au temps  $n$ ).

En divisant (5.3.1) par  $S_n^0$  on trouve

$$H_n \cdot \tilde{S}_n = H_{n+1} \cdot \tilde{S}_n$$

puis, comme dans le cas non actualisé,

$$\Delta \tilde{V}_{n+1} = H_{n+1} \cdot \Delta \tilde{S}_{n+1}.$$

On peut résumer les choses en énonçant la proposition suivante.

**Proposition 5.3.1.** *Il y a équivalence entre :*

- 1) *Le portefeuille  $H = (H^0, H^1)^T$  est autofinancé.*
- 2) *On a  $\Delta V_n(H) = H_n \cdot \Delta S_n$ ,  $1 \leq n \leq N$ .*
- 3) *On a  $\Delta \tilde{V}_n(H) = H_n \cdot \Delta \tilde{S}_n$ ,  $1 \leq n \leq N$ .*

En effet on a déjà montré par les discussions précédentes que 1) implique 2) et 3).

Pour voir par exemple que 2) implique 1) il suffit de remarquer que

$$\begin{aligned} H_{n+1} \cdot S_n - H_n \cdot S_n &= H_{n+1} \cdot (S_n - S_{n+1} + S_{n+1}) - H_n \cdot S_n \\ &= -\Delta V_{n+1} + H_{n+1} \cdot S_{n+1} - H_n \cdot S_n = 0. \end{aligned}$$

**Remarque 5.3.2.** Une stratégie autofinancée  $((H_n^0, H_n^1)^T)_{0 \leq n \leq N}$  est définie de façon équivalente par la donnée de  $V_0(H)$  et  $(H_n^1)_{0 \leq n \leq N}$ .

En effet si on connaît toute la trajectoire de  $(H^0, H^1)^T$  on a accès à  $V_0(H) = H_0^0 + H_0^1 S_0$  et à  $(H_n^1)_{0 \leq n \leq N}$ , et remarquons que ça permet de calculer pour tout  $1 \leq n \leq N$  la quantité  $\tilde{V}_n(H) = V_0(H) + \sum_{k=1}^n H_k^1 (\tilde{S}_k - \tilde{S}_{k-1})$ ; le même raisonnement vaut pour  $V_n(H)$ .

(noter qu'on a noté  $H_k^1(\tilde{S}_k - \tilde{S}_{k-1})$  et non  $H_k \cdot (\tilde{S}_k - \tilde{S}_{k-1})$ ; c'est que  $H_k^1(\tilde{S}_k - \tilde{S}_{k-1}) = H_k \cdot (\tilde{S}_k - \tilde{S}_{k-1})$ , puisque  $H_k^0(\tilde{S}_k^0 - \tilde{S}_{k-1}^0) = 0$ .)

Réciproquement, si on connaît  $V_0$  et  $(H_n^1)_{0 \leq n \leq N}$ , on considère  $\tilde{v}_n = V_0 + \sum_{k=1}^n H_k^1(\tilde{S}_k - \tilde{S}_{k-1})$ . Vérifions qu'on peut construire  $H^0$  processus prévisible tel que la stratégie  $(H^0, H^1)^T$  est autofinancée et qu'on a  $H_n^0 + H_n^1 \tilde{S}_n = \tilde{v}_n$  pour tout  $0 \leq n \leq N$  (en version non actualisée on aura donc  $H_n^0 S_n^0 + H_n^1 S_n = \tilde{v}_n S_n^0 =: v_n$ ; et en particulier  $H_0^0 S_0^0 + H_0^1 S_0 = V_0$ ).

Il suffit de poser

$$H_n^0 = \tilde{v}_n - H_n^1 \tilde{S}_n, \quad \forall 0 \leq n \leq N.$$

Remarquons que le processus  $H^0$  ainsi défini est prévisible. En effet  $H_n^0 = V_0 + \sum_{k=1}^{n-1} H_k^1 \Delta \tilde{S}_k - H_n^1 \tilde{S}_{n-1}$ , d'où on voit que  $H_n^0$  est  $\mathcal{F}_{n-1}$ -mesurable pour tout  $1 \leq n \leq N$  (et  $H_0^0$  est déterministe par définition, donc  $\mathcal{F}_0$ -mesurable).

Par construction la stratégie  $(H^0, H^1)^T$  vérifie  $\tilde{V}_n(H) = H_n^0 + H_n^1 \tilde{S}_n = \tilde{v}_n$  et  $V_n(H) = H_n^0 S_n^0 + H_n^1 S_n = v_n$  et est autofinancée par la proposition 5.3.1.

