Martingales et Chaînes de Markov

Dominique Bakry, Laure Coutin, Thierry Delmotte
Laboratoire de Statistique et Probabilits,
Universit Paul Sabatier,
118 route de Narbonne,
31062 Toulouse,
FRANCE

10 mai 2004

Table des matières

1 Introduction

Un processus stochastique est une famille de variables aléatoires (X_t) indexées par le temps.Ce temps peut être discret ($t \in \mathbb{N}$) ce sera le cas dans l'étude des martingales et des chaînes de Markov, ou continu ($t \in \mathbb{R}^+$), ce sera le cas des processus de Poisson.

Ce cours comprendra quatre chapitres :

1. Les martingales à temps discret .
2. Les chaînes de Markov à espace d'états fini.
3. Les chaînes de Markov à espace d'états dénombrable ;

4. Les processus de Poisson et les processus de Markov à temps continu et espace d'état fini ou dénombrable.

Il s'agit dans tous les cas de modèles mathématiques rendant compte de phénomènes concrets, physiques ou économiques par exemple. Ces sujets figurent également au programme de l'agrégation, dans l'option orale de mathématiques appliquées.

La théorie des martingales est l'un des outils les plus puissants de la théorie moderne des probabilités. Les chaînes de Markov, quant à elles, sont les modèles les plus utilisés pour représenter les phénomènes aléatoires. Le processus de Poisson est le plus simple des processus de Markov à temps continu, et sert de "brique élémentaire" pour en construire de plus généraux sur les espaces finis ou dénombrables. Il y a bien sûr d'autres modèles, en particulier les processus de Markov à temps continu et à trajectoires continues (les diffusions), dont le modèle le plus simple est le mouvement brownien, mais nous n'en parlerons pas dans ce cours.

2 Plan

1. Quelques rappels : théorème des classes monotones, espérances conditionnelles, intégrabilité uniforme
2. Martingales à temps discret.

   Définitions, exemples, inégalités maximales, théorèmes de convergence, temps d'arrêt, théorème d'arrêt, problème de l'arrêt optimal, théorèmes de convergence, exemple de l'algorithme de Robbins Monro, loi des grands nombres et théorèmes de limite centrale.
3. Chaînes de Markov à espace d'états fini :

   Définitions, exemples, classification des états, mesure invariante, convergence vers la mesure invariante, théorème ergodique, lois des temps d'atteinte ; liens avec les martingales et certains problèmes de résolutions d'EDP, algorithmes MCMC.
4. Chaînes de Markov à espace d'états dénombrables : exemples, récurrence et transience, exemples avec les files d'attente, les marches aléatoires, les processus de vie et de mort.
5. Processus de Poisson :

   Définition par processus de comptage/temps de saut, propriété de Markov, propriété d'accroissements indépendants stationnaires. Construction de Processus de Markov à états finis ou dénombrable à partir du générateur ; description du comportement des trajectoires.

Chapitre 1

Rappels

Dans ce chapitre, nous énonçons en vrac quelques définitions et propriétés dont nous nous servirons dans la suite. Tous ces points sont donnés sans démonstrations. La plupart d'entre elles se trouvent dans le livre [1].

Notation 0.1. *Dans tout ce cours, nous utiliserons la notation $\sigma(E)$ pour désigner la plus petite tribu qui rend mesurable tous les éléments de E, que E soit formé de parties d'un ensemble, de fonctions définies sur cet ensemble, ou bien des deux.*

Ainsi, la tribu borélienne de $\mathbb{R}$ est-elle par exemple $\sigma([a,b];\ a<b;\ a,b\in\mathbb{Q})$.

1 Théorème des classes monotones.

Les théorèmes de classes monotones sont l'un des outils essentiels de la théorie de la mesure. Ils permettent d'étendre facilement à toute une tribu des propriétés établies pour des classes de parties stables par intersection ou des familles de fonctions stables par multiplication. Traditionnellement, ils existent sous forme ensembliste et sous forme fonctionnelle. Nous ne livrons ici que la forme fonctionnelle, qui est de loin la plus souple à l'usage. On trouvera des formes ensemblistes diverses dans le livre [1], par exemple. La forme fonctionnelle présentée ici est extraite du livre [2], sous une forme légèrement différente.

Définition 1.1. *(PCM) Soit Ω un ensemble, et $\mathcal{H}$ un sous-espace vectoriel de l'ensemble des fonctions* **bornées** *définies sur Ω et à valeurs réelles. On dit que $\mathcal{H}$ a la Propriété de Classe Monotone (PCM) si, pour toute suite (f_n)*

d'éléments de $\mathcal{H}$, croissante, positive et bornée, la limite simple f de la suite (f_n) est encore dans $\mathcal{H}$. En d'autres termes,

$$(f_n \in \mathcal{H}\ ;\ \forall n,\ 0 \leq f_n \leq f_{n+1} \leq 1) \Longrightarrow \lim_n f_n \in \mathcal{H}.$$

On notera $\mathcal{M}(\mathcal{A})$ l'espace vectoriel de toutes les fonctions mesurables bornées par rapport à la tribu $\mathcal{A}$, alors le théorème des classes monotones s'énonce

Théorème 1.2. *(TCM) Soit Ω un ensemble et $\mathcal{H}$ un espace vectoriel satisfaisant PCM. Si $\mathcal{C}$ est une partie de $\mathcal{H}$ stable par multiplication, alors*

$$\mathcal{M}(\sigma(\mathcal{C})) \subset \mathcal{H}.$$

Nous donnerons une démonstration de ce théorème en appendice (voir 11.1).

2 Espérance conditionnelle.

Nous donnons dans cette section quelques définitions et propriétés qui nous seront utiles dans la suite de ce cours.

Définition 2.1. *Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité et $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Soit X une variable aléatoire réelle positive (resp. intégrable), $\mathcal{A}$ mesurable. Alors, il existe une variable aléatoire Y positive (resp. intégrable), $\mathcal{B}$-mesurable, unique (à l'égalité presque sûre près), telle que, $\forall B \in \mathcal{B}$,*

$$\mathbf{E}(Y\mathbf{1}_B) = \mathbf{E}(X\mathbf{1}_B). \tag{2.1}$$

Cette variable aléatoire Y est appelée l'espérance conditionnelle de X sachant $\mathcal{B}$ et on note

$$Y = \mathbf{E}(X/\mathcal{B}).$$

Lorsque $\mathcal{B}$ est la triby $\sigma(Z)$, où Z est une variable aléatoire, alors on note plus simplement $\mathbf{E}(X/\mathcal{B}) = \mathbf{E}(X/Z)$. De même, on note $\mathbf{E}(\mathbf{1}_A/\mathcal{B}) = \mathbf{P}(A/\mathcal{B})$. Il faut alors faire attention que c'est une variable aléatoire (qui prend ses valeurs entre 0 et 1) mais pas un nombre : c'est une probabilité aléatoire.

L'espérance conditionnelle se comporte comme l'espérance, à ceci près qu'elle associe à une variable aléatoire non pas un nombre mais une nouvelle variable aléatoire.

La proposition suivante résume les propriétés élémentaires de l'espérance conditionnelle.

Proposition 2.2. *Dans tous les cas (positif ou intégrable), nous avons*

1. *L'application* $X \mapsto \mathbf{E}(X/\mathcal{B})$ *est linéaire.*
2. *Si* $X \leq Y$, $\mathbf{E}(X/\mathcal{B}) \leq \mathbf{E}(Y/\mathcal{B})$, *(ps).*
3. *Pour toute variable aléatoire* X *positive ou intégrable,*
$$\mathbf{E}[\mathbf{E}(X/\mathcal{B})] = \mathbf{E}[X]. \tag{2.2}$$
4. *Plus généralement, si* X *est positive (resp. intégrable), alors, pour toute variable aléatoire* Z, $\mathcal{B}$*-mesurable, positive (resp. bornée), on a*
$$\mathbf{E}[Z\mathbf{E}(X/\mathcal{B})] = \mathbf{E}[ZX]. \tag{2.3}$$
5. *(Conditionnement successifs) Si* $\mathcal{C} \subset \mathcal{B}$ *est une sous-tribu de* $\mathcal{B}$, *alors*
$$\mathbf{E}(\mathbf{E}(X/\mathcal{B})/\mathcal{C}) = \mathbf{E}(X/\mathcal{C}).$$
On notera $\mathbf{E}(\mathbf{E}(X/\mathcal{B})/\mathcal{C}) = \mathbf{E}(X/\mathcal{B}/\mathcal{C})$.

Démonstration. — La propriété de linéarité 1 découle immédiatement de la définition.

La monotonie (propriété 2) est également immédiate ; il faut cependant faire un peu attention lorsque les variables X et Y sont seulement positives, et donc peuvent prendre les valeurs $+\infty$. On utilise la propriété suivante : si deux variables X_1 et Y_1 sont $\mathcal{B}$-mesurables, alors $X_1 \leq Y_1$ si et seulement si, pour tout $B \in \mathcal{B}$,
$$\mathbf{E}(X_1\mathbf{1}_B) \leq \mathbf{E}(Y_1\mathbf{1}_B).$$
Cette caractérisation est aussi bien valable pour des variables positives éventuellement infinies que pour des variables intégrables. On l'applique à $X_1 = \mathbf{E}(X/\mathcal{B})$ et $Y_1 = \mathbf{E}(Y/\mathcal{B})$, puis on applique la définition de l'espérance conditionnelle.

La propriété 3 est immédiate (il suffit d'appliquer la définition avec $B = \Omega$), mais c'est sans doute la propriété la plus utile. En effet, très souvent, pour des variables aléatoires compliquées, il est assez facile de calculer leur espérance conditionnelle par rapport à une sous-tribu judicieusement choisie, et l'espérance de cette espérance conditionnelle est alors simple à calculer.

La propriété 4 est un peu plus délicate. L'équation (2.3) n'est rien d'autre que l'extension de l'équation (2.1) des variables de la forme $Z = \mathbf{1}_B$ aux variables $\mathcal{B}$-mesurables quelconques. C'est une application directe du TCM (1.2).

Pour s'en convaincre, commençons par le cas où Z est intégrable. On désigne alors par $\mathcal{H}$ l'espace vectoriel des variables Z bornées qui vérifient l'équation (2.3) : c'est un espace vectoriel qui a la propriété PCM (définition 1.1), comme on le voit en appliquant le théorème de convergence dominée. On pose $\mathcal{C} = \{\mathbf{1}_B,\ B \in \mathcal{B}\}$. C'est bien une classe stable par multiplication, et la définition de l'éspérance conditionnelle (équation 2.1) nous dit justement que $\mathcal{C} \subset \mathcal{H}$. sûr, $\sigma(\mathcal{C}) = \mathcal{B}$, et donc $\mathcal{H}$ contient donc toutes les variables $\mathcal{B}$-mesurables bornées.

Pour étendre l'équation (2.3) au cas où X et Z sont positives, on commence par remplacer X par $X_n = X \wedge n$ et Z par $Z_n = Z \wedge n$ dans l'équation (2.3). Puis on utilise le théorème de convergence monotone des espérances conditionnelles (voir le point 8 de la proposition 2.4 plus bas), puis on applique le théorème de convergence monotone ordinaire aux deux membres pour passer à la limite.

Pour établir la propriété 5 de conditionnement successifs, il suffit de remarquer que la variable $Y = \mathbf{E}(X/\mathcal{B}/\mathcal{C})$ vérifie les propriétés qu'on attend de $\mathbf{E}(X/\mathcal{C})$, ce qui est immédiat.

Remarquons que la propriété de conditionnement successifs est fausse si $\mathcal{C}$ n'est pas incluse dans $\mathcal{B}$. En général, pour deux sous-tribus quelconques, les espérances conditionnelles sachant $\mathcal{B}$ et sachant $\mathcal{C}$ ne commutent pas : le fait de commuter est équivalent à la propriété d'indépendance conditionnelle de $\mathcal{B}$ et $\mathcal{C}$ par rapport à leur intersection. (Voir l'exercice **??**) et la définition précise de l'indépendance conditionnelle (section 4).

∎

La proposition suivante donne quelques outils élémentaires (mais fort pratiques) pour calculer des espérances conditionnelles.

Proposition 2.3.

1. *Si Z est $\mathcal{B}$-mesurable, alors $\mathbf{E}(Z/\mathcal{B}) = Z$, et plus généralement*

$$\mathbf{E}(XZ/\mathcal{B}) = Z\mathbf{E}(X/\mathcal{B}).$$

2. *Si X est indépendante de $\mathcal{B}$, alors $\mathbf{E}(X/\mathcal{B})$ est constante et vaut $\mathbf{E}(X)$. De plus, si $\mathbf{E}(f(X)/\mathcal{B})$ est constante, pour toute fonction borélienne*

bornée f, alors cette constante vaut $\mathbf{E}(f(X))$ et X est indépendante de $\mathcal{B}$.

3. *Si $\sigma(X, \mathcal{B}_1)$ est indépendante de la tribu $\mathcal{B}_2$, alors $\mathbf{E}(X/\sigma(\mathcal{B}_1, \mathcal{B}_2) = \mathbf{E}(X/\mathcal{B}_1)$.*
4. *Si X est indépendante de $\mathcal{B}$ et si Y est $\mathcal{B}$-mesurable, alors $\mathbf{E}(f(X,Y)/\mathcal{B}) = k(Y)$, où $k(y) = \mathbf{E}(f(X,y))$.*

Démonstration. — Les propriétés 1 et 2 sont immédiates.

Pour démontrer de point 3, on utilise le TCM (1.2). On considère l'équation (4), et on note $\mathcal{H}$ l'ensemble des variables aléatoires mesurables Z bornées pour lesquelles l'équation (4) est vraie. On vérifie immédiatement qu'il satisfait la propriété de classe monotone (PCM) de la définition 1.1. L'ensemble $\mathcal{C}$ est l'ensemble des variables de la forme $Z_1 Z_2$, où Z_1 est $\mathcal{B}_1$-mesurable bornée et Z_2 est $\mathcal{B}_2$-mesurable bornée. L'indépendance de $\sigma(X, \mathcal{B}_1)$ et de $\mathcal{B}_2$ montre que l'équation (2.3) est satisfaite pour $\mathcal{C}$. D'autre part, $\mathcal{C}$ est stable par multiplication et $\sigma(\mathcal{C}) = \sigma(\mathcal{B}_1, \mathcal{B}_2)$.

Le point 4 s'établit comme le précédent grâce au TCM à partir du cas où $f(x,y) = f_1(x) f_2(y)$, f_1 et f_2 étant des fonctions mesurables bornées.

∎

Enfin, l'espérance conditionnelle a toutes les propriétés d'une espérance. En particulier

Proposition 2.4.

1. *Pour tout $p \in [1, \infty]$,*
$$\|\mathbf{E}(X/\mathcal{B})\|_p \leq \|X\|_p.$$
L'espérance conditionnelle est donc une contraction dans tous les espaces L^p.
2. *L'espérance conditionnelle est dans L^2 un projecteur sur le sous-espace fermé $L^2(\Omega, \mathcal{B}, \mathbf{P})$. En particulier, c'est un opérateur symétrique : pour tout couple (X_1, X_2) de variables de L^2, on a*
$$\mathbf{E}[X_1 E(X_2/\mathcal{B})] = \mathbf{E}[\mathbf{E}(X_1/\mathcal{B})X_2] = \mathbf{E}[\mathbf{E}(X_1/\mathcal{B})\mathbf{E}(X_2/\mathcal{B})].$$
3. *(Inégalité de* Jensen *conditionnelle). Si ϕ est convexe, alors*
$$\phi(\mathbf{E}(X/\mathcal{B})) \leq \mathbf{E}(\phi(X)/\mathcal{B}),$$

dès que les deux membres ont un sens. En particulier

$$|\mathbf{E}(X/\mathcal{B})| \leq \mathbf{E}(|X|/\mathcal{B}).$$

4. *(Inégalité de Markov conditionnelle). Si Y est une variable $\mathcal{B}$-mesurable positive, et si X est une variable positive intégrable, alors*

$$\mathbf{E}(\mathbf{1}_{X\geq Y}/\mathcal{B}) \leq \frac{1}{Y}\mathbf{E}(X/\mathcal{B}).$$

5. *Si X est de carré intégrable, sa variance conditionnelle est la variable aléatoire (presque sûrement positive)*

$$\sigma^2(X/\mathcal{B}) = \mathbf{E}(X^2/\mathcal{B}) - \mathbf{E}(X/\mathcal{B})^2 = \mathbf{E}[(X - \mathbf{E}(X/\mathcal{B}))^2/\mathcal{B}].$$

On a de plus, en posant $\sigma^2(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$,

$$\sigma^2(X) = \sigma^2(\mathbf{E}(X/\mathcal{B})) + \mathbf{E}(\sigma^2(X/\mathcal{B})), \tag{2.4}$$

et même, lorsque $\mathcal{B}_1 \subset \mathcal{B}_2$

$$\sigma^2(X/\mathcal{B}_1) = \sigma^2(\mathbf{E}(X/\mathcal{B}_2)/\mathcal{B}_1) + \mathbf{E}(\sigma^2(X/\mathcal{B}_2)/\mathcal{B}_1). \tag{2.5}$$

6. *(Inégalité de* Chebychef *conditionnelle) Si X est de carré intégrable, et si Y est $\mathcal{B}$-mesurable positive, alors*

$$\mathbf{E}(\mathbf{1}_{|X|\geq Y}/\mathcal{B}) \leq \frac{1}{Y^2}\mathbf{E}(X^2/\mathcal{B}).$$

7. *(Inégalité de* Hölder *conditionnelle). Si $p \in [1, \infty]$ et si $\frac{1}{p} + \frac{1}{q} = 1$, alors*

$$|\mathbf{E}[XY/\mathcal{B}]| \leq [\mathbf{E}(|X|^p/\mathcal{B})]^{1/p}[\mathbf{E}(|Y|^q/\mathcal{B})]^{1/q}.$$

8. *(Théorème de convergence monotone conditionnelle). Si (X_n) est une suite de variables aléatoires positives qui converge en croissant vers X, alors la suite $(\mathbf{E}(X_n/\mathcal{B}))$ converge (en croissant), presque sûrement, vers $\mathbf{E}(X/\mathcal{B})$.*
9. *(Lemme de* Fatou *conditionnel)*
   Si (X_n) est une suite de variables aléatoires positives, alors

$$\mathbf{E}(\liminf_n X_n/\mathcal{B}) \leq \liminf_n[\mathbf{E}(X_n/\mathcal{B})].$$

10. *(Convergence dominée conditionnelle) Si X_n est une suite de variables qui converge presque sûrement vers X, et s'il existe une variable Y positive (non nécessairement intégrable) telle que, pour tout n, $|X_n| \leq Y(ps)$, alors $\mathbf{E}(X_n/\mathcal{B})$ converge presque sûrement vers $\mathbf{E}(X/\mathcal{B})$ sur l'ensemble $\{\mathbf{E}(Y/\mathcal{B}) < \infty\}$.*

Les démonstrations de tous ces points sont élémentaires, en utilisant les propriétés analogues des espérances ordinaires. Nous les laissons au lecteur à titre d'exercice.