Moralité : la valeur initiale du portefeuille et la quantité d'actif risqué imposent, par autofinancement, la quantité d'actif sans risque, c'est à dire ce qu'il nous faut emprunter ou placer à la banque.

Cette remarque sera utilisée à diverses reprises (notamment dans les démonstrations des théorèmes 5.5.1 et 5.6.1).

**Remarque 5.3.3.** Les valeurs de  $H^0$  et  $H^1$  peuvent être négatives : cela signifie que l'on a une dette (en l'actif sans risque ou risqué). Par contre il n'est pas raisonnable que dans un portefeuille de couverture  $H^0$  et  $H^1$  deviennent négatifs en même temps, d'où la définition suivante.

**Définition 5.3.1.** On dit d'un portefeuille autofinancé  $H$  qu'il est admissible si  $V_n(H) \geq 0$  pour tout  $n$ .

La définition ci-dessus traduit le fait que le détenteur d'un portefeuille admissible peut à tout instant rembourser sa dette (la part négative du portefeuille n'excède pas la part positive, en valeur absolue).

**Remarque 5.3.4.** Dans certains ouvrages (ex : [6]) le terme "stratégie admissible" désigne une stratégie qui est à la fois autofinancée et admissible au sens de la définition 5.3.1. Dans ce polycopié on préfère pour des raisons techniques séparer sémantiquement le caractère autofinancé et le caractère admissible d'une stratégie. On parlera souvent de "stratégie autofinancée et admissible".

On peut maintenant préciser la définition du prix d'une option, déjà évoquée dans la section 5.1.

**Définition 5.3.2.** Le prix d'une option payant  $h$  en  $N$  (en toute généralité  $h$  est une v.a.  $\mathcal{F}_N$ -mesurable) est, à l'instant  $0 \leq n \leq N$ , la valeur  $V_n(H)$  d'un portefeuille autofinancé et admissible tel que  $V_N(H) = h$  p.s. (on dit que le portefeuille réplique le pay-off  $h$ ).

**Remarque 5.3.5.** 1) En particulier le prix à l'instant de vente  $n = 0$  est  $V_0$ . C'est le prix  $p_0$  qu'on cherche à calculer.

2) Le fait que le portefeuille de réplique est autofinancé traduit le "sans apport d'argent extérieur" de la section 5.1. Le caractère admissible est une mesure de sécurité raisonnable.

## 5.4 Mesure de probabilités risque-neutre

**Définition 5.4.1.** Une mesure de probabilités  $\mathbb{Q}$  sur  $(\Omega, \mathcal{F})$  est dite équivalente à  $\mathbb{P}$  si pour tout  $A \in \mathcal{F}$  on a  $\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0$ .

**Remarque 5.4.1.** 1) On note  $\mathbb{P} \sim \mathbb{Q}$ .

2) Dans le cadre discret dans lequel on est placé ici on a en fait  $\mathbb{P} \sim \mathbb{Q} \Leftrightarrow \mathbb{Q}(\omega) > 0, \forall \omega \in \Omega$  (on rappelle que  $\mathbb{P}(\omega) > 0, \forall \omega \in \Omega$ ).

En effet : pour " $\Rightarrow$ " il suffit de remarquer que s'il existe  $\omega \in \Omega$  t.q.  $\mathbb{Q}(\omega) = 0$  alors  $\mathbb{P}(\omega) = 0$  par  $\mathbb{P} \sim \mathbb{Q}$ , ce qui est absurde.

Pour " $\Leftarrow$ " : soit  $A \in \mathcal{F}$ . Si  $\mathbb{Q}(A) = 0$  alors c'est que  $A = \emptyset$  et on a donc  $\mathbb{P}(A) = 0$ . De même on montre que si  $\mathbb{P}(A) = 0$  alors  $\mathbb{Q}(A) = 0$ . D'où  $\mathbb{P} \sim \mathbb{Q}$ .

**Définition 5.4.2.** On appelle mesure de probabilités risque-neutre une mesure  $\mathbb{P}^*$  équivalente à  $\mathbb{P}$  telle que les processus  $(\tilde{S}_n^0)$  et  $(\tilde{S}_n)$  sont des  $(\mathcal{F}_n)$ -martingales sous  $\mathbb{P}^*$ .

**Remarque 5.4.2.** Comme  $\tilde{S}^0 \equiv 1$ , c'est bien sûr une  $(\mathcal{F}_n)$ -martingale sous n'importe quelle mesure de probabilités : la vraie question c'est donc de trouver  $\mathbb{P}^*$  sous laquelle  $\tilde{S}$  est une martingale.

La proposition suivante est au coeur de la théorie du pricing par calcul de martingales.