Il faut faire attention à ce que, dans le cas positif, l'espérance conditionnelle d'une variable finie presque sûrement peut fort bien être infinie partout. Il suffit pour s'en convaincre de considérer, sur $[0, 1]$, muni de la mesure de Lebesgue, la sous-tribu engendrée par les deux intervalles $[0, 1/2[$ et $[1/2, 1]$, ainsi que la variable $X(\omega) = \frac{1}{\omega} + \frac{1}{1-\omega}$.

3 Noyaux de transition, lois conditionnelles : rappels.

Nous commençons par quelques rappels sur les lois conditionnelles et l'indépendance conditionnelle de tribus. Les lois conditionnelles sont des outils pour calculer les espérances conditionnelles. Elles se définissent à l'aide de la notion de noyau de transition.

Définition 3.1. *(Noyaux de transition.) Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. On appelle* **probabilité de transition**, *ou* **noyau de transition**, *de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ une application Q de $E \times \mathcal{F}$ dans $[0, 1]$ telle que :*

1. *Pour tout $B \in \mathcal{F}$, l'application $x \mapsto Q(x, B)$ est $\mathcal{E}$-mesurable,*
2. *Pour tout $x \in E$, l'application $B \mapsto Q(x, B)$ est une probabilité sur $(F, \mathcal{F})$.*

De plus, si $\mu(dx)$ est une probailité sur $(E, \mathcal{E})$, alors deux noyaux $Q(x, dy)$ et $Q'(x, dy)$ sur $E \times \mathcal{F}$ seront dit égaux μ-presque partout s'il existe un ensemble $A \in \mathcal{E}$ tel que $\mu(A) = 0$ et tel que, pour tout $x \notin A$, les probalités $Q(x, dy)$ et $Q'(x, dy)$ coïncident.

Ce n'est rien d'autre qu'une probabilité sur $\mathcal{F}$ dépendant mesurablement d'un paramètre x dans E. On la notera $Q(x, dy)$.

Ainsi, si E est un ensemble fini ou dénombrable, muni de la tribu $\mathcal{E} = \mathcal{P}(E)$ (ce qui sera le cas dans la plupart des cas qui vont nous intéresser), un noyau de transition de transition de F dans E , n'est rien d'autre qu'une famille de probabilités sur $(F, \mathcal{F})$ indexée par un paramètre $x \in E$.

Exemple. — On obtient souvent des noyaux de la façon suivante : si $f(x, y)$ est une fonction mesurable de deux variables sur $E \times G$, et si Y est une variable aléatoire à valeurs dans G, alors la loi de $f(x, Y)$ est un noyau de transition $Q(x, dy)$ de E dans F.

On peut faire sur les noyaux toutes les opérations habituelles sur les mesures de probabilité. La philosophie sous-jacente est que, si les mesures dépendent mesurablement d'un paramètre, ces opérations habituelles sur les mesures dépendent à nouveau mesurablement de ce paramètre.

La proposition suivante découle immédiatement du théorème des classes monotones, et permet de voir que les propriétés de mesurabilité sont bien conservées.

Proposition 3.2. *Si $Q(x, dy)$ est une probabilité de transition de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ et f une fonction mesurable positive (ou bornée) sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$, la fonction*

$$G(x) = \int_F f(x, y) Q(x, dy)$$

est mesurable sur $(E, \mathcal{E})$. Lorsque f ne dépend pas de x, on la notera $Qf(x)$ ou parfois $Q(f)(x)$:

$$Qf(x) = \int_F f(y) Q(x, dy).$$

Ainsi, l'application $f \mapsto Q(f)$ transforme les fonctions boréliennes bornées sur F en fonctions boréliennes bornées sur E.

Lorsqu'on a un noyau sur $E \times \mathcal{F}$ et une mesure sur E, on peut construire une mesure sur le produit $E \times F$. C'est l'intérêt fondamental des noyaux.

Proposition 3.3. *Si μ est une mesure positive σ-finie sur $(E, \mathcal{E})$, on définit une mesure $m(dx, dy)$ sur $E \times F$ par*

$$\forall B \in \mathcal{E} \otimes \mathcal{F}, \; m(B) = \int_E \int_F Q(x, B_x) \, \mu(dx),$$

où $B_x = \{y | \; (x, y) \in B\}$.

Alors m est une mesure positive σ-finie sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$. C'est une probabilité si μ en est une. On la note $Q(x,dy)\mu(dx)$, ou encore $Q \times \mu$, ou bien $\mu \times Q$.

Si $f(x,y)$ est une fonction mesurable sur le produit, positive ou bornée , on a

$$\int f(x,y)Q(x,dy)\mu(dx) = \int_E [\int_F f(x,y)Q(x,dy)]\mu(dx).$$

L'image de m par la projection $(x,y) \mapsto x$ est μ.

Toutes ces assertions sont élémentaires si on utilise le théorème des classes monotones.

Ceci nous amène à la définition suivante.

Définition 3.4. *(***Conditionnement***) Soit $\rho(dx,dy)$ une mesure de probabilité sur $E \times F, \mathcal{E} \otimes \mathcal{F})$, et $\mu(dx)$ la mesure image de ρ par la projection $\pi : (x,y) \in E \times F \mapsto x \in E$. si $Q(x,dy)$ est un noyau sur $E \times \mathcal{F}$ tel que*

$$\rho(dx,dy) = Q(x,dy)\mu(dx),$$

alors $Q(x,dy)$ s'appelle le conditionnement de ρ par rapport à x. Ce noyau est unique à l'égalité μ-presque partout près.

Remarque. — Rien ne dit que dans des espaces de probabilité très généraux, un tel noyau existe toujours pour toutes les lois de probabilité sur le produit. En fait, il faut quelques hypothèses minimales sur la nature des espaces mesurables pour s'assurer de leur existence en général. Néanmoins, dans toutes les situations que nous rencontrerons, ces noyaux nous seront explicitement donnés.

Notation 3.5. *La deuxième marginale de la mesure $\mu \times Q$ est notée μQ. L'application $\mu \mapsto \mu Q$ transporte donc les probabilités définies sur $(E, \mathcal{E})$ en probabilités définies sur $(F, \mathcal{F})$.*

L'application $\mu \mapsto \mu Q$ est l'application duale de l'application $f \mapsto Qf$: si f est une fonction $\mathcal{F}$-mesurable, positive ou bornée, alors

$$\int f(y)\mu Q(dy) = \int_F f(y)Q(x,dy) = \int_E Q(f)(x)\mu(dx).$$

En d'autres termes, si on note $\langle \mu, f \rangle = \int f(x)\mu(dx)$, alors

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

Exemple. — On comprendra sans doute un peu mieux cette dualité si l'on regarde sur l'exemple le plus simple, où E et F sont des ensembles finis. $E = \{1, \ldots, n\}$ et $F = \{1, \ldots, p\}$. Un mesure de probabilité ν sur F se représente par un vecteur ligne $[\nu] = (\nu_1, \ldots, \nu_p)$, où $\nu_i = \nu(i)$.

Un noyau Q est la donnée de n telles mesures, que nous représentons par une matrice

$$[Q] = \begin{pmatrix} \nu_1^1 & \ldots & \nu_p^1 \\ \nu_1^2 & \ldots & \nu_p^2 \\ & \ldots & \\ \nu_1^n & \ldots & \nu_p^n \end{pmatrix}.$$

Alors, en représentation matricielle, $[\nu Q] = [\nu][Q]$.

De même, si nous convenons de représenter une fonction définie sur E par un vecteur colonne

$$[f] = \begin{pmatrix} f_1 \\ \cdot \\ \cdot \\ f_n \end{pmatrix},$$

alors, $\langle \nu, f\rangle$ est le réel $[\nu][f]$, et $[Qf] = [Q][f]$. On voit donc que la matrice $[Q]$ agit sur les mesures comme la transposée de $[Q]$ sur les fonctions.

Opérations sur les noyaux.

On peut faire sur les noyaux toutes les opérations habituelles que l'on fait sur les probabilités, et même un peu plus. En voici quelques exemples :

1. **Produit de noyaux** : Si Q est une probabilité de transition de $(E, \mathcal{E})$ dans $(F, \mathcal{F})$ et si R est une probabilité de transition de $(F, \mathcal{F})$ dans $(G, \mathcal{G})$, on note pour tout $C \in \mathcal{F} \times G$:

$$(Q \times R)(x, C) = \int_F [\int_G R(y, dz)\mathbf{1}_C(y, z)]Q(x, dy).$$

   Alors $Q \times R$ est une probabilité de transition de $(E, \mathcal{E})$ dans $(F \times G, \mathcal{F} \otimes \mathcal{G})$. On le notera aussi $Q(x, dy)R(y, dz)$.

   En d'autres termes, si dans la construction de $\mu \times R$, nous remplaçons la mesure μ par un noyau Q, nous obtenons un noyau sur le produit au lieu d'une mesure.

2. **Composition des noyaux** Si Q et R sont deux noyaux comme plus haut, la deuxième marginale de la mesure $Q(x, dy)R(y, dz)$ sera notés

$(QR)(x,dz)$. En d'autres termes, pour une fonction $f : G \mapsto \mathbb{R}$ positive ou bornée,

$$\int_G f(z)(QR)(x,dz) = \int_F [\int_G f(z)R(z,dy)]Q(x,dy).$$

Si μ est une sur E, on a $\mu(QR) = (\mu Q)R$. Si f est une fonction positive ou bornée de G dans $\mathbb{R}$, alors $(QR)(f) = Q(R(f))$. On note

$$QR(x,dz) = \int_y Q(x,dy)R(y,dz).$$

Il s'agit de l'application $\mu \mapsto \mu R$ étendue aux noyaux.

3. **Produit tensoriel de noyaux** Si Q est un noyau de $(E,\mathcal{E})$ dans $(F,\mathcal{F})$ et Q_1 un autre noyau de $(E,\mathcal{E})$ dans $(F_1,\mathcal{F}_1)$, alors il existe un (unique) noyau $Q \otimes Q_1$ sur $(F \times F_1, \mathcal{F} \otimes \mathcal{F}_1)$ tel que, pour tout $B \in \mathcal{F}$ et tout $B_1 \in \mathcal{F}_1$, on ait

$$Q \otimes Q_1(x, B \times B_1) = Q(x,B)Q_1(x,B_1).$$

On note aussi $Q \otimes Q_1(x,(dy,dy_1)) = Q(x,dy)Q_1(x,dy_1)$.
Cette opération est l'extension aux noyaux du produit tensoriel des mesures.

4. On pourrait de même définir des noyaux de $E \times E_1$ dans $F \times F_1$ par

$$Q(x,dy)Q_1(x_1,dy_1),$$

qui dépend cette fois d'un paramètre $(x,x_1) \in E \times E_1$.

Définition 3.6. (**Lois conditionnelles.**) *Soit un couple de variables aléatoires (X,Y) est à valeurs dans $E \times F$, et soit μ la loi de X. On appelle loi conditionnelle de X sachant Y le noyau obtenu en conditionnant la loi $\rho(dx,dy)$ du couple (X,Y) par rapport à y, et on la note $\mathcal{L}(X/Y)$. La valeur de ce noyau (définie à un ensemble de mesure nulle près pour la loi de Y) est notée $Q(y,dx) = \mathcal{L}(X/Y = y)$.*

On a peut bien sûr définir de manière symétrique la loi $\mathcal{L}(Y/X)$.

Remarques

1. Si Q est la loi conditionnelle de Y sachant X et si μ est la loi de X, alors la loi de Y est μQ.
Cela se voit immédiatement sur la définition.

2. Bien que la loi conditionnelle $\mathcal{L}(Y/X = x)$ ne soit définie qu'à un ensemble de mesure nulle près pour la loi de X, il y a de nombreux cas où on peut lui donner un sens pour toutes les valeurs de x.

   Tout d'abord, c'est le cas lorsque X ne prend qu'un nombre fini ou dénombrable de valeurs (ce qui sera le cas dans la majeure partie de la suite).

   Cest encore le cas lorsqu'on dispose sur E et F de topologies telles qu'on puisse choisir une version de la loi conditionnelle telle que $x \mapsto Q(x, dy)$ soit continue pour la convergence étroite des lois sur F. C'est en particulier le cas lorsque $Y = F(X, Z)$, où la fonction F est continue en x, et que Z est indépendante de X.

L'intérêt des lois conditionnelles est qu'elles servent à calculer des espérances conditionnelles.

Proposition 3.7. *Si (X, Y) est un couple de variables aléatoires à valeurs dans $E \times F$, et si f est une fonction bornée, et si $Q(x, dy)$ désigne la loi conditionnelle de Y sachant X, alors*

$$\mathbf{E}(f(Y)/\sigma(X)) = Q(f)(X).$$

En d'autres termes, l'espérance pour la loi conditionnelle est l'espérance conditionnelle.

Remarque. — Considérons $Q(X(\omega), dy)$: c'est un noyau sur $(\Omega, \sigma(X)) \times \mathcal{F})$. Alors

$$\int f(y) Q(X(\omega), dy) = \mathbf{E}(f(Y)/\sigma(X)).$$

On aurait donc pu aussi définir la loi conditionnelle de Y par rapport à une sous-tribu $\mathcal{B}$ de $\mathcal{A}$, comme un noyau $Q(\omega, dy)$ sur $(\Omega, \mathcal{B}) \times \mathcal{F}$ qui soit tel que

$$\mathbf{E}(f(Y)/\mathcal{B}) = \int f(y) Q(\omega, dy),$$

pour toutes les fonctions $f : F \mapsto \mathbb{R}$ qui soient mesurables et bornées. Comme nous ne nous servirons que de tribus $\sigma(X)$, ce serait une généralisation inutile.

Calculs de lois conditionnelles.

Les points suivants donnent des moyens explicites de calculer des lois conditionnelles, qui sont suffisants dans la plupart des situations qu'on rencontrera.

1. X est indépendante de Y si et seulement si $\mathcal{L}(Y/X = x)$ ne dépend pas de x. Dans ce cas, c'est la loi de Y.

   (Ceci signifie plus exactement qu'en dehors d'un ensemble négligeable pour la loi de X, le noyau $Q(x, dy)$ est égal à une mesure de probabilité fixe $\nu(dy)$, qui est alors nécessairement la loi de Y.)
2. Si X et Y sont indépendantes et si $Z = F(X, Y)$, la loi de Z sachant $X = x$ est la loi de $F(x, Y)$. C'est faux si Y n'est pas indépendante de X. Dans ce cas, il faut calculer la loi de $F(x, Y^x)$, où Y^x est une variable dont la loi est $\mathcal{L}(Y/X = x)$.
3. Si $Y = f(X)$, la loi de Y sachant $X = x$ est la masse de Dirac $\delta_{f(x)}$.
4. Soit f une application mesurable de $(F, \mathcal{F})$ dans $(G, \mathcal{G})$. Convenons de noter $f^*\mu$ l'image d'une probabilité μ sur F par l'application f : $f^*\mu(B) = \mu(f^{-1}(B))$. Alors, la loi conditionnelle $\mathcal{L}(f(Y)/X = x)$ est $f^*(\mathcal{L}(Y/X = x))$. En d'autres termes, la loi conditionnelle de l'image par f est l'image par f de la loi conditionnelle.
5. Si (X, Y) est un couple de variables aléatoires à valeurs dans $E \times F$, si B est un ensemble $\mathcal{F}$ mesurable, si $\mathcal{L}(Y/X) = Q$, on notera $\mathbf{P}(Y \in B/X)$ la variable aléatoire $Q(\mathbf{1}_B)(X)$.
6. (**Principe des conditionnements successifs**). Soient (X, Y, Z) trois variables à valeurs dans E, F, G respectivement. Soit $R(z; dx, dy)$ la loi conditionnelle $\mathcal{L}((X, Y)/Z)$, et soit $Q(z, y; dx$ de conditionnement de $R(z; dx, dy)$ par rapport à y. Alors $Q(z, y; dx) = \mathcal{L}(X/(Z, Y)$.

   En d'autres termes, si on veut conditionner une variable par rapport à un couple d'autres variables, on peut déjà la conditionner par rapport à l'une d'entre elle, et reconditionner le résultat par rapport à la seconde.

4 Indépendance conditionnelle.

La définition qui suit est fondamentale dans l'étude des chaînes de Markov
.

Définition 4.1. *(***Indépendance conditionnelle.***) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, et soit $(\mathcal{A}_1, \mathcal{A}_2)$ deux sous-tribus de $\mathcal{A}$, et soit $\mathcal{A}_3$ une sous tribu de $\mathcal{A}_1 \cap \mathcal{A}_2$. On dit que $\mathcal{A}_1$ est conditionnellement indépendante de $\mathcal{A}_2$ sachant $\mathcal{A}_3$, et on note $\mathcal{A}_1 \perp_{\mathcal{A}_3} \mathcal{A}_2$ si, pour toute fonction $\mathcal{A}_1$ mesurable bornée Z_1, on a*

$$\mathbf{E}(Z_1/\mathcal{A}_2) = \mathbf{E}(Z_1/\mathcal{A}_3).$$

On a les propriétés suivantes :

1. Cette relation est symétrique en $(\mathcal{A}_1, \mathcal{A}_2)$.
2. $\mathcal{A}_1 \perp_{\mathcal{A}_3} \mathcal{A}_2$ si et seulement si, pour toute variable Z_1 $\mathcal{A}_1$-mesurable bornée, pour toute variable Z_2 $\mathcal{A}_2$ -mesurable bornée,
$$\mathbf{E}(Z_1 Z_2/\mathcal{A}_3) = \mathbf{E}(Z_1/\mathcal{A}_3)\mathbf{E}(Z_2/\mathcal{A}_3).$$
3. Si $(\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3)$ sont trois sous-tribus, que $\mathcal{A}_3 \subset \mathcal{A}_2$, et que, $\forall X$ $\mathcal{A}_1$-mesurable bornée, on ait $\mathbf{E}(X/\mathcal{A}_2) = \mathbf{E}(X/\mathcal{A}_3)$, alors $\sigma(\mathcal{A}_1, \mathcal{A}_3) \perp_{\mathcal{A}_3} \mathcal{A}_2$. On n'a donc pas besoin a priori que la tribu $\mathcal{A}_3$ soit une sous-tribu de $\mathcal{A}_1$ pour avoir l'équivalence précédente. Mais, il est nécessaire que $\mathcal{A}_3$ soit une sous-tribu de $\mathcal{A}_2$.

   En effet, si ε_1 et ε_2 sont deux variables de Bernoulli centrées et indépendantes, que $\varepsilon_3 = \varepsilon_1\varepsilon_2$, et si $\mathcal{A}_i = \sigma(\varepsilon_i)$, alors
$$\mathbf{E}(f(\varepsilon_1)/\mathcal{A}_2) = \mathbf{E}(f(\varepsilon_1)/\mathcal{A}_3) = \mathbf{E}(f(\varepsilon_1)),$$
   tandis que $\mathbf{E}(\varepsilon_1\varepsilon_2/\mathcal{A}_3) = \varepsilon_3$.
4. Si (X, Y, Z) sont trois variables aléatoires telles que $\mathcal{A}_1 = \sigma(X, Z)$, $\mathcal{A}_2 = \sigma(Y, Z)$ et $\mathcal{A}_3 = \sigma(Z)$, alors $\mathcal{A}_1 \perp_{\mathcal{A}_3} \mathcal{A}_2$ si et seulement si la loi conditionnelle de X sachant $(Y, Z) = (y, z)$ ne dépend que de z. Dans ce cas, on note $X \perp_Z Y$.
5. $X \perp_Z Y$ si et seulement si la loi conditionnelle de (X, Y) sachant $Z = z$ est le produit de la loi de X sachant $Z = z$ et de la loi de Y sachant $Z = z$:
$$\mathcal{L}((X, Y)/Z = z) = \mathcal{L}(X/Z = z) \otimes \mathcal{L}(Y/Z = z).$$
6. Si (X_1, X_2, X_3) sont trois variables aléatoires indépendantes, si $X = F(X_1, X_2)$ et $Y = G(X_2, X_3)$, alors $X \perp_{X_2} Y$.
7. Si la loi $\mu(dx, dy, dz)$ du triplet (X, Y, Z) admet une densité par rapport à une mesure produit $\mu_1(dx)\mu_2(dy)\mu_3(dz)$, alors $X \perp_Z Y$ si et seulement si
$$\mu(dx, dy, dz) = f(x, z)g(z, y)\mu_1(dx)\mu_2(dy)\mu_3(dz).$$
8. Si les variables (X, Y, Z) ne prennent qu'un nombre fini ou dénombrable de valeurs, alors $X \perp_Z Y$ si et seulement si $\forall(x, y, z)$,
$$\mathbf{P}(X = x, Y = y, Z = z)\mathbf{P}(Z = z) = \mathbf{P}(X = x, Z = z)\mathbf{P}(Y = y, Z = z).$$

5 Intégrabilité uniforme.

Définition 5.1. *On dit qu'une famille de v.a.* $(X_\alpha, \alpha \in I)$ *est* **uniformément intégrable** *si*

$$\lim_{n\to\infty} \sup_{\alpha\in I} \int_{\{|X_\alpha|\geq n\}} |X_\alpha| d\mathbf{P} = 0. \tag{5.6}$$

Il est équivalent d'avoir :

$$\sup_\alpha E[|X_\alpha|] < +\infty \; ; \; \forall \varepsilon > 0, \exists \delta > 0, \sup_{\alpha\in I, A\in\mathcal{A}, \mathbf{P}(A)\leq\delta} \{\int_A |X_\alpha| d\mathbf{P}\} < \varepsilon. \tag{5.7}$$

Une partie finie de L^1 est uniformément intégrable, ainsi qu'une famille dominée dans L^1. Plus généralement, si $\mathcal{H}$ est uniformément intégrable, et si une famille $\mathcal{H}_1$ est telle que

$$\forall X \in \mathcal{H}_1, \; \exists Y \in \mathcal{H}, \; |X| \leq |Y|,$$

alors $\mathcal{H}_1$ est uniformément intégrable.

Le critère suivant est en pratique très utile pour s'assurer de l'intégrabilité uniforme d'une famille.

Proposition 5.2. *(Théorème de* De La Vallée Poussin*) Une famille* $X_\alpha, \alpha \in I$ *est uniformément intégrable si et seulement si il existe une fonction* $\Phi(x)$: $\mathbb{R}_+ \to \mathbb{R}_+$*, croissante, telle que* $\Phi(x)/x$ *converge vers* $+\infty$ *lorsque* $x \to +\infty$*, et telle que*

$$\sup_{\alpha\in I} \mathbf{E}(\Phi(|X_\alpha|)) < \infty.$$

Ainsi, une partie bornée dans L^p est uniformément intégrable dès que $p > 1$.

Une autre classe de familles uniformément intégrables est donnée de la façon suivante.

Proposition 5.3. *Si* X *est dans* L^1*, l'ensemble des variables* $\mathbf{E}(X/\mathcal{B})$*, où* $\mathcal{B}$ *parcourt les sous-tribus de* $\mathcal{A}$*, est uniformément intégrable. Il en va de même de l'ensemble de toutes les espérances conditionnelles par rapport à toutes les sous-tribus des éléments d'une famille uniformément intégrable.*

La notion d'intégrabilité uniforme est principalement utilisée à cause de la propriété suivante :

Proposition 5.4. *Si (X_n) est une suite de variables aléatoires intégrables, qui converge vers X en probabilité lorsque n converge vers ∞, alors il est équivalent de dire*

1. *(X_n) est uniformément intégrable.*
2. *X_n converge vers X dans L^1.*

Dans ce cas, on a bien sûr $\mathbf{E}(X_n) \to \mathbf{E}(X)$.

Remarquons qu'on retrouve ainsi comme cas particulier le théorème de convergence dominée.

Chapitre 2

Filtrations, temps d'arrêt.

Dans ce chapitre, on introduit le vocabulaire élémentaire de la théorie des processus à temps discrets que nous utiliserons dans la suite du cours.

Un **processus** (à temps discret) sur $(\Omega, \mathcal{A}, \mathbf{P})$ est une suite de variables aléatoires $(X_n(\omega))_{n\geq 0}$, à valeurs dans un espace mesuré $(E, \mathcal{E})$. Lorsque E est l'ensemble des réels, on parlera de processus réel, ou scalaire. Il est souvent utile de le considérer comme une variable aléatoire $X(\omega, n)$ défini sur l'espace produit $\Omega \times \mathbb{N}$, muni de la tribu produit.

Une **filtration** $(\mathcal{F}_n)$ est la donnée d'une suite croissante de sous-tribus de $\mathcal{A}$. Sauf si c'est spécifié par ailleurs, on pose en général $\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n = \sigma(\bigcup_n \mathcal{F}_n)$. C'est la plus petite tribu qui contienne toutes les $\mathcal{F}_n$, où $\bigcup_n \mathcal{F}_n$. Remarquons que, puisque la famille de tribus $\mathcal{F}_n$ est croissante, alors $\bigcup_n \mathcal{F}_n$ est une algèbre, mais ce n'est pas en général une σ-algèbre.

Un exemple typique de filtration est obtenu de la façon suivante : on observe une suite (X_n) de variables aléatoires, X_n étant l'observation à l'instant n. On note alors

$$\mathcal{F}_n^X = \sigma(X_0, \cdots, X_n). \tag{0.1}$$

C'est la plus petite tribu qui rende mesurable les variables $(X_0, \cdots, X_n)$. On l'appelle la **filtration naturelle** du processus X_n. On la notera $\mathcal{F}^X$.

Il est souvent commode (mais pas indispensable) de supposer que $\mathcal{F}_0$ contient tous les négligeables de $\mathcal{F}_\infty$, c'est à dire tous les ensembles A qui sont inclus dans un ensemble de $\mathcal{F}_\infty$ qui est de probabilité nulle. On peut toujours augmenter toutes les tribus de ces négligeables (on dit qu'on les complète). Cela permet de supposer qu'un ensemble de mesure nulle est toujours dans $\mathcal{F}_0$, et d'avoir la même notion d'égalité ps dans toutes les tribus.

On dit qu'un processus (X_n) est **adapté** à $(\mathcal{F}_n)$ si, pour tout n, X_n est $\mathcal{F}_n$-mesurable.

Ainsi, la filtration naturelle du processus (X_n) est la plus petite filtration par rapport à laquelle le processus (X_n) soit adapté.

On dit qu'un processus (X_n) est **prévisible** si X_0 est $\mathcal{F}_0$-mesurable, et que, pour tout $n \geq 1$, X_n est $\mathcal{F}_{n-1}$-mesurable. Ainsi, $(Y_n) = (X_{n-1})$ est prévisible dans la filtration naturelle de (X_n).

De façon générale, on pourra dire qu'une partie $A \subset \Omega \times \mathbb{N}$ est adaptée si la suite $Y_n(\omega) = \mathbf{1}_A(\omega, n)$ est adaptée. L'ensemble des parties adaptées forme une tribu, et un processus $X(\omega, n)$ est adapté s'il est mesurable par rapport à cette tribu.

On peut faire de même avec les ensembles prévisibles.

La définition qui suit est sans doute la plus importante :

Définition 0.5. *Un temps d'arrêt est une application de Ω dans $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ qui a la propriété suivante :*

$$\forall n \in \mathbb{N},\ \{T \leq n\} \in \mathcal{F}_n. \tag{0.2}$$

Remarques

1. Il est important d'autoriser aux *temps d'arrêt* de prendre la valeur $+\infty$. Noter que l'événement $\{T = \infty\}$ est automatiquement $\mathcal{F}_\infty$ mesurable.
2. La variable aléatoire $T : \Omega \longrightarrow \bar{\mathbb{N}}$ est un *temps d'arrêt* si et seulement si, pour tout n, l'événement $\{T = n\}$ est $\mathcal{F}_n$ mesurable. La démonstration est laissée à titre d'exercice.
3. Si T est une variable à valeurs dans $\bar{\mathbb{N}}$, on peut considérer l'ensemble

$$[T, \infty[= \{(\omega, n) \ \mid\ T(\omega) \leq n\}$$

   ou l'ensemble $[0, T] = \{(\omega, n) \mid n \leq T(\omega)\}$. Alors, T est un *temps d'arrêt* si et seulement si $[T, \infty[$ est adapté, ou encore si et seulement si $[0, T]$ est prévisible. Les ensembles $[0, T]$, $[0, T[$, $[T, \infty[$, etc. sont des intervalles de temps aléatoires. On les appelle des **intervalles stochastiques**.

L'exemple fondamental de *temps d'arrêt* est le suivant : soit (X_n) un processus adapté et A_n une famille d'ensembles mesurables dans l'espace où (X_n)

prend ses valeurs. Alors, la variable

$$T(\omega) = \inf\{n |\ X_n(\omega) \in A_n\} \tag{0.3}$$

est un *temps d'arrêt*. (Par convention, on pose $\inf(\emptyset) = +\infty$.)

Pour le voir, il suffit d'écrire

$$\{T = n\} = \cap_{k=0}^{n-1}\{X_k \notin A_k\} \cap \{X_n \in A_n\},$$

dont on voit immédiatement qu'il est dans $\mathcal{F}_n$.

Réciproquement, tout temps d'arrêt peut s'écrire comme cela, en considérant le processus $\mathbf{1}_{[T,\infty[}$ et $A_n = \{1\}$.

De façon plus générale, si A est une partie de $\Omega \times \mathbb{N}$, on note

$$D_A(\omega) = \inf\{n \ \mid \ (\omega, n) \in A\}. \tag{0.4}$$

On appelle cette variable aléatoire le début de A. Alors, si A est adaptée, son début D_A est *temps d'arrêt*. L'exemple fondamental (0.3) se ramène à cela, en prenant $A = \{(\omega, n) \ \mid \ X_n(\omega) \in A_n\}$.

Les *temps d'arrêt* vérifient les propriétés élémentaires suivantes :

si S et T sont des *temps d'arrêt*, $\sup(S,T)$, $\inf(S,T)$ sont des *temps d'arrêt*, $S+T$ est un *temps d'arrêt*, mais en général $T-1$ n'est pas un *temps d'arrêt*. Si S_n est une suite de *temps d'arrêt*, alors $\limsup_n S_n$ et $\liminf_n S_n$ sont des *temps d'arrêt*.

Notation 0.6. *Dans toute la suite, nous désignerons le* sup *par le symbole* $\vee$ *et l'*inf *par le symbole* $\wedge$.

Ainsi

$$\sup(S,T) = S \vee T, \ \inf(S,T) = S \wedge T.$$

Définition 0.7. Tribus associées aux temps d'arrêt *: si T est un temps d'arrêt, on note*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty, \ \forall n \in \mathbb{N}, \ A \cap \{T \le n\} \in \mathcal{F}_n.\} \tag{0.5}$$

C'est une tribu, qu'on appelle la tribu des événements antérieurs à T.

Pour vérifier qu'une variable est $\mathcal{F}_T$-mesurable, on a la proposition suivante :

Proposition 0.8. *Une variable réelle X est $\mathcal{F}_T$ mesurable si et seulement si, pour tout n, $X\mathbf{1}_{T\leq n}$ est $\mathcal{F}_n$ mesurable.*

Démonstration. — Pour qu'une variable réelle X soit $\mathcal{B}$ mesurable, quelle que soit la tribu $\mathcal{B}$, il faut et il suffit que pour tout borélien B de $\mathbb{R}$ qui ne contienne pas 0, l'événement $\{X \in B\}$ soit $\mathcal{B}$-mesurable. Soit B un borélien de $\mathbb{R}$, qui ne contient par 0. Alors,

$$\{X\mathbf{1}_{T\leq n} \in B\} = \{X \in B\} \cap \{T \leq n\}.$$

Au vu de la définition 0.7, cette identité donne immédiatement la condition nécessaire et suffisante.

∎

Remarque. — Il ne faut pas confondre la tribu $\mathcal{F}_T$ et la tribu $\sigma(T)$, la plus petite tribu rendant T mesurable : pour tout k, on a $\{T = k\} \in \mathcal{F}_T$, et donc $\sigma(T) \subset \mathcal{F}_T$, mais l'inverse est faux en général.

Proposition 0.9.

1. *Si S et T sont des temps d'arrêt, et si $A \in \mathcal{F}_S$, alors $A \cap \{S \leq T\} \in \mathcal{F}_T$.*
2. *Les événements $\{S \leq T\}$, $\{S < T\}$, $\{S = T\}$ appartiennent à $\mathcal{F}_S \cap \mathcal{F}_T$.*
3. *Si $S \leq T$, alors $\mathcal{F}_S \subset \mathcal{F}_T$.*
4. *Si S et T sont des temps d'arrêt, alors $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S\wedge T}$.*
5. *Si T est un temps d'arrêt, alors*

   $$\mathcal{F}_T = \sigma(A \cap \{n \leq T\},\ A \in \mathcal{F}_n). \tag{0.6}$$

6. *Si T est un temps d'arrêt, alors*

   $$\mathcal{F}_T = \vee_n \mathcal{F}_{T\wedge n}. \tag{0.7}$$

7. *Si T et S sont des temps d'arrêt, alors $\mathcal{F}_S \vee \mathcal{F}_T = \mathcal{F}_{S\vee T}$, où $\mathcal{F}_S \vee \mathcal{F}_T$ désigne la plus petite tribu qui contienne à la fois F_T et $\mathcal{F}_S$.*
8. *Si (X_n) est adapté et que T est un temps d'arrêt, alors la variable X_T définie par*

   $$X_T(\omega) = X_{T(\omega)}(\omega)\mathbf{1}_{T<\infty} = \sum_n X_n \mathbf{1}_{T=n}$$

   est $\mathcal{F}_T$ mesurable. Remarquons que nous avons pris la précaution de poser $X_T = 0$ sur $\{T = \infty\}$. Si une variable X_∞ $\mathcal{F}_\infty$-mesurable est définie, alors on a le même résultat en posant $X_T = X_\infty$ sur $\{T = \infty\}$.

9. *Si T est un temps d'arrêt, et si A est un ensemble mesurable, la variable $T^A = T\mathbf{1}_A + \infty\mathbf{1}_{A^c}$ est un temps d'arrêt si et seulement si $A \in \mathcal{F}_T$, puisque $\{T^A \leq n\} = A \cap \{T \leq n\}$.*

Démonstration. — La plupart des propriétés énoncées sont des conséquences faciles des définitions, et les démonstrations sont laissées au lecteur à titre d'exercice. Nous ne détaillons que les points qui ne sont pas immédiats.

Le point 5 : on remarque tout d'abord que les générateurs proposés sont bien dans $\mathcal{F}_T$. D'autre part, si A est dans $\mathcal{F}_T$, on écrit

$$A = A \cap \{T = \infty\} \bigcup_n A \cap \{T = n\}.$$

Puis $A \cap \{T = n\}[(A \cap \{T \leq n\}) \cap \{n \leq T\}]$, et, puisque $(A \cap \{T \leq n\}) \in \mathcal{F}_n$, on a donc bien écrit $A \cap \{T < \infty\}$ à l'aide des générateurs proposés.

Reste à traiter le cas de $A \cap \{T = \infty\}$. Appelons $\mathcal{F}_T^1$ la tribu engendrée par les ensembles de la forme $B \cap \{n \leq T\}$, avec $B \in \mathcal{F}_n$. Considérons l'ensemble $\mathcal{H}$ des variables aléatoires Z $\mathcal{F}_\infty$-mesurables bornées telles que $Z\mathbf{1}_{T=\infty}$ soit $\mathcal{F}_T^1$-mesurables. Alors $\mathcal{H}$ est un espace vectoriel de classes monotones. Pour montrer que $\mathcal{H}$ contient toutes les variables $\mathcal{F}_\infty$ mesurables bornées, il suffit de voir que $\mathcal{H}$ contient les variables $\mathbf{1}_A$, où $A \in \cup_n \mathcal{F}_n$.

Soit donc un tel A et n tel que $A \in \mathcal{F}_n$. On a

$$A \cap \{T = \infty\} = \cap_{k \geq n} A \cup \{k \leq T\},$$

et on voit sur cette écriture que $A \in \mathcal{F}_T^1$.

Le point 6 : on voit immédiatement que $\vee_n \mathcal{F}_{T\wedge n} \subset \mathcal{F}_T$, puisque $\mathcal{F}_{T\wedge n} \subset \mathcal{F}_T$. Par ailleurs, l'identité (0.6) nous montre que $\mathcal{F}_T$ est engendrée par des éléments qui sont tous dans $\cup_n \mathcal{F}_{T\wedge n}$.

Le point 7 : on a $\mathcal{F}_T \vee \mathcal{F}_S \subset \mathcal{F}_{T\vee S}$. Pour voir l'inverse, on utilise les générateurs de la formule (0.6). Si $A \in \mathcal{F}_n$,

$$A \cap \{n \leq T \vee S\} = (A \cap \{n \leq T\}) \cup (A \cap \{n \leq S\}) \in \mathcal{F}_T \vee \mathcal{F}_S.$$

Le point 8 : cette propriété est immédiate si on utilise la proposition 0.8 qui caractérise les variables $\mathcal{F}_T$-mesurables.