**Proposition 5.4.1.** Soit  $\mathbb{P}^*$  une mesure de probabilités risque-neutre et  $H = (H^0, H^1)^T$  une stratégie autofinancée.

Alors le processus  $(\tilde{V}_n(H))_{0 \leq n \leq N}$  est une  $(\mathcal{F}_n)$ -martingale sous  $\mathbb{P}^*$ .

*Démonstration.* Grâce à la Proposition 5.3.1-3) on a

$$\forall 0 \leq n \leq N, \quad \tilde{V}_n(H) = \tilde{V}_0(H) + \sum_{k=1}^n H_k^1 (\tilde{S}_k - \tilde{S}_{k-1})$$

Or  $(\tilde{S}_n)$  est une  $(\mathcal{F}_n)$ -martingale sous  $\mathbb{P}^*$  et  $(H_n^1)$  est  $(\mathcal{F}_n)$ -prévisible.

Donc d'après la remarque 4.2.2-2) le processus  $(\tilde{V}_n(H))_{0 \leq n \leq N}$  est une  $(\mathcal{F}_n)$ -martingale sous  $\mathbb{P}^*$  (partant de  $\tilde{V}_0(H)$ ).  $\square$

Conséquence : Si on trouve  $\mathbb{P}^*$  une mesure de probabilités risque-neutre et  $H = (H^0, H)^T$  une stratégie autofinancée admissible qui réplique le pay-off  $h$  on a pour tout  $0 \leq n \leq N$ ,

$$V_n(H) = S_n^0 \tilde{V}_n(H) = S_n^0 \mathbb{E}^* [\tilde{V}_N(H) | \mathcal{F}_n] = (1+r)^n \mathbb{E}^* \left[ \frac{V_N(H)}{(1+r)^N} | \mathcal{F}_n \right] = \mathbb{E}^* [(1+r)^{n-N} h | \mathcal{F}_n], \quad (5.4.1)$$

où on a noté  $\mathbb{E}^* := \mathbb{E}_{\mathbb{P}^*}$  (notation conservée pour la suite).

En particulier le prix de vente  $p_0$  en  $n = 0$  du produit payant  $h$  en  $N$  est

$$V_0 = \mathbb{E}^* [(1+r)^{-N} h].$$

Les enjeux maintenant sont : cette mesure de probabilités risque-neutre existe-t-elle ? Est-elle unique (ce qui garantira l'unicité des prix) ? Peut-on trouver une stratégie autofinancée admissible de réplication ?

On va étudier les deux "théorèmes fondamentaux", qui vont apporter des réponses à ces questions.

## 5.5 Premier théorème fondamental

**Définition 5.5.1.** Le modèle (ou le marché) est dit viable, si pour toute stratégie autofinancée et admissible  $H$  avec  $V_0(H) = 0$  on a  $\mathbb{P}(V_N(H) > 0) = 0$  (autrement dit il n'y a pas de stratégie autofinancée admissible d'arbitrage).

**Théorème 5.5.1.** Le marché est viable si et seulement si il existe  $\mathbb{P}^*$  mesure de probabilités risque-neutre sur  $(\Omega, \mathcal{F})$ .

*Démonstration.* " $\Leftarrow$ " : Soit  $H$  autofinancée et admissible t.q.  $V_0(H) = 0$ .

Par la proposition 5.4.1 le processus  $(\tilde{V}_n(H))_{0 \leq n \leq N}$  est une martingale sous  $\mathbb{P}^*$ .

Donc  $\mathbb{E}^*[\tilde{V}_N(H)] = \mathbb{E}^*[V_0(H)] = 0$  (une martingale est constante en espérance).

Comme  $S_N^0 > 0$  et  $V_N(H) \geq 0$  (puisque  $H$  est admissible) on a  $\tilde{V}_N(H) \geq 0$ .

Donc  $\tilde{V}_N(H) = V_N(H)/S_N^0 = 0$ ,  $\mathbb{P}^*$ -p.s. Comme  $S_N^0 > 0$  c'est que  $V_N(H) = 0$ ,  $\mathbb{P}^*$ -p.s., i.e.  $\mathbb{P}^*(V_N(H) > 0) = 0$ .

On a donc  $\mathbb{P}(V_N(H) > 0) = 0$ , puisque  $\mathbb{P}^* \sim \mathbb{P}$ . Donc le marché est viable.

" $\Rightarrow$ " : Ici la preuve va se décomposer en plusieurs étapes. En préambule on a besoin d'introduire quelques notations.

On note  $\Gamma = \{X \text{ v.a. avec } X > 0\}$ .

Comme le marché est viable on a, pour toute stratégie autofinancée admissible  $H$  t.q.  $V_0(H) = 0$ , que  $V_N(H) \notin \Gamma$  et  $\tilde{V}_N(H) \notin \Gamma$ .

Pour tout  $H^1 = (H_n^1)_{0 \leq n \leq N}$  processus prévisible (à valeurs réelles) on note  $\tilde{G}(H)$  le processus défini par

$$\tilde{G}_0(H) = 0 \quad \text{et} \quad \tilde{G}_n(H) = \sum_{k=1}^n H_k^1 \Delta \tilde{S}_k, \quad \forall 1 \leq n \leq N.$$

Plusieurs remarques :

1)  $\tilde{G}_n(H)$  peut être vue comme la valeur actualisée à l'instant  $n$  d'un portefeuille autofinancé de valeur initiale nulle, contenant  $H_n^1$  parts d'actif risqué à tout instant  $0 \leq n \leq N$  (cf remarque 5.3.2).

2) Pour  $H^1$  prévisible quelconque rien ne dit que la stratégie autofinancée correspondante est admissible.

**Étape 1 :** On m.q. pour tout  $(H_n^1)$  prévisible  $\tilde{G}_N(H) \notin \Gamma$ . La difficulté de ce résultat est qu'on ne sait pas si  $H^1$  correspond à une stratégie admissible.

Supposons  $\tilde{G}_N(H) \in \Gamma$ . Si  $\tilde{G}_n(H) \geq 0, \forall 0 \leq n \leq N$  c'est que la stratégie correspondant à  $H^1$  est autofinancée et admissible et c'est en contradiction avec la viabilité (cf remarque 1) qui précède). Dans ce cas le résultat est donc montré par l'absurde.

Si tel n'est pas le cas on pose

$$n := \sup\{0 \leq k \leq N : \mathbb{P}(\tilde{G}_k(H) < 0) > 0\}.$$

On a  $1 \leq n \leq N - 1, \mathbb{P}(\tilde{G}_n(H) < 0) > 0$  et  $\forall m > n, \tilde{G}_m(H) \geq 0$ .

Posons alors  $A = \{\tilde{G}_n(H) < 0\}$  et définissons  $\psi^1 = (\psi_j^1)_{0 \leq j \leq N}$  par

$$\psi_j^1(\omega) = \begin{cases} 0 & \text{si } j \leq n \\ \mathbf{1}_A(\omega) H_j^1(\omega) & \text{si } j > n. \end{cases}$$

Le processus  $\psi^1$  est prévisible. En effet pour tout  $j \leq n$  la v.a.  $\psi_j^1$  est  $\mathcal{F}_0$ -mesurable, donc  $\mathcal{F}_{j-1}$ -mesurable. Pour  $j \geq n+1$ , la v.a.  $H_j^1$  est  $\mathcal{F}_{j-1}$ -mes. et  $\mathbf{1}_A$  est  $\mathcal{F}_n$ -mes. donc  $\mathcal{F}_{j-1}$ -mes. Donc  $\psi_j^1$  est  $\mathcal{F}_{j-1}$ -mesurable.

Par ailleurs par sommation

$$\tilde{G}_j(\psi) = \begin{cases} 0 & \text{si } j \leq n \\ \mathbf{1}_A(\tilde{G}_j(H) - \tilde{G}_n(H)) & \text{si } j > n. \end{cases}$$

Ainsi  $\tilde{G}_j(\psi) \geq 0, \forall 0 \leq j \leq N$ . En effet pour  $j > n$  on a  $\tilde{G}_j(H) \geq 0$ , et on a  $-\tilde{G}_n(H)\mathbf{1}_A \geq 0$ , car cette dernière quantité vaut zéro sur  $A^c$  et est strictement positive sur  $A$ .

En particulier notons que  $\tilde{G}_N(\psi) > 0$  sur l'évènement  $A$ .

Ainsi  $\psi^1$  définit une stratégie autofinancée et admissible d'arbitrage. À nouveau cela contredit la viabilité et le résultat est montré.

**Étape 2 :** On va utiliser le théorème suivant (pour la preuve, voir dans [6] les théorèmes 8.3.1 et 8.3.2).

**Théorème de Hahn-Banach (forme géométrique).** Soit  $K$  un convexe compact et  $V$  un sous-espace vectoriel de  $\mathbb{R}^n$ , disjoint de  $K$ .

Il existe une forme linéaire  $\xi$  sur  $\mathbb{R}^n$ , vérifiant :

1) Pour tout  $x \in K$ , on a  $\xi(x) > 0$ .

2) Pour tout  $x \in V$ , on a  $\xi(x) = 0$ .