∎

Chapitre 3

Martingales à temps discret.

Comme nous l'avons dit dans l'introduction, la théorie des martingales est l'un des outils les plus puissants de la théorie des probabilités. Nous présentons ici les résultats élémentaires de cette théorie. On pourra consulter les ouvrages de J.Neveu [5] ou de P.A. Meyer et C. Dellacherie [3] pour un exposé plus approfondi. Une exposition plus succinte de la théorie se trouve également dans les livres de P. Barbe et M.Ledoux [1], ou de P.S. Toulouse [6].

La théorie des martingales a son origine dans l'étude des jeux : elle modélise d'une part le caractère aléatoire d'un phénomène mais aussi son évolution dans le temps. On étudie ici le temps discret, c'est à dire lorsque le paramètre de temps est un entier. Il existe aussi une théorie des matringales en temps continu (lorsque le paramètre de temps est un réel positif), mais elle présente des difficultés techniques considérables dans sa forme la plus générale.

1 Définitions et premiers exemples

Définition 1.1. *Soit un processus adapté X sur l'espace de probabilité filtré $(\Omega, \mathcal{A}, (\mathcal{F}_n, n \in \mathbb{N}), \mathbf{P})$ tel que pour tout entier n, X_n est intégrable. On dit que X est une* **martingale** *si pour tout entier n,*

$$\mathbf{E}[X_{n+1}/\mathcal{F}_n] = X_n, \quad \textit{presque sûrement.}$$

On dit que X est une **sous-martingale** *si pour tout entier n,*

$$\mathbf{E}[X_{n+1}/\mathcal{F}_n] \geq X_n, \quad \textit{presque sûrement.}$$

On dit que X est une **sur-martingale** *si pour tout entier n,*

$$\mathbf{E}[X_{n+1}/\mathcal{F}_n] \leq X_n, \quad \textit{presque sûrement.}$$

Exemples

(Les démonstrations des propriétés qui suivent sont laissées au lecteur à titre d'exercice.)

1. Si $H \in L^1(\Omega, \mathcal{A})$, $X_n = \mathbf{E}[H/\mathcal{F}_n]$ est une martingale. D'après le critère (5.3), cette martingale est uniformément intégrable.
2. (**Exemple fondamental.**) Soit $(Z_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires indépendantes et intégrables. X_0 est une variable aléatoire donnée et indépendante de la suite (Z_n). (La plupart du temps, X_0 est constante.) On pose $X_n = X_0 + \sum_{i=1}^n Z_i$. Alors les filtrations $\mathcal{F}_n^X$ et $\mathcal{F}_n = \sigma(X_0, Z_1, \ldots, Z_n)$ sont égales et pour cette filtration :
   (a) si pour tout entier n, $\mathbf{E}(Z_n) = 0$, X est une martingale ;
   (b) si pour tout entier n, $\mathbf{E}(Z_n) \geq 0$, X est une sous- martingale ;
   (c) si pour tout entier n, $\mathbf{E}(Z_n) \leq 0$, X est une sur-martingale ;
   (d) dans tous les cas, si toutes les variables Z_i ont même espérance m, $X_n - nm$ est une martingale.
3. Un cas particulier de l'exemple 2 issu de la théorie des jeux est celui où la loi des variables aléatoires Z_n est la loi de Bernoulli de paramètre p : $\mathbf{P}(Z_i = 1) = p$, $\mathbf{P}(Z_i = -1) = 1 - p$. avec valeurs $+1$ et -1. Dans ce cas, X_n représente la fortune d'un joueur après n paris, lorsque sa fortune initiale est X_0. Alors $M_n = X_n - n(2p-1)$ est une martingale pour la filtration naturelle $\mathcal{F}^X$.
4. Dans l'exemple 2, si l'on suppose que $\mathbf{E}[\exp(aZ_n)] := \exp(r_n)$ existe et est finie, on pose alors $R_n = r_1 + \cdots + r_n$. (Ici, par convention, $R_0 = 0$.) Alors, $M_n = \exp(aX_n - R_n)$ est une martingale pour la filtration naturelle $\mathcal{F}^X$.

Remarquons enfin qu'on peut avoir le choix de la filtration par rapport à laquelle X est une martingale, (resp. une sous-martingale, une sur-martingale). En effet :

Proposition 1.2. *Si X est une martingale (resp. une sur-martingale, une sous-martingale) par rapport à une filtration $(\mathcal{F}_n)$ et qu'elle est adaptée à une filtration $(\mathcal{G}_n)$ plus petite que $(\mathcal{F}_n)$ (au sens où, pour tout n, $\mathcal{G}_n \subset \mathcal{F}_n$), alors X*

est une martingale (resp. une sur-martingale, une sous-martingale) par rapport à la filtration $\mathcal{G}_n$. En particulier, X est une martingale (resp. une sur-martingale, une sous-martingale) par rapport à sa filtration naturelle.

Démonstration. — Il suffit d'appliquer le principe des conditionnements successifs (5). ∎

De même, nous pouvons augmenter les filtrations en rajoutant à chaque tribu $\mathcal{F}_n$ une tribu indépendante :

Proposition 1.3. *Soit (X_n) une martingale (resp. une sous-martingale, une sur-martingale), par rapport à une filtration $\mathcal{F}_n$. Soit $\mathcal{B}$ une tribu indépendante de $\mathcal{F}_\infty$, et soit $\mathcal{G}_n = \mathcal{F}_n \vee \mathcal{B}$. Alors (X_n) est une martingale (resp. une sous-martingale, une sur-martingale) par rapport à la filtration $\mathcal{G}_n$.*

Démonstration. — Il suffit d'appliquer le point (3) de la section 2. ∎

Notation 1.4. *Dans tout ce qui suit, nous noterons*

$$(\Delta X)_n := X_n - X_{n-1} \tag{1.1}$$

le processus des accroissements de la suite (X_n).

Proposition 1.5. *Soit X une $\mathcal{F}$-martingale. Alors*

1. *$\forall n \geq 0$, $\forall k \geq 0$, $\mathbf{E}[X_{n+k}/\mathcal{F}_n] = X_n$; $\mathbf{E}[X_n] = \mathbf{E}[X_0]$.*
2. *Si la martingale est de carré intégrable, les accroissements $(\Delta X)_n$ de X sont orthogonaux :*
$$n \neq m \implies \mathbf{E}[(\Delta X)_n (\Delta X)_m] = 0.$$
3. *Si X est une sur-martingale, $-X$ est une sous-martingale.*
4. *L'espace des martingales est un espace vectoriel réel.*
5. *Si X est une martingale et ϕ une application convexe telle que $Y_n = \phi(X_n)$ est intégrable, alors Y_n est une sous-martingale.*
6. *De même, si X est une sous-martingale, et si ϕ est croissante et convexe, $\phi(X)$ est une sous-martingale dès que $\phi(X_n)$ est intégrable.*

Démonstration. — Nous laissons les détails au lecteur à titre d'exercice. Nous donnons juste quelques indications.

Le point 1 repose sur le principe des conditionnements successifs 5 de la section 2, par une récurrence immédiate.

Le point 2 s'obtient en conditionnant par $\mathcal{F}_{m-1}$ lorsque $n < m$.

Les points 3 et 4 sont immédiats.

Les point 5 et 6 reposent sur l'inégalité de JENSEN conditionnelle 3 de la section 2.

∎

Pour les martingales de carré intégrable, nous avons en outre

Proposition 1.6. *Si M_n est une martingale de carré intégrable,*

$$\forall n \leq p, \ \mathbf{E}[(M_p - M_n)^2] = \sum_{n+1}^{p} \mathbf{E}[(\Delta M)_k^2].$$

Démonstration. — Il suffit d'appliquer la propriété d'accroissements orthogonaux 2 de la proposition précédente 1.5. ∎

Corollaire 1.7. *Une martingale bornée dans L^2 converge dans L^2.*

Démonstration. — Par définition, dire que martingale est bornée dans L^2 revient à dire qu'il existe une constante C telle que, pour tout n,

$$\mathbf{E}(X_n^2) \leq C^2.$$

Dans ce cas,

$$\mathbf{E}(X_n - X_0)^2 \leq 4C^2,$$

et la proposition précédente 1.6 montre que la série

$$\sum_k \mathbf{E}[(\Delta M)_k^2]$$

est convergente. Par conséquent,

$$\lim_{n\to\infty} \sup_{p\geq q\geq n} \sum_{q}^{p} \mathbf{E}[(\Delta M)_k^2] = 0.$$

En réutilisant la proposition précédente, on obtient

$$\lim_n \sup_{p,q\geq n} \mathbf{E}(M_p - M_q)^2 = 0,$$

et la suite est donc de Cauchy dans L^2 : elle y converge.

∎

Nous verrons dans la section 8 des théorèmes beaucoup plus forts.

2 Décompositions

La première proposition décompose une sous-martingale en la somme d'une martingale et d'une suite croissante de variables aléatoires :

Théorème 2.1. Décomposition de DOOB *: soit X une sous-martingale ; il existe une martingale M et un processus croissant prévisible A, nul en 0, uniques, tels que pour tout entier n, $X_n = M_n + A_n$.*

Le processus A est appelé "compensateur' ' de X.
Démonstration. — On part de $A_0 = 0$ et $M_0 = X_0$. Puis pour $n \geq 1$, on définit A_n de la façon suivante : on pose $\Delta_n = \mathbf{E}(X_n/\mathcal{F}_{n-1}) - X_{n-1}$, et

$$A_n = \Delta_1 + \cdots + \Delta_n.$$

On pose enfin $M_n = X_n - A_n$. Par construction A_n est prévisible, et puisque X_n est une sous-martingale, $\Delta_n \geq 0$, et donc A_n est croissante. Ensuite,

$$\mathbf{E}(M_{n+1}/\mathcal{F}_n) = \mathbf{E}(X_{n+1}/\mathcal{F}_n) - A_{n+1} = X_n + \Delta_n - A_{n+1} = M_n.$$

et donc M_n est une martingale.

L'unicité de la décompositon s'écrit de la même façon, et remarquant que, si une telle décomposition a lieu, on doit avoir

$$\mathbf{E}(X_{n+1} - X_n/\mathcal{F}_n) = A_{n+1} - A_n,$$

ce qui caractérise A_n si l'on sait que $A_0 = 0$. ∎

Explicitons cette décomposition dans le cas particulier des martingales de carré intégrable.

Proposition 2.2. *Soit M_n une martingale de carré intégrable. Rappelons (notation 1.4) que $(\Delta M)_n = M_n - M_{n-1}$ et posons*

$$U_n = \mathbf{E}[(\Delta M)_n^2/\mathcal{F}_{n-1}).$$

Alors $M_n^2 - \sum_{k=1}^n U_k$ est une martingale.

Démonstration. — C'est la décomposition précédente dans le cas de la sous-martingale M_n^2, en remarquant que

$$\mathbf{E}[(\Delta M)_n^2/\mathcal{F}_{n-1}) = \mathbf{E}(M_n^2/\mathcal{F}_{n-1}] - M_{n-1}^2.$$

Il suffit de développer le double produit et d'utiliser la propriété de martingale.

∎

La seconde décomposition est beaucoup moins utilisée que la première, mais rend quelques services :

Théorème 2.3. Décomposition de Krickeberg *Soit X une martingale bornée dans L^1 ; il existe deux martingales positives Y et Z telles que $X_n = Y_n - Z_n$ pour tout entier n.*

Démonstration. — (On en verra une démonstration élémentaire plus bas dans le cas des martingales uniformément intégrables : voir la remarque 3 qui suit le théorème 8.7.)

On peut tout d'abord supposer que $X_0 = 0$.

L'hypothèse est qu'il existe une constante C telle que pour tout n, $E[|X_n|] \leq C$. Or, la suite $(|X_n|, n \in \mathbb{N})$ est une sous martingale. Par la décomposition de Doob (2.1) il existe une martingale M et un processus croissant prévisible A tels que pour tout entier $n, |X_n| = M_n + A_n$. On a $\mathbf{E}(A_n) = \mathbf{E}(|X_n|) \leq C$. La suite croissante A_n est donc bornée dans L^1, et par conséquent converge vers presque sûrement vers une limite intégrable que nous notons A_∞ . On a $|X_n| \leq M_n + A_\infty$, et on peut donc en prendre l'espérance conditionnelle :

$$\forall n, |X_n| \leq M_n + E[A_\infty/\mathcal{F}_n] = Y_n.$$

La suite Y est une martingale positive, $Y - X$ est aussi une martingale. Elle est positive puisque $Y \geq |X| \geq X$. ∎

3 Transformation des martingales

La propriété de martingale (et également de sur- et sous-martingale) a l'énorme avantage d'être stable par un certain nombre de transformations élémentaires qui ne seraient pas possibles dans le cadre de l'étude des sommes de variables aléatoires indépendantes.

Ces transformations reposent sur le principe suivant :

Proposition 3.1. *Soit (X_n) un processus adapté et (V_n) un processus prévisible tel que, pour tout n, la variable $V_n(X_n - X_{n-1})$ soit intégrable. Alors, on définit le processus $(V.X)$ de la façon suivante*

$$(V.X)_n = V_0X_0 + \sum_{k=1}^{n} V_k(X_k - X_{k-1}).$$

Alors, si X est une martingale, il en va de même de $(V.X)$. Si X est une sur- (resp. sous-) martingale, et si V est de plus positif, alors $(V.X)$ est une sur- (resp. sous-) martingale.

Démonstration. — En reprenant la notation 1.4 , le processus $(V.X)$ vérifie

$$(\Delta(V.X))_n = V_n(\Delta X)_n.$$

La démonstration est alors immédiate et laissée au lecteur. ∎

Dans un casino par exemple, un processus V correspond à une stratégie du joueur : en fonction de toutes les observations dont il dispose à l'instant n, il va miser à l'instant $n+1$ une somme V_{n+1}, pour obtenir un gain $V_{n+1}(X_{n+1} - X_n)$.

Un cas particulier important de la proposition 3.1 est le suivant :

Corollaire 3.2. *Soit (X_n) une martingale (resp. une sous-, une sur-martingale), et soit T un temps d'arrêt. Alors le processus X^T défini par $X_n^T = X_{T\wedge n}$ est une martingale (resp. une sous-, une sur-martingale).*

Démonstration. — Il suffit de considérer le processus $V = \mathbf{1}_{[0,T]}$, dont on a déjà vu qu'il est prévisible. Dans ce cas, le processus $(V.X)$ n'est rien d'autre que X^T :

$$(V.X)_n = X_0 + \sum_{k=1}^{T\wedge n}(X_k - X_{k-1}) = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})\mathbf{1}_{k\leq T}.$$

∎

remarquons au passage que le processus $\mathcal{F}_{T\wedge n}$ est en fait adapté à la filtration $\mathcal{G}_n = \mathcal{F}_{T\wedge n}$ plus petite que $\mathcal{F}_n$.

De même, en choisissant le processus prévisible $V = \mathbf{1}_A\mathbf{1}_{[T,\infty[}$, avec $A \in \mathcal{F}_T$, nous obtenons

Corollaire 3.3. *Si T est un temps d'arrêt, et nous voyons que $\mathbf{1}_A(X_{T\vee n} - X_{T)})$ est une martingale (resp. une sous-, une sur-martingale).*

4 Théorème d'arrêt : le cas des temps bornés.

Dans cette section, nous énonçons le théorème d'arrêt pour les *temps d'arrêt* bornés : il sera étendu plus tard aux *temps d'arrêt* quelconques, mais uniquement pour les martingales uniformément intégrables.

Théorème 4.1. *(***Théorème d'arrêt***.)*

Soit $(X_n, n \in \mathbb{N})$ *une martingale et* S *et* T *deux temps d'arrêt bornés (c'est à dire qu'il existe un entier* n *tel que* $S \vee T \leq n$*, presque sûrement). Alors*

$$\mathbf{E}(X_T/\mathcal{F}_S) = X_{S\wedge T}. \tag{4.2}$$

De même, si X *est une sous- (resp une sur-)martingale, on a*

$$X_{S\wedge T} \leq \text{ (resp } \geq \text{) } \mathbf{E}(X_T/\mathcal{F}_S). \tag{4.3}$$

En particulier, si $(X_n, n \in \mathbb{N})$ *est une sous-martingale et* S *et* T *sont deux temps d'arrêt bornés, alors*

$$\mathbf{E}(X_S \mathbf{1}_{S\leq T}) \leq \mathbf{E}(X_T \mathbf{1}_{S\leq T}). \tag{4.4}$$

On a bien sûr une inégalité inverse pour les sur-martingales.

Démonstration. — Nous n'allons démontrer le résultat que pour les martingales, le cas des sur et sous- martingales se traitant de la même façon.

Nous allons d'abord montrer le résultat dans le cas particulier où $T = n$ et $S \leq n$. L'égalité (4.2) à établir s'écrit alors

$$X_S = \mathbf{E}(X_n/\mathcal{F}_S).$$

Par définition,

$$X_S = \sum_{k=0}^{n} X_k \mathbf{1}_{S=k}.$$

Nous savons déjà que X_S est $\mathcal{F}_S$ mesurable ; elle est de plus intégrable comme combinaison linéaire finie de variables intégrables.

Il nous suffit donc de démontrer que, pour tout $A \in \mathcal{F}_S$,

$$\mathbf{E}(X_S \mathbf{1}_A) = \mathbf{E}(X_n \mathbf{1}_A).$$

Ceci s'écrit
$$\sum_{k=0}^{n} \mathbf{E}(X_k \mathbf{1}_{A\cap\{S=k\}}) = \sum_{k=0}^{n} \mathbf{E}(X_n \mathbf{1}_{A\cap\{S=k\}}).$$
Mais, puisque $A \in \mathcal{F}_S$, alors $A \cap \{S = k\} \in \mathcal{F}_k$, et la propriété de martingale s'applique pour donner, pour tout $k \leq N$,
$$\mathbf{E}(X_k \mathbf{1}_{A\cap\{S=k\}}) = \mathbf{E}(X_n \mathbf{1}_{A\cap\{S=k\}}).$$

Pour passer au cas général, nous choisissons un entier n qui majore à la fois S et T. Appliquons le cas particulier qui précède à la martingale arrêtée X^T, et au temps d'arrêt S. Nous avons $X_n^T = X_T$ puisque $T \leq n$, $X_S^T = X_{S\wedge T}$. Nous avons donc
$$\mathbf{E}(X_T/\mathcal{F}_S) = X_{S\wedge T}.$$

Remarquons que la variable $X_{S\wedge T}$ est $\mathcal{F}_{S\wedge T}$ mesurable, et par conséquent on a aussi
$$X_{S\wedge T} = \mathbf{E}(X_T/\mathcal{F}_{S\wedge T}).$$

Pour établir l'inégalité (4.4), il suffit de voir que l'inégalité (4.3) s'écrit, pour tout $A \in \mathcal{F}_S$
$$\mathbf{E}(X_{S\wedge T}\mathbf{1}_A) \leq \mathbf{E}(X_T \mathbf{1}_A).$$
On l'applique alors à l'ensemble $A = \{S \leq T\}$.

∎

Corollaire 4.2. *Soit (T_n) une suite croissante de temps d'arrêt bornés, et X une martingale (resp. une sous-martingale, une sur-martingale) ; alors $(X_{T_n}, n \in \mathbb{N})$ est une martingale (resp. une sous-martingale, une sur-martingale) pour la filtration $(\mathcal{F}_{T_n}, n \in \mathbb{N})$.*

Démonstration. — (en exercice) ∎

Corollaire 4.3. *Soit X est une variable intégrable, et soit (X_n) la martingale $\mathbf{E}(X/\mathcal{F}_n)$. Si T est un temps d'arrêt borné, alors*
$$\mathbf{E}(X/\mathcal{F}_T) = X_T.$$

Si S et T sont deux temps d'arrêt bornés,
$$\mathbf{E}(X/\mathcal{F}_S/\mathcal{F}_T) = \mathbf{E}(X/\mathcal{F}_T/\mathcal{F}_S) = \mathbf{E}(X/\mathcal{F}_{S\wedge T}) = X_{S\wedge T}.$$
En d'autres termes, $\mathcal{F}_S$ et $\mathcal{F}_T$ sont conditionnellement indépendantes sachant $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.

Démonstration. — Soit N un majorant commun à T et S. On considère la martingale $X_n = \mathbf{E}(X/\mathcal{F}_n)$. Par le principe des conditionnements successifs, $\mathbf{E}(X/\mathcal{F}_T) = \mathbf{E}(X_N/\mathcal{F}_T)$. Puis le théorème d'arrêt nous donne $\mathbf{E}(X_N/\mathcal{F}_T) = X_T$.

Si S et T sont deux temps d'arrêt bornés,

$$\mathbf{E}(X_T/\mathcal{F}_S) = X_{S\wedge T} = \mathbf{E}(X/\mathcal{F}_{S\wedge T}).$$

Ensuite, on retrouve la caractérisation de l'indépendance conditionnelle : deux tribus sont conditionnellement indépendantes par rapport à leur intersection si et seulement si les espérances conditionnelles par rapport à ces tribus commutent (section 4).

∎

5 Un exemple d'application : le problème de l'arrêt optimal.

Le problème de l'arrêt optimal se rencontre dans la pratique lorsque, connaissant la loi d'une suite de variables aléatoires X_n, on cherche à maximiser un profit ou à minimiser un coût.

Dans une filtration $(\mathcal{F}_n)$, on se donne un entier N et une suite adaptée de variables aléatoires intégrables $(V_n)_{n\leq N}$, puis on cherche à trouver un *temps d'arrêt* T inférieur à N qui maximise $\mathbf{E}(V_T)$.

Pour répondre on peut d'abord déterminer l'espérance optimale sachant $\mathcal{F}_n$ que l'on notera $\hat{V}_n$ et qui se définit naturellement par récurrence descendante :

$$\hat{V}_N = V_N;\ \hat{V}_n = \max\{V_n, \mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n)\},\quad (n = 0, \cdots, N-1).$$

Autrement dit, on choisit au temps n la meilleure option entre s'arrêter sur la valeur V_n et continuer avec comme perspective $\mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n)$ qui ne peut se calculer que si l'on a déjà déterminé $\hat{V}_{n+1}$. Remarquons que $\hat{V}_n$ est une sur-martingale. En fait, c'est la plus petite sur-martingale qui majore (V_n). (Elle n'est définie que pour $n \leq N$, mais le lecteur formaliste pourra la prolonger à tous les entiers en la supposant arrêtée à N.) Elle s'appelle **l'enveloppe de Snell** de la suite (V_n).

Un premier *temps d'arrêt* optimisant $\mathbf{E}(V_T)$ est

$$T_1 = \inf\{n \mid V_n = \hat{V}_n\} = \inf\{n \mid V_n \geq \mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n)\}.$$

Mais quand l'égalité est possible il existe un autre *temps d'arrêt* optimal :

$$T_2 = \inf\{n \mid V_n > \mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n)\} \wedge N.$$

Et plus généralement nous montrons dans le prochain théorème qu'un *temps d'arrêt* $T \leq T_2$ tel que $V_T = \hat{V}_T$ est optimal. La différence entre les stratégies induites par T_1 et par T_2 est que dans le premier cas, si continuer donne exactement la même espérance que s'arrêter, on s'arrête, alors que dans le deuxième cas on continue, ce qui donne la même espérance $\mathbf{E}(V_T)$ mais peut donner des lois de V_T différentes.

A titre d'exercice, on pourra considérer $X_0 = 0$, $X_n = Z_1 + \ldots + Z_n$ où les Z_i sont indépendantes et de loi de Bernouilli valant $+1$ ou -1 avec probabilité $1/2$. Pour $0 < a \leq 1$ on définit la suite $V_n = a\mathbf{1}_{X_n>0} - \mathbf{1}_{X_n<0}$ qui est adaptée à la filtration $\mathcal{F}_n = \mathcal{F}_n^X$. Il est clair que dès que $X_n > 0$ il est judicieux de s'arrêter car on ne peut pas espérer mieux que a. Pour les mêmes raisons il ne coûte rien de continuer quand $X_n < 0$ car on ne peut pas espérer pire que -1 (encore que si l'on suit la stratégie de T_1 on s'arrêterait si $X_n < n - N$ car on ne pourrait pas espérer mieux non plus).

Il est plus délicat de décider si l'on s'arrête ou non quand $X_n = 0$, cela dépend du temps restant $N - n$. Remarquons que l'on n'a pas considéré les cas $a > 1$ pour lesquels avoir une chance sur deux d'obtenir a suffit pour se motiver à continuer. D'autre part si $a = 1$ et $N = 1$, on remarquera que $T_1 = 0$ et $T_2 = 1$, ce qui donne $\mathbf{E}(V_{T_1}) = \mathbf{E}(V_{T_2}) = 0$ mais $V_{T_1} = 0$ et $V_{T_2} = X_1$, et ont donc des lois différentes.

Enfin pour observer le calcul des $\hat{V}_n$, prenons $N = 2$ et répondons à la question : au temps 0, $X_0 = 0$, faut-il s'arrêter tout de suite ? On initialise la récurrence descendante par $\hat{V}_2 = V_2$ puis

$$\begin{array}{rcl} V_1 & = & a\mathbf{1}_{X_1>0} - \mathbf{1}_{X_1<0} \\ \text{et} \quad \mathbf{E}(\hat{V}_2/\mathcal{F}_1) & = & a\mathbf{1}_{X_1>1} + \dfrac{a}{2}\mathbf{1}_{X_1=1} + \dfrac{a-1}{2}\mathbf{1}_{X_1=0} + \dfrac{-1}{2}\mathbf{1}_{X_1=-1} - \mathbf{1}_{X_1<-1} \\ \text{donc} \quad \hat{V}_1 & = & a\mathbf{1}_{X_1>0} + \dfrac{-1}{2}\mathbf{1}_{X_1=-1} - \mathbf{1}_{X_2<-1}. \end{array}$$

Ainsi $\mathbf{E}(\hat{V}_1/\mathcal{F}_0) = \mathbf{E}(\hat{V}_1) = \dfrac{2a-1}{4}$, à comparer avec $V_0 = 0$. Au temps 0, il faut s'arrêter si $a < 1/2$ et continuer si $a > 1/2$. On trouvera une étude plus approfondie de cet exemple dans l'exercice **??**.

Théorème 5.1. *Soit* $(V_n)_{0\leq n\leq N}$ *une suite de variables aléatoires intégrables adaptée à une filtration* $(\mathcal{F}_n)$. *On considère deux temps d'arrêt* S *et* T *bornés* *par N. Si avec les notations ci-dessus $T \leq T_2$ et $V_T = \hat{V}_T$, alors*

$$\mathbf{E}(V_S) \leq \mathbf{E}(V_T).$$

Démonstration. — Remarquons que le processus arrêté $\hat{V}^T$ est une martingale, si $n \leq N-1$ on a

$$\mathbf{E}(\hat{V}_{(n+1)\wedge T}/\mathcal{F}_n) = \hat{V}_{n\wedge T}.$$

Pour le voir, nous décomposons

$$\begin{aligned}\mathbf{E}(\hat{V}_{(n+1)\wedge T}/\mathcal{F}_n) &= \mathbf{E}(\hat{V}_T \mathbf{1}_{T\leq n}/\mathcal{F}_n) + \mathbf{E}(\hat{V}_{n+1}\mathbf{1}_{T\geq n+1}/\mathcal{F}_n)\\ &= \hat{V}_T\mathbf{1}_{T\leq n} + \hat{V}_n\mathbf{1}_{T\geq n+1} = \hat{V}_{n\wedge T}.\end{aligned}$$

En effet $\hat{V}_T\mathbf{1}_{T\leq n}$ est $\mathcal{F}_n$-mesurable, cela se voit directement ou par la propriété 8 et la proposition 0.8. Pour interpréter la seconde espérance conditionnelle on en a d'abord extrait $\mathbf{1}_{T\geq n+1}$ qui est $\mathcal{F}_n$-mesurable puis on a observé que sur $\{T \geq n+1\} \subset \{T_2 \geq n+1\}$, $\mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n) \geq V_n$ donc $\mathbf{E}(\hat{V}_{n+1}/\mathcal{F}_n) = \hat{V}_n$.

On peut maintenant commencer la majoration de $\mathbf{E}(V_S)$,

$$\mathbf{E}(V_S) \leq \mathbf{E}(\hat{V}_S) = \mathbf{E}(\hat{V}_S\mathbf{1}_{S\geq T}) + \mathbf{E}(\hat{V}_S\mathbf{1}_{S<T}).$$

Pour le premier terme, on applique (4.4) à la sur-martingale $\hat{V}$:

$$\mathbf{E}(\hat{V}_S\mathbf{1}_{S\geq T}) \leq \mathbf{E}(\hat{V}_T\mathbf{1}_{S\geq T}) = \mathbf{E}(V_T\mathbf{1}_{S\geq T}).$$

Le second terme se réécrit avec V_T en appliquant (4.4) à la martingale $\hat{V}^T$:

$$\mathbf{E}(\hat{V}_S\mathbf{1}_{S<T}) = \mathbf{E}(\hat{V}_S^T\mathbf{1}_{S<T}) = \mathbf{E}(\hat{V}_T^T\mathbf{1}_{S<T}) = \mathbf{E}(V_T\mathbf{1}_{S<T}).$$

D'où la majoration de $\mathbf{E}(V_S)$ par $\mathbf{E}(V_T)$. ∎

6 Théorème d'arrêt : le cas des temps finis.

Nous allons étendre dans cette section les théorèmes d'arrêt de la section 4 au cas des temps finis, moyennant quelques conditions supplémentaires d'intégrabilité sur les martingales (resp. les sous-martingales, les sur-martingales).

Proposition 6.1. *Soit $(X_n, n \in \mathbb{N})$ une martingale (resp. une sous-martingale) et T et S deux temps d'arrêt presque sûrement finis.*

Si les suites $(X_{T\wedge n})$ *et* $(X_{S\wedge n})$ *sont uniformément intégrables, alors*

$$X_{S\wedge T} = E[X_T/\mathcal{F}_S]. \quad \textit{(resp. } X_{S\wedge T} \leq E[X_T/\mathcal{F}_S]).$$

Cela se produit en particulier lorsqu'il existe $Y \in L^1$ *telle que pour tout* n, $|X_{T\wedge n}| \leq Y$*), ou bien lorsque la suite* (X_n) *elle même est uniformément intégrable*

En particulier, dans le cas des martingales, on a $\mathbf{E}(X_T) = \mathbf{E}(X_0)$ *pour tout temps d'arrêt fini satisfaisant à l'hypothèse ci-dessus. Ce résultat sera étendu plus bas au cas des temps d'arrêt quelconques (finis ou non). (Voir les remarques 2 à la suite de la proposition 8.7).*

Démonstration. — Prenons le cas des martingales, le cas des sous-martingales se traitant de la même manière.

On remarque tout d'abord que la variable X_T est intégrable, puisque limite presque sûre de la suite $(X_{T\wedge n})$ qui est uniformément intégrable par hypothèse. (On se sert ici du fait que T est presque sûrement fini, ne serait-ce que pour pouvoir définir X_T.) Il en va de même de X_S. Nous voulons démontrer que, pour toute variable Z $\mathcal{F}_S$-mesurable bornée, on a

$$\mathbf{E}(X_S Z) = \mathbf{E}(X_T Z).$$

On peut en utilisant le TCM (1.2) se rammener au cas où Z est dans l'une des tribus $\mathcal{F}_{S\wedge n}$ grâce à la formule (0.7).

Nous fixons alors un tel n et appliquons le théorème d'arrêt pour les *temps d'arrêt* bornés (4.1) aux *temps d'arrêt* $S\wedge p$ et $T\wedge p$, pour $p \geq n$. On a alors

$$\mathbf{E}(ZX_{S\wedge p}) = \mathbf{E}(ZX_{T\wedge p}).$$

On peut alors passer à la limite en utilisant l'intégrabilité uniforme.

Pour voir que l'hypothèse s'applique en particulier lorsque la suite (X_n) est elle même uniformément intégrable, il suffit de voir que $X_{T\wedge n} = \mathbf{E}(X_n/\mathcal{F}_T)$, et de se rappeler que les espérances conditionnelles d'une famille uniformément intégrable est elle-même uniformément intégrable (cf. proposition 5.3).

∎

Exemple : le problème de la ruine du joueur.

Nous reprenons l'exemple fondamental 2 de la section 3. On dispose d'une suite de variables Z_n, qui est le résultat d'un jeu au temps n, et qui suit une loi de Bernoulli valant $+1$ ou -1 avec probabilité $p, 1-p$. La fortune initiale du joueur est a et il espère ateindre la quantité $b > a$. Le jeu est terminé lorsque la fortune du joueur est soit nulle, auquel cas il a perdu, soit égale à b, auquel cas il a gagné. On note T l'instant de la fin du jeu.

On souhaite répondre aux questions suivantes :

1. Quelle est la probabilité $\mathbf{P}(G)$ que le joueur gagne ?
2. Quelle est l'espérance de T ?

Traitons d'abord le cas où $\mathbf{E}(Z_i) = m = 2p - 1$ est non nul.

Tout d'abord, il faut montrer que le temps d'atteinte T est fini presque sûrement. Si X_n est la fortune du joueur à l'instant n, $X_n - mn$ est une martingale. En considérant le temps $T \wedge n$, et la martingale précédente, puisque $|X_{T\wedge n}| \leq b$, on obtient $\mathbf{E}(T \wedge n) \leq b/m$, lorsque $m > 0$, puis en passant à la limite $\mathbf{E}(T) \leq b/m$.

On en déduit que le temps T est fini (ps) lorsque $m > 0$. Par symétrie, on déduit de même que T est fini (ps) lorsque $m < 0$ (On obtient alors $\mathbf{E}(T) \leq -b/m$.

On remarque ensuite que, si $r = \frac{1-p}{p}$, alors $\mathbf{E}(r^{Y_n}) = 1$, et donc que r^{X_n} est aussi une martingale ; en lui appliquant le théorème d'arrêt au temps $T \wedge n$, on obtient $\mathbf{E}(r^{T\wedge n}) = r^a$. Si $m > 0$, alors $r > 1$ et le premier membre est croissant en n, et si $m < 0$, $r < 1$ et le premier membre est borné. On peut donc selon les cas appliquer le théorème de convergence monotone ou dominée pour passer à la limite en n et obtenir $\mathbf{E}(r^T) = r^a$. Ceci donne $r^b\mathbf{P}(G) + 1 - \mathbf{P}(G) = r^a$, d'où

$$\mathbf{P}(G) = \frac{r^a - 1}{r^b - 1}.$$

Par ailleurs, puisque $X_T = b\mathbf{1}_G$, on obtient, grâce à la première martingale, $\mathbf{E}(T) = (b\mathbf{P}(G) - a)/m$.

Lorsque $m = 0$ est nul, X_n est elle même une martingale, mais on remplace la deuxième martingale par $X_n^2 - n$, qui est encore une martingale. On utilise alors un raisonnement analogue, et on trouve

$$\mathbf{P}(G) = a/b \quad \text{et} \quad \mathbf{E}(T) = a(b - a).$$

On peut déduire de ce qui précède d'autres informations. Par exemple, dans le cas où $m = 0$, que le temps d'arrêt $T_b = \inf\{n \mid X_n = b\}$ est fini presque sûrement. En effet, supposons pour simplifier que la martingale X_n parte de

0 à l'instant n, et appelons $T_{b,k} = T_b \wedge T_{-k} = \inf\{n \mid X_n \in \{-k, b\}\}$. En décalant de k le résultat précédent, nous trouvons que $\mathbf{P}(T_b < T_{-k}) = \frac{k}{b+k}$. Puisque T_{-k} croît vers l'infini avec k, nous obtenons en passant à la limite que $\mathbf{P}(T_b < \infty) = 1$. Mais observons qu'on ne peut pas appliquer le théorème d'arrêt au temps T_b, ce qui mènerait à une contradiction. Á titre d'exercice, le lecteur pourra calculer $\mathbf{P}(T_b < \infty)$, pour $b > 0$, dans le cas où $m < 0$.

Enfin, nous énonçons une dernière conséquence du théorème d'arrêt.

Corollaire 6.2. *(Identité de* WALD*) Soit* $(Y_n, n \in \mathbb{N})$ *une suite de variables aléatoires réelles adaptées, de même loi,d'espérance* m, *et pour tout* n, Y_n *est indépendante de* $\mathcal{F}_{n-1}$. *On pose* $S_n = Y_1 + \cdots + Y_n$. *Alors, si* T *un temps d'arrêt intégrable,* $E[S_T] = E[Y_1]E[T]$.

Démonstration. — On montre que le processus $(X_n = S_n - nm, n \geq 0)$ est une martingale. Alors, le théorème est immédiat pour des temps d'arrêt bornés.