Le sous-espace  $V$  est donc contenu dans un hyperplan qui ne rencontre pas  $K$ .

Ici posons donc  $K = \{X \in \Gamma : \sum_{\omega \in \Omega} X(\omega) = 1\}$ . L'ensemble  $K$  est convexe car pour tous  $X, Y \in K$ , et tout  $t \in [0, 1]$ , on a

$$\sum_{\omega} (tX(\omega) + (1-t)Y(\omega)) = t \sum_{\omega} X(\omega) + (1-t) \sum_{\omega} Y(\omega) = t + 1 - t = 1,$$

ce qui montre que  $tX + (1-t)Y \in K$ .

En outre  $K$  est un compact de  $\mathbb{R}^\Omega$  (dans ce qui suit on fait à chaque fois que nécessaire l'identification entre  $\mathbb{R}^\Omega$  et  $\mathbb{R}^{|\Omega|}$ , ces deux ensembles étant bijection de façon transparente;  $\mathbb{R}^\Omega$  représente l'ensemble des v.a.r définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , les éléments de  $\mathbb{R}^\Omega$  étant bien  $\mathcal{F}$ -mesurables puisque  $\mathcal{F} = \mathcal{P}(\Omega)$ ).

En effet considérons

$$\begin{aligned} \varphi : (\mathbb{R}_+^*)^\Omega &\rightarrow \mathbb{R} \\ X &\mapsto \varphi(X) = \sum_\omega X(\omega) \end{aligned}$$

(noter que  $\Gamma = (\mathbb{R}_+^*)^\Omega$ ). L'application  $\varphi$  est continue et donc  $K = \varphi^{-1}(\{1\})$  est fermé puisque  $\{1\}$  est fermé dans  $\mathbb{R}$ . De plus  $K$  est borné. En effet pour  $X \in K$  on a  $\sum_\omega X(\omega) = 1$  et  $X(\omega) \geq 0$  pour tout  $\omega$ , ce qui implique  $0 \leq X(\omega) \leq 1$  pour tout  $\omega$ .

Comme  $K$  peut être vu comme un fermé borné de  $(\mathbb{R}_+^*)^\Omega$  il est compact.

On définit maintenant

$$V = \{\text{v.a. de la forme } \tilde{G}_N(H) \text{ avec } H^1 \text{ prévisible}\}.$$

Comme  $\tilde{G}_N(H) + a\tilde{G}_N(\psi) = \tilde{G}_N(H + a\psi)$ , pour tous  $H^1, \psi^1$  prévisibles et tout  $a \in \mathbb{R}$ , on a que  $V$  est un sous-e.v. de  $\mathbb{R}^\Omega$ .

Or  $K$  et  $V$  ne se rencontrent pas, par l'étape 1 (et puisque  $K \subset \Gamma$ ).

On peut alors utiliser le théorème de Hahn-Banach : il existe  $\xi$  forme linéaire sur  $\mathbb{R}^\Omega$  t.q.  $\xi(X) > 0$  pour tout  $X \in K$  et  $\xi(X) = 0$  pour tout  $X \in V$ .

La forme linéaire  $\xi$  peut se représenter matriciellement à l'aide d'un vecteur  $(\lambda(\omega))_{\omega \in \Omega}$  t.q.  $\xi(X) = \sum_\omega \lambda(\omega)X(\omega)$ .

On a donc  $(\lambda(\omega))$  avec

- 1)  $\sum_\omega \lambda(\omega)X(\omega) > 0$  pour tout  $X \in K$ .
- 2)  $\sum_\omega \lambda(\omega)\tilde{G}_N(H)(\omega) = 0$  pour tout  $H^1$  prévisible.

La condition 1) ci-dessus implique que  $\lambda(\omega) > 0$  pour tout  $\omega$  (il suffit de considérer,  $\bar{\omega}$  par  $\bar{\omega}$ , l'élément  $X \in K$  défini par  $X(\omega) = \mathbf{1}_{\omega=\bar{\omega}}$ ,  $\omega \in \Omega$ ).

On définit donc  $\mathbb{P}^*$  par

$$\mathbb{P}^*(\omega) = \frac{\lambda(\omega)}{\sum_{\omega' \in \Omega} \lambda(\omega')}.$$

Comme  $\mathbb{P}^*(\omega) > 0$  pour tout  $\omega$  on a  $\mathbb{P}^* \sim \mathbb{P}$  (cf remarque 5.4.1-2)).

Reste donc à vérifier que  $(\tilde{S}_n)$  est une  $(\mathcal{F}_n)$ -martingale sous  $\mathbb{P}^*$ .