Puis on traite d'abord le cas où les variables Y_i sont positives, en appliquant le théorème à $T \wedge n$ et en passant à la limite par convergence monotone. Enfin, pour passer au cas général, on observe que $|S_{T\wedge n}|$ est majorée par $\hat{S}_T$, où $\hat{S}_n = \sum_0^n |Y_i|$, variable intégrable d'après ce qui précède. On peut donc appliquer le théorème de convergence dominée. (On aurait pu tout aussi bien appliquer la décomposition de KRICKEBERG.) ∎

Grâce à la proposition 1.3, on applique en fait souvent ce résultat au cas d'une variable T indépendante de la suite (X_n), car c'est alors automatiquement un *temps d'arrêt* de la filtration $\mathcal{G}_n = \mathcal{F}_n \vee \sigma(T)$.

Voici un exemple d'application de cette identité de WALD. Soit une suite $(X_n)_{n\geq 1}$ de variables aléatoires réelles, intégrables, indépendantes, de même loi, d'espérance $\mu > 0$.

On note $S_n = \sum_{i=1}^n X_i$. Soit N une variable aléatoire de loi géométrique de paramètre p indépendante de la suite $(X_n)_{n\geq 1}$. Alors, $\frac{S_N}{\mathbf{E}(S_N)}$ converge en loi vers une variable aléatoire exponentielle de paramètre 1.
En effet, nous pouvons appliquer l'identité de Wald à la suite (X_n) adaptée à la filtration $(\mathcal{F}_n)$ définie par $\mathcal{F}_n = \sigma(X_0, ..., X_n, N)$, et au temps d'arrêt N et ainsi $\mathbf{E}(S_N) = \frac{\mu}{p}$. Or d'après la loi des grands nombres, la suite $\frac{S_n}{n\mu}$ converge presque sûrement vers 1, et les variables aléatoires Np convergent en loi vers une variable aléatoire continue de loi exponentielle de paramètre 1. Soit $t > 0$ nous déduisons de ce qui précède,

$$\lim_{p\mapsto 0} \mathbf{P}(\frac{S_N}{\mathbf{E}(S_N)} \leq t) = \lim_{p\mapsto 0} \mathbf{P}(\frac{S_N}{N\mu} pN \leq t) = 1 - e^{-t}.$$

7 Inégalités.

Théorème 7.1. *(Inégalité Maximale de* Doob*.) Soit* $(X_n, n \in \mathbb{N})$ *une sous martingale positive et* $\lambda \geq 0$*. Appelons* $X_n^* = \sup_{k=0}^n X_k$*. Alors*

$$\forall n \in \mathbb{N},\ \lambda \mathbf{P}\{X_n^* \geq \lambda\} \leq \mathbf{E}[X_n \mathbf{1}_{\{X_n^* \geq \lambda\}}] \leq E[X_n].$$

Démonstration. — Considérons le *temps d'arrêt* $T = \inf\{k \in \mathbb{N}, X_k \geq \lambda\}$. Alors,

$$\{T \leq n\} = \{X_n^* \geq \lambda\}.$$

Posons $S = T \wedge (n+1)$, qui est un *temps d'arrêt* borné. On a

$$A = \{S \leq n\} = \{T \leq n\} \in \mathcal{F}_S.$$

Ecrivons le théorème d'arrêt pour cette sous-martingale, entre les temps n et $S \wedge n$.

$$\mathbf{E}(X_{S\wedge n} \mathbf{1}_A) \leq \mathbf{E}(X_n \mathbf{1}_A).$$

Cela s'écrit encore

$$\mathbf{E}(X_T \mathbf{1}_{T \leq n}) \leq \mathbf{E}(X_n \mathbf{1}_{T \leq n}). \tag{7.5}$$

Or, sur $\{T \leq n\}$, $X_T \geq \lambda$, et on obtient l'inégalité cherchée en minorant le membre de gauche de l'inégalité (7.5) par $\lambda \mathbf{P}(T \leq n)$. ∎

Corollaire 7.2. *Si* $(X_n, n \in \mathbb{N})$ *est une martingale,* $(|X_n|, n \in \mathbb{N})$ *est une sous martingale positive et on obtient donc*

$$\forall n \in \mathbb{N},\ \lambda \mathbf{P}\{\max_{k \leq n} |X_k| \geq \lambda\} \leq E[|X_n| \mathbf{1}_{\max_{k\leq n} |X_k| \geq \lambda}] \leq E[|X_n|].$$

Théorème 7.3. *Soit* $(X_n, n \in \mathbb{N})$ *une sous martingale positive et* $p > 1$*. Alors*

$$\|X_n^*\|_p \leq \frac{p}{p-1} \|X_n\|_p.$$

Démonstration. — Il suffit évidemment de se restreindre au cas où X_n est dans L^p, et il en ira de même de toutes les variables X_k, pour $k \leq n$.

Or, pour toute variable aléatoire positive U de L^p, on a

$$E(U^p) = p \int_0^\infty t^{p-1} \mathbf{P}(U \geq t) dt.$$

On a donc

$$\begin{aligned} E[(X_n^*)^p] &= p\int_0^\infty t^{p-1}\mathbf{P}(X_n^* \geq t)dt \\ &\leq p\int_0^\infty t^{p-2}E[X_n \mathbf{1}_{\{X_n^* \geq t\}}]dt \\ &= pE[X_n\int_0^\infty t^{p-2}\mathbf{1}_{\{X_n^* \geq t\}}dt] \\ &= \frac{p}{p-1}E[X_n(X_n^*)^{p-1}], \end{aligned}$$

que l'on majore par l'inégalité de Hölder

$$E[X_n(X_n^*)^{p-1}] \leq \|X_n\|_p \|X_n^*\|_p^{p-1}.$$

Puisque X_n^* est majorée par $\sum_0^n X_k$, c'est une variable de L^p, et on peut donc simplifier pour obtenir le résultat. ∎

En particulier, on en déduit

Corollaire 7.4. *Soit (X_n) une sous-martingale positive bornée dans L^1. Alors la variable $X^* = \sup_n X_n$ est finie presque sûrement. Si (X_n) est bornée dans L^p $(p > 1)$, alors X^* est dans L^p.*

(Ce dernier résultat est faux pour $p = 1$.)

Les mêmes conclusions sont vraies pour les martingales (non nécessairement positives).

Démonstration. — La suite X_n^* converge en croissant vers X^*. Il suffit donc d'appliquer l'inégalité de Doob pour obtenir

$$\lambda \mathbf{P}(X_n^* > \lambda) \leq \sup_n \mathbf{E}(|X_n|) = K < \infty.$$

En passant à la limite

$$\lambda \mathbf{P}(X^* > \lambda) \leq K,$$

et donc $\mathbf{P}(X^* > \lambda) \to 0$ $(\lambda \to \infty)$. Ceci montre la finitude de X^*.

La seconde assertion se traite de même en utilisant le théorème 7.3.

On passe des sous-martingales positives aux martingales en changeant X_n en $|X_n|$.

∎

Remarque. — La démonstration du théorème 7.3, on aurait pu remplacer la fonction t^p par la fonction $t\log_+ t$. Nous aurions vu que si

$$\sup_n \mathbf{E}(X_n \log_+ X_n) < \infty,$$

alors X^* est dans L^1. C'est la meilleure condition possible sur l'intégrabilité des variables X_n qui assure l'intégrabilité de X^*.

8 Convergence des martingales

Dans cette section, nous allons démontrer le théorème fondamental de convergence presque sûre des martingales.

Commençons par établir une inégalité qui contrôle le nombre d'oscillations d'une sous-martingale entre deux valeurs, et plus exactement son nombre de descentes à travers un intervalle.

Soit $x = (x_n)$ une suite de réels. Pour tout $n \geq 0$, pour tout couple de réels (a, b) tels que $a < b$, on définit le **nombre de descentes** $D_n^{a,b}(x)$ de la suite x sur $\{0, \cdots, n\}$ à travers l'intervalle (a, b) en comptant les passages au dessus de b et en dessous de a de la façon suivante : on pose $A_0 = B_0 = 0$, puis, par récurrence

$$B_n = \inf\{p \geq A_{n-1} \mid x_p > b\};\ A_n = \inf\{p \geq B_n \mid x_p < a\},$$

où nous convenons comme d'habitude que $\inf\{\emptyset\} = +\infty$. Alors,

$$D_n^{a,b}(x) = \sum_{p=1}^{\infty} 1_{\{A_p < n\}}.$$

Ce nombre compte exactement le nombre de passages au dessus de b de la suite qui sont suivis strictement avant n d'un passage en dessous de a.

Le nombre total de passages de la suite (x_n) entre a et b sera défini comme la limite croissante $D^{a,b}(x)$ des nombres $D_n^{a,b}(x)$ lorsque $n \to \infty$.

Remarquons que la somme qui apparaît dans la définition de D_n est toujours finie, car $A_n > B_{n-1} > A_{n-1}$, si bien que $A_n \geq A_{n-1} + 2$, et donc $A_n \geq 2n$.

D'autre part, si $X = (X_n)$ est une suite de variables aléatoires, alors $D_n^{a,b}(X)$ est une variable aléatoire ; lorsque n croît, ces variables forment une suite croissante. De plus, si la suite (X_n) est adaptée à une filtration $\mathcal{F}_n$, alors les variables aléatoires A_n et B_n intervenant dans la définition de $D_n^{a,b}(X)$ sont des *temps d'arrêt*.

L'intérêt de ces nombres de descentes est le suivant.

Si $x = (x_n)$ une suite de nombres réels, nous posons $x^* = \sup_n |x_n|$. Alors :

Lemme 8.1. *Soit (x_n) une suite de nombres réels. Une CNS pour que la suite converge est que x^* soit fini et que, pour tout couple de rationnels (a, b) tel que $a < b$, la quantité $D^{a,b}(x)$ soit finie.*

Remarque. — Nous nous sommes restreints aux couples de rationnels dans l'énoncé précédent pour des raisons de dénombrabilité. N'importe quel ensemble dense dans $\mathbb{R}$ aurait aussi bien fait l'affaire.

Démonstration. — Montrons que la condition est suffisante. Si x^* est fini, la suite est bornée : appelons M et m les limites supérieures et inférieures de x, qui sont finies. Si $m < M$, choisissons deux rationnels a et b tels que $m < a < b < M$. Alors, on voit immédiatement que $D^{a,b}(x) = \infty$. D'où une contradiction et donc $m = M$, et la suite converge.

Réciproquement, une suite convergente est bornée, et, si un couple (a, b) avec $a < b$ est tel que $D^{a,b}(x) = \infty$, alors $\liminf_n x_n \leq a$ et $\limsup_n x_n \geq b$, ce qui est impossible. ∎

Nous pouvons maintenant énoncer le résultat sur les descentes

Proposition 8.2. *Soit (X_n) une sous martingale positive. Considérons un couple (a, b) avec $a < b$ de réels et appelons $D_n^{a,b}$ le nombre de descentes de X à travers l'intervalle (a, b) sur $\{0, \cdots, n\}$. Alors,*

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

Ici, comme d'habitude, $(X_n - b)_+$ désigne la partie positive de $X_n - b$.

Démonstration. — Fixons n, et commençons par arrêter la sous-martingale à n (c'est à dire remplaçons X par X^n.) Considérons les *temps d'arrêt* intervenant dans la définition des descentes, que nous notons A_p et B_p pour simplifier. Rappelons que $B_p < A_p$ (on compte par convention les passages au dessus de b avant les passages en dessous de a). Posons $A_p^n = A_p \wedge n$ et $B_p^n = B_p \wedge n$, qui sont des *temps d'arrêt* bornés.

D'après le théorème d'arrêt, nous avons

$$\mathbf{E}(X_{A_p^n} - X_{B_p^n}) \geq 0.$$

Décomposons cette espérance sur les ensembles $\{A_p < n\}$ et $\{A_p \geq n\}$. Nous obtenons

$$\mathbf{E}[(X_{B_p^n} - X_{A_p^n})\mathbf{1}_{\{A_p<n\}}] \leq \mathbf{E}[(X_{A_p^n} - X_{B_p^n})\mathbf{1}_{\{A_p\geq n\}}].$$

Or, sur $\{A_p < n\}$,

$$X_{B_p^n} - X_{A_p^n} = X_{B_p} - X_{A_p} \geq b - a.$$

D'autre part, sur $\{A_p \geq n\}$, $X_{A_p^n} - X_{B_p^n} = X_n - X_{B_p^n}$. Cette dernière quantité est nulle sur $\{B_p \geq n\}$. Sur $\{B_p < n\}$, elle est inférieure à $X_n - b$; elle est donc toujours inférieure à $(X_n - b)_+\mathbf{1}_{\{B_p<n\leq A_p\}}$ sur $\{A_p \geq n\}$.

On obtient donc

$$(b-a)\mathbf{E}(\mathbf{1}_{\{A_p<n\}}) \leq \mathbf{E}[(X_n - b)_+\mathbf{1}_{\{B_p<n\leq A_p\}}].$$

Remarquons encore que les ensembles $\{B_p < n \leq A_p\}$ sont disjoints lorsque p varie, puis sommons en p l'inégalité précédente. Le membre de gauche devient $(b-a)\mathbf{E}(D_n^{a,b})$ par définition du nombre de descentes, tandis que le second est inférieur à $\mathbf{E}[(X_n - b)_+]$. ∎

Corollaire 8.3. *Soit (X_n) une sous-martingale positive, bornée dans L^1. Alors, pour tout couple (a,b), $a < b$, le nombre de descentes de la suite (X_n) à travers (a,b) est presque sûrement fini, et la suite (X_n) est presque sûrement convergente.*

Démonstration. — Montrons tout d'abord le premier point. Soit M un majorant des $\mathbf{E}(X_n)$, et un couple de réels (a,b) tel que $a < b$. En appelant D_n le nombre de descentes à travers (a,b) sur l'intervalle $\{0,n\}$, et $D^{a,b}$ le nombre total de descentes, on voit que

$$\mathbf{E}(D_n^{a,b}) \leq \frac{1}{b-a}(M + |b|).$$

En passant à la limite en n, nous voyons qu'alors la variable $D^{a,b}$ est intégrable donc finie presque sûrement.

Appelons $N_{a,b}$ l'événement $\{D^{a,b} = \infty\}$, et

$$N = \bigcup_{(a,b)\in\mathbb{Q},a<b} N_{a,b}.$$

Puisque $\mathbf{P}(N_{a,b}) = 0$, alors $\mathbf{P}(N) = 0$.

D'autre part, nous savons d'après le corollaire 7.4 que la variable $X^* = \sup_n X_n$ est finie presque sûrement. Soit $N_1 = \{X^* = \infty\}$: $\mathbf{P}(N \cup N_1) = 0$, et, si $\omega \notin N \cup N_1$, la suite $(X_n(\omega))$ satisfait aux hypothèse du lemme (rappelons que (X_n) est positive par hypothèse) : elle y est donc convergente. ∎

Le résultat précédent se prolonge immédiatement aux martingales bornées dans L^1, aux sous-et sur-martingales bornées dans L^1 non nécessairement positives :

Corollaire 8.4. *Soit X_n une martingale, ou une sous-martingale, ou une sur-martingale, bornée dans L^1. Alors, X_n converge presque sûrement vers une variable X_∞.*

Démonstration. — Commençons par le cas des martingales : les fonctions $x_+ = \sup(x, 0)$ et $x_- = \sup(-x, 0)$ sont convexes : donc, si $M_.$ est une martingale bornée dans L^1, les processus M_{n+} et M_{n-} sont des sous-martingales positives bornées dans L^1. Elles convergent donc presque sûrement et il en va de même de la martingale $M_n = M_{n+} - M_{n-}$.

Traitons ensuite le cas d'une sous-martingale (X_n) bornée dans L^1 : nous avons vu plus haut qu'elle peut s'écrire $X_n = M_n + A_n$, où (M_n) est une martingale et (A_n) est une suite croissante nulle en 0.

Si nous posons $K = \sup_n \mathbf{E}(|X_n|)$, on a $E(|X_0|) \leq K$, et donc

$$|\mathbf{E}(M_n)| = |\mathbf{E}(M_0)| = |\mathbf{E}(X_0)| \leq K,$$

et donc

$$\mathbf{E}(|A_n|) = \mathbf{E}(A_n) = \mathbf{E}(X_n - M_n) \leq 2K,$$

et A_n est bornée dans L^1, et donc également M_n par différence.

Il s'ensuit que M_n converge presque sûrement, ainsi que A_n (une suite croissante positive bornée dans L^1 est presque sûrement convergente vers une limite finie et même intégrable, par le théorème de convergence monotone). Par différence, X_n est presque sûrement convergente.

Le cas des sur-martingales est une conséquence du cas des sous-martingales, par changement de signe. ∎

Remarquons qu'une conséquence immédiate est qu'une sur-martingale positive est toujours presque sûrement convergente, (car elle est automatiquement bornée dans L^1), ainsi qu'une sur-martingale minorée (par une constante), ou qu'une sous-martingale majorée.

Par le lemme de Fatou, il est clair que la limite d'une martingale bornée dans L^1 est une variable intégrable M_∞. Il n'est pas sûr par contre que $M_n = \mathbf{E}(M_\infty/\mathcal{F}_n)$. C'est le cas des martingales uniformément intégrables, comme le montre la proposition suivante. (Nous renvoyons le lecteur à la section 5 pour les propriétés élémentaires des familles uniformément intégrables.)

Proposition 8.5. *Soit M_n une martingale bornée dans L^1, et soit M_∞ la limite de M_n lorsque $n \to \infty$. Les propositions suivantes sont équivalentes*

1. *M_n converge dans L^1 vers M_∞.*
2. *M_n est uniformément intégrable.*
3. *$M_n = \mathbf{E}[M_\infty/\mathcal{F}_n]$.*
4. *Il existe une variable intégrable M telle que $M_n = \mathbf{E}[M/\mathcal{F}_n]$. De plus, dans ce cas, $M_\infty = \mathbf{E}[M/\mathcal{F}_\infty]$.*

(Rappelons que $\mathcal{F}_\infty = \vee_n \mathcal{F}_n$.)

Démonstration. — Nous savons que, pour une suite de variables intégrables qui converge presque sûrement, il est équivalent de converger dans L^1 ou d'être uniformément intégrable. De plus, pour toute variable aléatoire M intégrable, l'ensemble des $\mathbf{E}(M/\mathcal{B})$, où $\mathcal{B}$ parcourt les sous tribus de $\mathcal{A}$ forme une famille uniformément intégrable. Il ne reste donc à montrer que les points suivants :

1. Si (M_n) est uniformément intégrable, alors $M_n = \mathbf{E}(M_\infty/\mathcal{F}_n)$;
2. Si M est intégrable, alors la martingale $M_n = \mathbf{E}(M/\mathcal{F}_n)$ converge vers $\mathbf{E}(M/\mathcal{F}_\infty)$.

Commençons par le premier point : nous écrivons, pour $p \geq n$ $M_n = \mathbf{E}(M_p/\mathcal{F}_n)$, puis nous passons à la limite en p : l'espérance conditionnelle étant continue dans L^1, et puisque M_p converge vers M_∞ dans L^1 par hypothèse, nous obtenons le résultat.

Pour le second point, remarquons d'abord que M_∞ est $\mathcal{F}_\infty$ mesurable par construction. Il suffit donc de montrer que, pour tout $A \in \mathcal{F}_\infty$, nous avons $\mathbf{E}(M_\infty \mathbf{1}_A) = \mathbf{E}(M\mathbf{1}_A)$. Or, ceci est vrai lorsque A est dans l'une des sous-tribus $\mathcal{F}_n$, puisqu'alors

$$\mathbf{E}(M\mathbf{1}_A) = \mathbf{E}(M_n\mathbf{1}_A) = \mathbf{E}(M_\infty\mathbf{1}_A).$$

Alors, l'identité cherchée, qui est vraie pour tous les éléments de l'algèbre $\cup_n \mathcal{F}_n$, reste vraie pour tous les éléments de la σ -algèbre engendrée par cette algèbre, qui est $\mathcal{F}_\infty$. (Argument de classes monotones.) ∎

Remarques

1. Un énoncé analogue à la proposition 8.5 est vrai pour les sous- et sur-martingales ; nous en laissons le soin au lecteur.
2. Si la martingale est bornée dans L^p pour un $p > 1$, alors elle est dominée par une variable de L^p et elle converge alors dans L^p.

Un des intérêts principaux des martingales uniformément intégrables est que les résultats énoncés auparavant pour les temps d'arrêt bornés ou finis presque sûrement s'étendent aux temps d'arrêt quelconques. C'est en particulier vrai pour le plus important d'entre eux, le théorème d'arrêt :

Théorème 8.6. *(***Théorème d'arrêt,** *temps d'arrêt* **quelconques.***) Soit M_n une martingale uniformément intégrable et soit T un temps d'arrêt quelconque ; alors, en définissant $M_T = M_\infty$ sur $\{T = \infty\}$, on a*

1. $M_T = \mathbf{E}(M_\infty / \mathcal{F}_T)$.
2. *L'ensemble des variables (M_T), où T est un temps d'arrêt quelconque, est uniformément intégrable.*
3. *Si S et T sont des temps d'arrêt quelconques, on a*
$$\mathbf{E}(M_T / \mathcal{F}_S) = M_{S \wedge T}.$$
4. *Pour une variable $\mathcal{A}$-mesurable M intégrable, en posant $M_n = \mathbf{E}(M/\mathcal{F}_n)$, on a $M_T = \mathbf{E}(M/\mathcal{F}_T)$.*

Démonstration. — Pour le premier point, il suffit de réécrire la démonstration du théorème d'arrêt dans ce cas. Si $\mathcal{A}$ est dans $\mathcal{F}_T$, alors

$$\begin{aligned} \mathbf{E}(M_T \mathbf{1}_A) &= \sum_{k \in \mathbb{N} \bigcup \infty} \mathbf{E}(M_k \mathbf{1}_{A \bigcap \{T=k\}}) \\ &= \sum_{k \in \mathbb{N} \bigcup \infty} \mathbf{E}(M_\infty \mathbf{1}_{A \bigcap \{T=k\}}) = \mathbf{E}(M_\infty \mathbf{1}_A). \end{aligned}$$

Ensuite, la famille des M_T est une sous-famille de la famille uniformément intégrable des $\mathbf{E}(M_\infty / \mathcal{B})$, où $\mathcal{B}$ parcourt les sous-tribus de $\mathcal{A}$.

La martingale arrêtée M^T est donc uniformément intégrable. En lui appliquant le théorème d'arrêt au temps S, on obtient

$$\mathbf{E}(M_T/\mathcal{F}_S) = M_{S\wedge T}.$$

Enfin, il suffit d'écrire

$$\mathbf{E}(M/\mathcal{F}_T) = \mathbf{E}(\mathbf{E}(M/\mathcal{F}_\infty)/\mathcal{F}_T) = \mathbf{E}(M_\infty/\mathcal{F}_T) = M_T.$$

∎

Enfin, la convergence dans L^2 des martingales bornées dans L^2 est beaucoup plus simple à obtenir, comme nous l'avons déjà remarqué, et n'exige pas l'inégalité de DOOB. Bien que ce résultat soit inclus dans ceux qui précède, cela vaut sans doute la peine de l'énoncer séparément.

Théorème 8.7. *Soit* $(X_n, n \in \mathbb{N})$ *une sous martingale positive bornée dans* L^2*. Alors la suite* $(X_n, n \in \mathbb{N})$ *est dominée dans* L^2*, et converge p.s. et dans* L^2*.*

Remarques

1. Il est important de remarquer que la condition d'intégrabilité uniforme est bien nécessaire pour obtenir le théorème d'arrêt. En effet, reprenons l'exemple de la ruine du joueur $M_n = 1+\sum_1^n X_p$, avec des variables X_p de BERNOULLI, indépendantes et centrées. Appelons $T = \inf\{n| \ X_n = 0\}$. Il n'est pas difficile de voir que T est fini presque sûrement (exercice !). Alors, la martingale arrêtée M^T converge presque sûrement vers 0, et donc $M_\infty = 0$, alors que $M_0 = 1$.
   Or, elle est bien bornée dans L^1 puisque $\mathbf{E}(|M_n^T|) = \mathbf{E}(|M_{T\wedge n}|) = \mathbf{E}(M_{T\wedge n})$ (puisque $M_{T\wedge n} \geq 0) = \mathbf{E}(M_0) = 1$.
2. Par contre pour appliquer le théorème d'arrêt (aux martingales et aux sous-martingales), il suffit que la suite $(X_{n\wedge T})$ soit uniformément intégrable. Nous l'avons déjà vu dans la proposition 6.1 pour les *temps d'arrêt* finis, mais cela se prolonge immédiatement aux *temps d'arrêt* quelconques. En effet, on peut dans ce cas appliquer le théorème d'arrêt 8.6. à $(X_{n\wedge T})$ et passer ensuite à la limite en n.
3. La décomposition de KRICKEBERG des martingales uniformément intégrables est immédiate : on décompose M_∞ en ses parties positives et négatives M_∞^+ et M_∞^-, et on écrit

$$M_n = \mathbf{E}(M_\infty^+/\mathcal{F}_n) - \mathbf{E}(M_\infty^-/\mathcal{F}_n).$$

Enfin, notons un corollaire parfois utile du théorème de convergence.

Corollaire 8.8. *(***Loi du 0-1 conditionnelle.***) Soit A un élément $\mathcal{F}_\infty$ mesurable. Si* $\limsup_n \mathbf{P}(A/\mathcal{F}_n) > 0$*, alors* $\mathbf{P}(A) = 1$.

Démonstration. — La suite $M_n = \mathbf{P}(A/\mathcal{F}_n)$ converge presque sûrement vers $\mathbf{1}_A$. Si la limite supérieure de cette suite est presque sûrement positive, alors $\mathbf{1}_A > 0$, p.s. Donc $\mathbf{1}_A = 1$ (p.s.) et par suite $\mathbf{P}(A) = 1$.

Dans la situation classique de la loi du $0-1$, nous disposons d'une suite $\mathcal{A}_n$ de sous-tribus indépendantes. On note $\mathcal{B}_n = \sigma(\mathcal{A}_k,\ k \geq n)$ et $\mathcal{B}_\infty = \cap_n \mathcal{B}_n$. Un événement asymptotique est un élément de $\mathcal{B}_\infty$. Si l'on pose $\mathcal{F}_n = \sigma(\mathcal{A}_0, \ldots, \mathcal{A}_n)$, alors $\mathcal{B}_\infty \subset \mathcal{F}_\infty$, et $\mathcal{B}_\infty$ est indépendante de $\mathcal{F}_n$ pour tout n. Si $A \in \mathcal{B}_\infty$, alors $\mathbf{P}(A/\mathcal{F}_n) = \mathbf{P}(A)$, et on voit donc que $\mathbf{P}(A) = 0$ ou bien $\mathbf{P}(A) = 1$. L'intérêt de cette extension ici est qu'on n'a pas besoin de se limiter aux événements asymptotiques. ∎

9 Conditionnement

Tous les résultats précédents sur les martingales et sous-martingales admettent une version conditionnelle qui est parfois fort utile. Le principe est le suivant :

Proposition 9.1. *Soit (X_n) une sous-martingale uniformément intégrable, et T un temps d'arrêt. Soit $A \in \mathcal{F}_T$ avec $\mathbf{P}(A) > 0$. Considérons la suite $Y_n = \mathbf{1}_A X_{n+T}$ et la filtration $\mathcal{G}_n = \{A \cap B,\ B \in \mathcal{F}_{T+n}\}$, et la plus petite tribu engendrée par tous les $\mathcal{G}_n$ qu'on note $\mathcal{G}_\infty$. On définit $\mathbf{P}_A(B) = \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(A)}$. Alors, sur l'espace de probabilité $(A, \mathcal{G}_\infty, \mathbf{P}_A)$, muni de la filtration $(\mathcal{G}_n)$, la suite (Y_n) est encore une sous-martingale uniformément intégrable.*

On peut bien sûr remplacer la sous-martingale (X_n) de la proposition précédente par une martingale, et la suite (Y_n) sera alors une martingale. On peut aussi enlever la condition d'être uniformément intégrable, la seule condition étant qu'on puisse appliquer le théorème d'arrêt à X_{n+T}.

Démonstration. — Il suffit d'appliquer le théorème d'arrêt à X_{n+1+T}

$$\mathbf{E}(X_{n+1+T}/\mathcal{F}_{T+n}) \leq X_{n+T}.$$

Prenons alors $B \in \mathcal{F}_{n+T}$, et appliquons la définition de l'espérance conditionnelle à $A \cap B \in \mathcal{F}_{T+n}$. Il vient

$$\mathbf{E}(\mathbf{1}_A \mathbf{1}_B X_{n+1+T}) \geq \mathbf{E}(\mathbf{1}_A \mathbf{1}_B X_{n+T}),$$

ce qui, divisé par $\mathbf{P}(A)$ s'écrit

$$\mathbf{E}_A(Y_{n+1}/\mathcal{G}_n) \geq Y_n.$$

∎

Avec cette remarque, on peut alors conditionner tous les résultats précédents. Faisons le par exemple avec l'inégalité maximale de Doob. On obtient

Proposition 9.2. *Soit (X_n) une sous-martingale uniformément intégrable et posons $X_T^* = \sup_{n \geq T} |X_n - X_T|$. Alors, si λ est une variable positive $\mathcal{F}_T$-mesurable, on a*

$$\lambda \mathbf{P}(X_T^* \geq \lambda / \mathcal{F}_T) \leq \mathbf{E}(|X_\infty - X_T| / \mathcal{F}_T), \ (p.s.).$$

Démonstration. — On veut démontrer que, pour tout $A \in \mathcal{F}_T$, on a

$$\mathbf{E}(\mathbf{1}_{X_T^* \geq \lambda} \mathbf{1}_A) \leq \mathbf{E}(\mathbf{1}_A \frac{|X_\infty - X_T|}{\lambda}).$$

On peut bien sûr supposer que $\mathbf{P}(A) \neq 0$ sinon il n'y a rien à démontrer. On peut aussi supposer que λ est minorée presque sûrement par une constante positive (sinon, remplacer λ par $\lambda + \epsilon$ et faire converger ϵ vers 0). Alors on peut appliquer ce qui précède : $Y_n = \mathbf{1}_A \frac{X_{n+T} - X_T}{\lambda}$ est une sous-martingale pour la filtration $\mathcal{G}_n$ de la proposition précédente, et en multipliant les deux membres par $\mathbf{P}(A)$ on a donc

$$\mathbf{P}(\sup_n |Y_n| \geq 1) = \mathbf{P}(A \cap \{X_T^* \geq \lambda\}) \leq \mathbf{E}(\mathbf{1}_A \frac{|X_\infty - X_T|}{\lambda}).$$

C'est exactement la proposition énoncée.

∎

On peut bien sûr utiliser ce principe sur d'autres inégalités, sur le nombre de descentes par exemple.

10 Martingales inverses

Les martingales inverses jouent un rôle important en théorie des martingales à temps continu (voir [3]). Mais elles ont aussi une utilité directe (voir plus bas la preuve de la loi forte des grands nombres). De plus, elles sont beaucoup plus simples à étudier que les martingales ordinaires. On pourra en trouver un exposé plus détaillé dans [5], pages 115-119.

Sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$, on considère une suite décroissante de tribus : $(\mathcal{F}_n, n \in \mathbb{N})$. On pose alors $\mathcal{F}_\infty = \cap_n \mathcal{F}_n$.

Définition 10.1. *Une suite de variables aléatoires* $(X_n, n \in \mathbb{N})$ *sur* $(\Omega, \mathcal{A}, \mathbf{P})$ *est une* **martingale inverse** *(resp. une sous-martingale inverse) si pour tout* n*,* X_n *est* $\mathcal{F}_n$*-mesurable, intégrable, et* $E(X_n/\mathcal{F}_{n+1}) = X_{n+1}$ *(resp.*$E(X_n/\mathcal{F}_{n+1}) \geq X_{n+1}$*.*

Ces processus ont des propriétés analogues à celles des martingales comme par exemple : une fonction convexe d'une martingale est une sous-martingale, l'espérance d'une martingale est constante, l'espérance d'une sous martingale est décroissante. Remarquons qu'une sous-martingale inverse positive est automatiquement bornée dans L^1.

En fait, il suffit de changer n en $-n$ et de changer l'ordre du temps pour retrouver les propriétés des martingales ordinaires : cette fois-ci l'ensemble des temps est $\mathbb{Z}_-$, l'ensemble des entiers négatifs.

Ainsi, nous avons une décomposition de Doob :

Proposition 10.2. *Soit* (X_n) *une sous-martingale inverse bornée dans* L^1 *; il existe un processus décroissant* (A_n)*, avec* $\lim_{n\to\infty} A_n = 0$*,* $\mathbf{E}(A_0) < \infty$*, tel que* A_n *soit* $\mathcal{F}_{n+1}$ *mesurable pour tout* n*, et tel que* $X_n - A_n$ *soit une martingale inverse. Dans ce cas, on a* $X_n - A_n = \mathbf{E}(X_0 - A_0/\mathcal{F}_n)$*.*

Démonstration. — On fait comme dans le cas des sous-martingales ordinaires, mais cette fois-ci en partant de ∞ : si l'on pose $D_n = \mathbf{E}(X_n - X_{n+1}/\mathcal{F}_{n+1})$, la propriété de sous-martingale inverse nous dit que D_n est positive, tandis que

$$\mathbf{E}(\sum_{p=0}^{n} D_p) = \mathbf{E}(X_0 - X_{n+1}).$$

La suite (X_n) étant bornée dans L^1, on a $\sum_n \mathbf{E}(D_n) < \infty$, et donc la série $\sum_n D_n$ converge, presque sûrement et dans L^1. On pose alors $A_n = \sum_{p=n}^{\infty} D_p$, et on voit immédiatement qu'elle vérifie les propriétés de l'énoncé.

∎

Si M est une martingale inverse, alors $M_n = \mathbf{E}(M_0/\mathcal{F}_n)$. Une martingale inverse est donc toujours uniformément intégrable. C'est également le cas d'une sous-martingale inverse :

Proposition 10.3. *Une sous-martingale inverse bornée dans L^1 est uniformément intégrable.*

Démonstration. —

On écrit la décomposition de Doob $X_n = M_n - A_n$, où la suite A_n est décroissante positive et bornée dans L^1. Elle est donc uniformément intégrable. Ensuite, M_n est une martingale inverse, avec $M_0 \in L^1$, et est donc uniformément intégrable. L'intégrabilité uniforme de la suite (X_n) s'ensuit.

∎

Tous les résultats obtenus sur les martingales et sous-martingales ordinaires vont s'appliquer (presque) immédiatement aux martingales et sous-martingales inverses, en retournant le temps. Plus précisément, considérons une martingale (resp. une sous-martingale) inverse (X_n) et fixons un entier N : Notons $\mathcal{G}_n = \mathcal{F}_{(N-n)\vee 0}$ et $Y_n = X_{(N-n)\vee 0}$. Alors, $\mathcal{G}_n$ est une filtration ordinaire par rapport à laquelle Y_n est une martingale (resp. une sous-martingale) ordinaire.

On peut lui appliquer alors les inégalités maximales et les inégalités sur les descentes (qui deviennent des montées dans ce cas) : on obtient les mêmes inégalités qu'avant, mais c'est la variable X_0 qui joue le rôle qui était alors dévolu à X_n.

Théorème 10.4. *Soit (X_n) une sous-martingale inverse positive ou une martingale inverse. Soit $X_N^* = \sup_{k=0}^N |X_k|$.*

Alors, pour tout $\lambda > 0$,

$$\lambda \mathbf{P}(X_N^* \geq \lambda) \leq \mathbf{E}[|X_0| \mathbf{1}_{X_N^* \geq \lambda}].$$

Théorème 10.5. *Soit (X_n) une sous-martingale inverse positive. Considérons deux réels $a < b$ et appelons $U_n^{(a,b)}$ le nombres de montées de la suite X_n à travers l'intervalle $[a, b]$. Alors*

$$(b-a)\mathbf{E}(U_n^{(a,b)}) \leq \mathbf{E}[(X_0 - b)_+].$$

Les théorèmes de convergence s'en déduisent alors immédiatement.

Théorème 10.6. *Soit X_n une martingale inverse, (resp. une sous-martingale inverse) bornée dans L^1.*

Alors X_n converge presque sûrement et dans L^1 vers $X_\infty = \mathbf{E}(X_0/\mathcal{F}_\infty)$ (resp. $X_\infty \leq \mathbf{E}(X_0/\mathcal{F}_\infty)$).

Démonstration. — Prenons le cas des martingales. Il suffit de suivre le shéma de démonstration donné pour les martingales ordinaires. La seule chose à démontrer est que

$$X_\infty = \mathbf{E}(X_0/\mathcal{F}_\infty).$$

C'est immédiat en utilisant l'intégrabilité uniforme. ∎

Remarque. — Si nous mettons ensemble les deux résultats de convergence pour les martingales ordinaires et inverses uniformément intégrables, nous obtenons la régularité de l'espérance conditionnelle par rapport à la "variable tribu" : si $\mathcal{F}_n$ est une suite croissante ou décroissante de tribus qui converge vers $\mathcal{F}_\infty$ (au sens où $\mathcal{F}_\infty = \vee_n \mathcal{F}_n$ dans le cas croissant et $\mathcal{F}_\infty = \cap_n \mathcal{F}_n$ dans le cas décroissant), alors, pour toute variable dans X de L^1, on a

$$\mathbf{E}(X/\mathcal{F}_n) \to \mathbf{E}(X/\mathcal{F}_\infty), \text{ ps et dans } L^1.$$

Exemple : L'application la plus spectaculaire de ce résultat est la loi forte des grands nombres : nous l'énonçons comme exercice :

Soit une suite (Z_n) de variables aléatoires indépendantes intégrables et de même loi. On pose $M_n = \dfrac{Z_1 + \cdots + Z_n}{n}$ et $\mathcal{F}_n = \sigma(M_n, M_{n+1}, ...)$. Alors (M_n) est une $\mathcal{F}$-martingale inverse, qui converge vers $\mathbf{E}(Z_1)$, presque sûrement et dans L^1.