On a

$$\mathbb{E}^*[\tilde{G}_N(H)] = \frac{1}{\sum_{\omega' \in \Omega} \lambda(\omega')} \sum_\omega \lambda(\omega)\tilde{G}_N(H)(\omega) = 0,$$

et donc

$$\mathbb{E}^* \left[ \sum_{k=1}^N H_k^1(\tilde{S}_k - \tilde{S}_{k-1}) \right] = 0 \quad (5.5.1)$$

pour tout  $H^1$  prévisible.

Soit  $1 \leq k \leq N-1$  et  $A \in \mathcal{F}_k$ , définissons  $H^1$  par  $H_{k+1}^1 = \mathbf{1}_A$  et  $H_n^1 = 0$  pour  $n \neq k+1$ .

Le processus  $H^1$  est bien prévisible car  $H_n^1$  est  $\mathcal{F}_0$ -mesurable donc  $\mathcal{F}_{n-1}$ -mesurable pour tout  $1 \leq n \leq N$  avec  $n \neq k+1$  (et  $H_0^1$  est bien  $\mathcal{F}_0$ -mesurable). Et  $H_{k+1}^1$  est bien  $\mathcal{F}_k$ -mesurable.

En utilisant ce processus  $H^1$  dans (5.5.1) il vient  $\mathbb{E}^*[\mathbf{1}_A(\tilde{S}_{k+1} - \tilde{S}_k)] = 0$ .

On a donc montré que  $\mathbb{E}^*[\mathbf{1}_A\tilde{S}_{k+1}] = \mathbb{E}^*[\mathbf{1}_A\tilde{S}_k]$ ,  $\forall A \in \mathcal{F}_k$ , or  $\tilde{S}_k$  est  $\mathcal{F}_k$ -mesurable, donc par définition de l'espérance conditionnelle il vient  $\mathbb{E}^*[\tilde{S}_{k+1} | \mathcal{F}_k] = \tilde{S}_k$ .

Or  $k$  était quelconque dans  $\{1, \dots, N-1\}$ . On a donc montré que  $(\tilde{S}_n)$  est une  $(\mathcal{F}_n)$ -martingale sous  $\mathbb{P}^*$ .  $\square$

## 5.6 Second théorème fondamental

**Définition 5.6.1.** On dit que le marché est complet si pour toute v.a.  $h \geq 0$  (p.s.),  $\mathcal{F}_N$ -mesurable, il existe  $H = (H^0, H^1)^T$  stratégie autofinancée admissible telle que  $V_N(H) = h$  p.s.

**Remarque 5.6.1.** Si le marché est viable et que  $H = (H^0, H^1)^T$  autofinancée réplique  $h \geq 0$  alors  $H$  est automatiquement admissible. En effet il existe alors  $\mathbb{P}^*$  mesure de probabilités risque-neutre, et même si on ne sait pas si  $H$  est admissible on peut appliquer le raisonnement qui mène à (5.4.1). On a donc

$$\forall 0 \leq n \leq N, \quad V_n(H) = \mathbb{E}^*[(1+r)^{n-N}h | \mathcal{F}_n] \geq 0,$$

par la Propriété 2.2.1-ii).

**Théorème 5.6.1.** *Un marché viable est complet si et seulement si il existe une unique mesure de probabilités risque-neutre sur  $(\Omega, \mathcal{F})$ .*

*Démonstration.* La viabilité assure l'existence d'une mesure de probabilités risque-neutre, on se préoccupe donc uniquement de son unicité.

" $\Rightarrow$ ": Soient  $\mathbb{P}_1$  et  $\mathbb{P}_2$  deux mesures de probabilités risque-neutre, on va montrer que la complétude implique  $\mathbb{P}_1 = \mathbb{P}_2$ .

Soit  $\bar{\omega} \in \Omega$  et  $h(\bar{\omega}) = \mathbf{1}_{\omega=\bar{\omega}}$ . Il existe  $H$  stratégie autofinancée admissible telle que  $V_N(H) = h$ ,  $\mathbb{P}$ ,  $\mathbb{P}_1$  et  $\mathbb{P}_2$ -p.s. (car  $\mathbb{P} \sim \mathbb{P}_1 \sim \mathbb{P}_2$ ).

Or  $(\tilde{V}_n(H))_{0 \leq n \leq N}$  est une martingale sous  $\mathbb{P}_1$  et  $\mathbb{P}_2$  (prop. 5.4.1) donc

$$\mathbb{E}_1[\tilde{V}_N(H)] = \mathbb{E}_1[\tilde{V}_0(H)] = V_0(H) = \mathbb{E}_2[\tilde{V}_N(H)]$$

car  $\tilde{V}_0(H) = V_0(H)$  est déterministe.

Or on a  $\mathbb{E}_1[\tilde{V}_N(H)] = \mathbb{E}_1[h/S_N^0] = \mathbb{P}_1(\bar{\omega})/S_N^0(\bar{\omega})$  et de la même façon  $\mathbb{E}_2[\tilde{V}_N(H)] = \mathbb{P}_2(\bar{\omega})/S_N^0(\bar{\omega})$ . On a donc  $\mathbb{P}_1(\bar{\omega}) = \mathbb{P}_2(\bar{\omega})$ .

Or  $\bar{\omega}$  était quelconque dans  $\Omega$ , on a donc montré  $\mathbb{P}_1 = \mathbb{P}_2$ .

" $\Leftarrow$ ": Soit  $\mathbb{P}_1$  mesure de probabilités risque-neutre. Supposons le marché viable et non complet et montrons qu'il existe  $\mathbb{P}_2$ , une autre mesure de probabilités risque-neutre, avec  $\mathbb{P}_2 \neq \mathbb{P}_1$ . Par contraposition on aura le résultat voulu.

Considérons

$$\mathcal{V} = \left\{ V_0 + \sum_{n=1}^N H_n^1 \Delta \tilde{S}_n \text{ avec } (H_n^1) \text{ prévisible, } V_0 \in \mathbb{R} \right\}.$$

C'est (cf remarque 5.3.2) le sous-e.v. de  $\mathbb{R}^\Omega$  constitué des valeurs terminales actualisées de toutes les stratégies autofinancées possibles (partant en outre de toutes les valeurs initiales possibles). Notons que les éléments de  $\mathcal{V}$  ne correspondent pas forcément à des stratégies admissibles.

Dire que le marché est non complet c'est dire qu'il existe  $h \geq 0$ ,  $\mathcal{F}_N$ -mesurable, qu'on ne peut pas toucher avec la valeur terminale d'un portefeuille autofinancé et admissible. En tenant compte de la remarque 5.6.1 c'est qu'il existe une telle v.a.  $h$  qu'on ne peut pas toucher avec la valeur terminale d'un portefeuille autofinancé. C'est donc qu'il existe  $h \geq 0$ ,  $\mathcal{F}_N$ -mesurable, avec  $h/S_N^0 \notin \mathcal{V}$ .

(Notons qu'en fait  $h > 0$  car 0 est toujours répliquable par un portefeuille constamment de valeur nulle).

On a donc

$$\mathcal{V} \subsetneq \mathbb{R}^\Omega. \quad (5.6.1)$$

Introduisons sur  $\mathbb{R}^\Omega$  le produit scalaire

$$(X, Y) \mapsto \mathbb{E}_1(XY) = \sum_{\omega} X(\omega)Y(\omega)\mathbb{P}_1(\omega),$$

et considérons  $\mathcal{V}^\perp$  qui est non réduit à zéro grâce à (5.6.1).

Prenons  $X \in \mathcal{V}^\perp$  avec  $X \neq 0$  et posons

$$\mathbb{P}_2(\omega) = \mathbb{P}_1(\omega) \left( 1 + \frac{X(\omega)}{2x} \right), \quad \forall \omega \in \Omega,$$

où  $x := \max_{\omega} |X(\omega)|$ . On a

$$\sum_{\omega} \mathbb{P}_2(\omega) = \sum_{\omega} \mathbb{P}_1(\omega) + \frac{1}{2x} \mathbb{E}_1(X) = 1 + \frac{1}{2x} (X, 1).$$

Or  $1 \in \mathcal{V}$ . En effet il suffit de prendre  $V_0 = 1$  et  $H^1 \equiv 0$  pour toucher le payoff actualisé 1 (remarquer que  $H^1$  constant est bien prévisible).

D'où  $(X, 1) = 0$ , puisque  $X \in \mathcal{V}^\perp$ , et finalement  $\sum_{\omega} \mathbb{P}_2(\omega) = 1$  : l'objet  $\mathbb{P}_2$  est bien une mesure de probabilités.

On a

$$\forall \omega \in \Omega, \quad \mathbb{P}_2(\omega) \geq \frac{\mathbb{P}_1(\omega)}{2} > 0.$$

Ceci implique  $\mathbb{P}^2 \sim \mathbb{P}$ .

Par ailleurs  $\mathbb{P}_2 \neq \mathbb{P}_1$  car  $X \neq 0$ . Reste donc à voir que  $(\tilde{S}_n)$  est une martingale sous  $\mathbb{P}_2$ .  
Fixons  $1 \leq n \leq N$  et soit  $Y$  v.a.  $\mathcal{F}_{n-1}$ -mesurable. Évaluons,

$$\mathbb{E}_2(Y(\tilde{S}_n - \tilde{S}_{n-1})) = \mathbb{E}_1(Y(\tilde{S}_n - \tilde{S}_{n-1})) + \frac{1}{2x} \mathbb{E}_1(Y(\tilde{S}_n - \tilde{S}_{n-1})X).$$

On a  $\mathbb{E}_1(Y(\tilde{S}_n - \tilde{S}_{n-1})) = 0$  car  $(\tilde{S}_n)$  est une martingale sous  $\mathbb{P}_1$ . Reste à voir que  $Y(\tilde{S}_n - \tilde{S}_{n-1}) \in \mathcal{V}$  pour conclure que  $\mathbb{E}_1(Y(\tilde{S}_n - \tilde{S}_{n-1})X) = 0$ . Mais pour cela il suffit de prendre  $V_0 = 0$  et  $H^1$  défini par  $H_n^1 = Y$  et  $H_k^1 = 0$  pour  $k \neq n$ .

On a donc montré  $\mathbb{E}_2(Y(\tilde{S}_n - \tilde{S}_{n-1})) = 0$  pour tout  $Y$  v.a.  $\mathcal{F}_{n-1}$ -mesurable, ce qui équivaut à  $\mathbb{E}_2(\tilde{S}_n | \mathcal{F}_{n-1}) = \tilde{S}_{n-1}$ . Comme  $n$  avait été choisi quelconque on a le résultat voulu.  $\square$

## 5.7 Conclusion : valorisation et couverture dans le cas d'un marché viable et complet

On travaille le plus souvent sous les hypothèses de viabilité et de complétude (quoique l'étude des "marchés incomplets" constitue un sujet en soi). Donc :

- 1)  $\mathbb{P}^*$  mesure de probabilités risque-neutre existe.
- 2) Elle est unique.
- 3) Pour tout  $h \geq 0$ ,  $\mathcal{F}_N$ -mesurable, il existe  $H = (H^0, H^1)^T$  stratégie autofinancée et admissible t.q.  $V_N(H) = h$   $\mathbb{P}$ -p.s. (remarquons que si on examine la preuve du théorème 5.6.1 on a plutôt l'impression qu'il existe  $V_0$  et  $(H_n^1)_{0 \leq n \leq N}$  qui réalisent ceci, mais c'est équivalent par la remarque 5.3.2).

Moralité : le prix à l'instant  $0 \leq n \leq N$  d'un produit payant  $h$  en  $N$  est bien donné par

$$V_n(H) = \mathbb{E}^* [(1+r)^{n-N} h | \mathcal{F}_n],$$

comme annoncé dans la formule (5.4.1).

Dans le cadre du modèle CRR avec  $a < r < b$  le modèle est viable et complet (cf Exercice 1 de la fiche de TD 9).

Pour certains pay-off  $h$ , de la forme  $h = h(S_N)$  (on parle d'option vanille), le prix est donc

$$V_n(H) = \mathbb{E}^* [(1+r)^{n-N} h(S_N) | \mathcal{F}_n].$$

On peut dans ce cas donner une formule explicite pour le prix, ainsi que pour la quantité d'actif risqué  $H_n^1$  (cf Exercice 2 de la fiche de TD 9).





# Bibliographie

- [1] P. Barbe and M. Ledoux, *Probabilité (l3m1)*, Enseignement SUP-Maths, EDP Sciences, 2012.
- [2] D. Foata, A. Fuchs, and J. Franchi, *Calcul des probabilités - 3e édition : Cours, exercices et problèmes corrigés*, Mathématiques, Dunod, 2012.
- [3] C. Graham, *Chaînes de markov : cours, exercices et corrigés détaillés*, Mathématiques appliquées pour le Master / SMAI, Dunod, 2008.
- [4] G. Grimmett, P.M.S.G. Grimmett, D. Stirzaker, M.I.D.R. Stirzaker, and S.L.G.R. Grimmett, *Probability and random processes*, Probability and Random Processes, OUP Oxford, 2001.
- [5] J. Jacod and P. Protter, *L'essentiel en théorie des probabilités*, Collection Enseignement des mathématiques, Cassini, 2003.
- [6] D. Lamberton and B. Lapeyre, *Introduction au calcul stochastique appliqué à la finance*, Ellipses Marketing, 1997.
- [7] Steven E. Shreve, *Stochastic calculus for finance I : The binomial asset pricing model : Binomial asset pricing model*, Springer-Verlag, New York, NY, 2003.
- [8] C. Wagschal, *Dérivation, intégration*, Collection Méthodes, Hermann, 1999.