Démonstration. — Indication : on montre d'abord que $\mathcal{F}_{n+1} = \sigma(M_{n+1}) \vee \sigma(Z_{n+2}, Z_{n+3}, ...)$. Puisque le couple (M_n, M_{n+1}) est indépendant de la tribu $\sigma(Z_{n+2}, Z_{n+3}, ...)$, on en déduit que $\mathbf{E}(M_n/\mathcal{F}_{n+1}) = \mathbf{E}(M_n/M_{n+1})$. On remarque ensuite que la loi de $(Z_1, \ldots, Z_{n+1})$ est invariante par permutation des indices (on dit que ces variables sont échangeables), et donc

$$\mathbf{E}(Z_1/M_{n+1}) = \mathbf{E}(Z_2/M_{n+1}) = \ldots = \mathbf{E}(Z_{n+1}/M_{n+1}),$$

puis en sommant que

$$\mathbf{E}(Z_1/M_{n+1}) = M_{n+1}.$$

On en déduit immédiatement la propriété de martingale inverse.

On voit donc que la suite (M_n) converge presque sûrement et dans $L^1(\Omega, \mathcal{A}, \mathbf{P})$ vers M_∞. On utilise ensuite la loi du $\{0,1\}$ pour montrer que $M_\infty = \mathbf{E}(Z_1)$. ∎

Remarque. — Le théorème de convergence de martingales permet de retrouver que la tribu $\vee_n \mathcal{F}_{T\wedge n}$ est égale à $\mathcal{F}_T$, aux ensembles de mesure nulle près. En effet En effet, en appelant $\mathcal{F}_T^1$ la tribu $\vee_n \mathcal{F}_{T\wedge n}$, il suffit de démontrer que, pour toute variable X intégrable, on a $\mathbf{E}(X/\mathcal{F}_T^1) = \mathbf{E}(X/\mathcal{F}_T)$.

Soit alors X_n la martingale (uniformément intégrable) $\mathbf{E}(X\mathcal{F}_n)$. On a

$$\mathbf{E}(X\mathcal{F}_{T\wedge n}) = X_{T\wedge n},$$

et $X_{T\wedge n}$ converge presque srement vers $X_T = \mathbf{E}(X/\mathcal{F}_T)$ lorsque $n \to \infty$.

Par ailleurs, le théorème de convergence des martingales appliqué à la filtration $\mathcal{G}_n = \mathcal{F}_{T\wedge n}$ nous montre que $X_{T\wedge n}$ converge vers $\mathbf{E}(X/\mathcal{F}_T^1)$.

11 Algorithme de Robbins-Monro

Exemple : le dosage.

Il s'agit de déterminer le dosage optimal d'un produit chimique pour obtenir l'effet voulu ; on sait, à cet effet, effectuer des tests.

Au dosage $z \in \mathbb{R}$, on associe l'effet $F(z,\eta)$, où η est une variable aléatoire dont la réalisation dépend du test effectué (pas du dosage). On connaît la loi de η et la fonction F.

L'effet moyen du dosage z est donné par $f(z) = E[F(z,\eta)]$; on suppose que f est strictement croissante et continue.

Soit $a \in \mathbb{R}$ l'effet désiré. On cherche à déterminer l'unique z^* tel que $f(z^*) = a$, et ceci sans avoir à calculer la fonction f, calcul qui peut s'avérer impossible dans la pratique, soit qu'on ne connaisse pas la loi de η, soit que l'espérance soit impossible à calculer.

Pour cela, on se donne une suite de variables aléatoires (η_n) indépendantes et de même loi que η, et une suite γ_n, dont on précisera les propriétés plus bas, mais la suite $\gamma_n = 1/n$ par exemple convient), et on construit une suite de variables aléatoires $(Z_n, n \in \mathbb{N})$ définie par récurrence :

$$Z_{n+1} = Z_n - \gamma_n[F(Z_n, \eta_{n+1}) - a].$$

Alors, si

$$E(F(z,\eta)^2) \leq c(1+z^2),$$

la suite Z_n converge presque sûrement vers la valeur z^*.

Le principe de cert algorithme est très général, et s'applique dans un grand nombre de situations. Cela s'appuie sur le résultat plus général suivant :

Théorème 11.1. *(de* ROBBINS-MONRO*) : Soit $h \in C(\mathbb{R},\mathbb{R})$, (γ_n) une suite de réels positifs et (ε_n) une suite de variables aléatoires de carré intégrable. On pose $\mathcal{F}_n = \sigma(\varepsilon_1, \cdots, \varepsilon_n)$, et on suppose que*

1. $E(\varepsilon_{n+1}/\mathcal{F}_n) = 0$;
2. $\sum_n \gamma_n = +\infty$; $\sum_n \gamma_n^2 < +\infty$;
3. $E(\varepsilon_{n+1}^2/\mathcal{F}_n) \leq c(1+Z_n^2)$;
4. $|h(z)|^2 \leq c(1+z^2)$;
5. *Il existe un unique* $z^* \in \mathbb{R} : h(z^*) = 0, (z-z^*)h(z) < 0, \forall z \neq z^*$.

Alors, si on pose

$$Z_{n+1} = Z_n + \gamma_n[h(Z_n) + \varepsilon_{n+1}], Z_0 \textit{ donnée constante,}$$

la suite (Z_n) converge presque sûrement vers z^*.

Pour le problème du dosage, on pose $\varepsilon_{n+1} = f(Z_n) - F(Z_n, \eta_{n+1})$, et $h(z) = a - f(z)$. Les conditions sont alors vérifiées.

Démonstration. — Nous ne donnons que des indications sur les grandes lignes : nous renvoyons au problème **??** et à son corrigé pour les détails.

Première étape : Il existe deux constantes α et β telles que, pour tout n, on ait la majoration

$$E[(Z_{n+1}-z^*)^2/\mathcal{F}_n] \leq (Z_n - z^*)^2(1+\alpha\gamma_n^2) + \gamma_n^2\beta.$$

En effet, d'une part les hypothèses montrent par récurrence que Z_n est de carré intégrable pour tout n, et d'autre part :

(11.6)
$$(Z_{n+1}-z^*)^2 = (Z_n-z^*)^2 + 2\gamma_n(Z_n-z^*)(h(Z_n)+\varepsilon_{n+1}) + \gamma_n^2(h(Z_n)+\varepsilon_{n+1})^2$$

que l'on conditionne par $\mathcal{F}_n$:

$$\begin{aligned} E[(Z_{n+1}-z^*)^2/\mathcal{F}_n] &= \\ &= (Z_n-z^*)^2 + 2\gamma_n(Z_n-z^*)h(Z_n) + \\ &\quad \gamma_n^2(h(Z_n)^2 + E[\varepsilon_{n+1}^2/\mathcal{F}_n]. \end{aligned}$$

Le deuxième terme est négatif par (iii) et le dernier se majore par (ii) et (iii) :

$$E[(Z_{n+1}-z^*)^2/\mathcal{F}_n] \leq (Z_n - z^*)^2 + \gamma_n^2 2c(1+Z_n^2).$$

Soit une majoration de la forme $(Z_n - z^*)^2(1+\alpha\gamma_n^2) + \gamma_n^2\beta$.

Deuxième étape : on pose $u_n = \Pi_{k=1}^{n-1}(1+\alpha\gamma_k^2)$, suite convergente puisque la série $\sum_n \gamma_n^2$ converge ; et $v_n = \sum_{k=1}^{n-1} \frac{\beta\gamma_k^2}{u_{k+1}}$ qui converge aussi puisque $u_{k+1} \geq 1$ et la série $\sum_n \gamma_n^2$ converge.

On définit la suite de variables aléatoires $X_n = \frac{(Z_n - z^*)^2}{u_n} - v_n$, soit

$$(Z_n - z^*)^2 = u_n(X_n + v_n).$$

On vérifie que X est une surmartingale : X_n est $\mathcal{F}_n$-mesurable par définition et intégrable puisque Z_n est de carré intégrable ; on a

$$\begin{aligned} E[X_{n+1}/\mathcal{F}_n] &= \frac{1}{u_{n+1}}(E[(Z_{n+1}-z^*)^2/\mathcal{F}_n]) - v_{n+1} \\ &\leq \frac{1}{u_{n+1}}[(Z_n - z^*)^2(1+\alpha\gamma_n^2) + \gamma_n^2\beta] - v_{n+1} \\ &= X_n. \end{aligned}$$

De plus, elle est minorée par la suite convergente $(-v_n, n \in \mathbb{N})$. C'est donc une sur-martingale bornée dans L^1 et elle converge presque sûrement.

Troisième étape : La suite $((Z_n - z^*)^2, n \in \mathbb{N})$ converge vers $(Z_\infty - z^*)^2$. Il faut montrer que cette limite est nulle. On reprend l'équation (11.6) en gardant le deuxième terme négatif que l'on passe à gauche et il vient :

$$0 \leq 2\gamma_n(z^* - Z_n)h(Z_n) \leq (Z_n - z^*)^2(1+\alpha\gamma_n^2) + \gamma_n^2\beta - E[(Z_{n+1}-z^*)^2/\mathcal{F}_n]$$

que l'on réexprime avec les suites (u_n) et (v_n) :

$$0 \leq 2\gamma_n(z^* - Z_n)h(Z_n) \leq u_n(X_n + v_n)\frac{u_{n+1}}{u_n} + \beta\gamma_n^2 - u_{n+1}(E[X_{n+1}/\mathcal{F}_n] + v_{n+1}),$$

ce qui donne

$$0 \leq 2\gamma_n(z^* - Z_n)h(Z_n) \leq u_{n+1}(X_n - E[X_{n+1}/\mathcal{F}_n]).$$

Ces termes sont tous positifs. La série des espérances $u_{n+1}(E[X_n]-E[X_{n+1}])$ est convergente, car le terme général de la série des espérances est qui est convergente puisque $E(X_n)$ est une suite décroissante et u_n est une suite convergente.

Donc la série de terme général $\gamma_n(z^* - Z_n)h(Z_n)$ est positive, la série de ses espérances converge, et elle converge donc ps et dans L^1.

La divergence de la série de terme général γ_n et l'unicité du zéro de h permettent alors de conclure. ∎

12 Compléments sur les martingales de carré intégrable.

Rappelons que, si M_n est une martingale de carré intégrable, il existe un unique processus croissant prévisible, nul en 0, tel que $M_n - A_n$ soit une martingale. Nous le noterons $\langle M \rangle_n$. Rappelons sa valeur

$$\langle M \rangle_n = \sum_{k=1}^{n} \mathbf{E}[(M_k - M_{k-1})^2 / \mathcal{F}_{k-1}].$$

Nous l'appelerons "la variation quadratique" de M.

Il est immédiat que, si T est un *temps d'arrêt*, la variation quadratique de la martingale arrêtée M^T est le processus arrêté $\langle M \rangle^T$, et, comme nous l'avons vu

$$\mathbf{E}(M_n - M_0)^2 = \mathbf{E}(\langle M \rangle_n).$$

Remarquons que cette identité est encore vraie pour $n = \infty$, par un passage à la limite (avec la convention que $\mathbf{E}[(M_\infty - M_0)^2] = \infty$ si M n'est pas bornée dans L^2.)

Cette identité s'étend immédiatement aux *temps d'arrêt*, bornés ou non.

Proposition 12.1. *Si T est un temps d'arrêt, alors*

$$\mathbf{E}[(M_T - M_0)^2] = \mathbf{E}(\langle M \rangle_T].$$

Démonstration. — C'est immédiat pour un temps borné, car on applique l'identité précédente à la martingale M^T, et à un temps n qui majore T.

Si T n'est pas borné, nous appliquons ce qui précède au *temps d'arrêt* $T \wedge n$, et on passe à la limite.

∎

On voit donc qu'il suffit de savoir que $\langle M\rangle_\infty$ est intégrable pour en déduire que la martingale est bornée dans L^2.

Dans la pratique, il est souvent possible que le processus $\langle M\rangle_n$ soit très facile à calculer : par exemple, si M_n est une somme de variables aléatoires indépendantes et centrées $\sum_0^n \varepsilon_k$, alors $\langle M\rangle_n = \sum_0^n \mathbf{E}(\varepsilon_k^2)$.

On peut améliorer ce résultat pour décider si la martingale arrêtée M^T est uniformément intégrable. Rappelons que l'inégalité de Doob nous dit que, si M est une martingale de carré intégrable nulle en 0, alors, en posant $M^* = \sup_n |M|_n$, on a

$$\mathbf{E}(M^{*2}) \leq 4\mathbf{E}(\langle M\rangle_\infty).$$

On a

Théorème 12.2. *(cf [5], p.148)*

1. *M_n converge presque sûrement vers une limite M_∞ sur l'événement $\{\langle M\rangle_\infty < \infty\}$.*
2. *Si M est une martingale de carré intégrable nulle en 0, et si $M^* = \sup_n |M_n|$, alors*

$$\mathbf{E}(M^*) \leq 3\mathbf{E}(\sqrt{\langle M\rangle_\infty}).$$

3. *En particulier, M^T est uniformément intégrable dès que $\sqrt{\langle M\rangle_T}$ est intégrable.*

Démonstration. — Commencons par traiter le cas des martingales bornées dans L^2. Il suffira ensuite d'arrêter la martingale en n, puis de passer à la limite pour obtenir le résultat. Fixons $a > 0$ et posons

$$T_a = \inf\{n|\ \langle M\rangle_{n+1} > a^2\},$$

avec comme d'habitude $\inf\{\emptyset\} = +\infty$. Nous profitons du fait que $\langle M\rangle_n$ est prévisible, ce quin fait que T_a est un temps d'arrêt. Nous avons donc

$$\{T_a = \infty\} = \{\langle M\rangle_\infty \leq a^2\}.$$

D'autre part, $\langle M\rangle_{T_a} \leq a^2$ sur $\{T_a < \infty\}$: on voit donc que $\langle M\rangle_{T_a} \leq \langle M\rangle_\infty \wedge a^2$, et que M^{T_a} est bornée dans L^2. Elle converge donc vers une limite. En faisant converger a vers l'infini, nous voyons déjà la convergence de M_n sur $\{\langle M\rangle_\infty < \infty\}$, puisqu'alors $T_a = \infty$ pour a assez grand.

Ensuite, pour démontrer l'inégalité

$$\mathbf{E}(M^*) \leq 3\mathbf{E}(\sqrt{\langle M \rangle_\infty}),$$

nous écrivons

$$\begin{aligned} \mathbf{P}(M^* > a) &\leq \mathbf{P}(T_a < \infty) + \mathbf{P}(T_a = \infty; M^* > a) \\ &\leq \mathbf{P}(T_a < \infty) + \mathbf{P}((M^{T_a})^* > a). \end{aligned}$$

On applique ensuite l'inégalité de Doob à la sous-martingale $(M^2)^{T_a}$, pour obtenir

$$\mathbf{P}(((M^{T_a})^*) > a) \leq \frac{1}{a^2}\mathbf{E}(M^2_{T_a}) = \frac{1}{a^2}\mathbf{E}(\langle M \rangle_{T_a}).$$

Mais $\mathbf{E}(\langle M \rangle_{T_a}) \leq \mathbf{E}(\langle M \rangle_\infty \wedge a^2)$;

En reportant cette majoration dans ce qui précède, on obtient

$$\mathbf{P}(M^* > a) \leq \mathbf{P}(\langle M \rangle_\infty > a^2) + \frac{1}{a^2}\mathbf{E}(\langle M \rangle_\infty \wedge a^2).$$

En intégrant cette inégalité par rapport à la mesure de Lebesgue en a, sur $[0, \infty[$, on obtient

$$\mathbf{E}(M^*_\infty) \leq 3\mathbf{E}(\sqrt{\langle M \rangle_\infty}).$$

Pour obtenir le dernier point, on applique le résultat à la martingale arrêtée M^T.

∎

Remarque. — En appliquant ces inégalités aux martingales $\mathbf{1}_A(M_{T\vee n} - M_T)$ où A est dans $\mathcal{F}_T$, et observant que leur crochet est $\mathbf{1}_A(\langle M \rangle_{T\vee n} - \langle M \rangle_T)$, nous obtenons les versions conditionnelles de ces inégalités :

$$\mathbf{E}(\sup_{n\geq T} |M_n - M_T|^2 / \mathcal{F}_T) \leq 4\mathbf{E}(\langle M \rangle_\infty - \langle M \rangle_T / \mathcal{F}_T)(p.s.)$$

et

$$\mathbf{E}(\sup_{n\geq T} |M_n - M_T| / \mathcal{F}_T) \leq 3\mathbf{E}(\sqrt{\langle M \rangle_\infty - \langle M \rangle_T} / \mathcal{F}_T)(p.s.)$$

Nous avons vu que M converge vers une valeur finie sur $\{\langle M \rangle_\infty < \infty\}$. Que se passe-t-il sur $\{\langle M \rangle_\infty = \infty\}$? La réponse est donnée par la proposition suivante

Proposition 12.3. (**Loi des grands nombres**) *Si $f : \mathbb{R}_+ \to \mathbb{R}_+$ est une fonction croissante, telle que*

$$\int_0^\infty \frac{dt}{[1+f(t)]^2} < \infty,$$

alors $M_n/f(\langle M\rangle_n)$ converge presque sûrement vers 0 lorsque $n \to \infty$ sur $\{\langle M\rangle_\infty = \infty\}$.

Remarquons que ce résultat s'applique en particulier avec $f(t) = t$, et aussi avec $f(t) = t^\alpha$ si $\alpha > 1/2$.

Démonstration. — Posons $A_n = (1 + f(\langle M\rangle_n))^{-1}$. Considérons la martingale $N_n = (A.M)_n$. Alors

$$\begin{aligned}
\langle N\rangle_n &= \sum_{k=1}^n A_k^2(\langle M\rangle_k - \langle M\rangle_{k-1}) \\
&= \sum_{k=1}^n \frac{\langle M\rangle_k - \langle M\rangle_{k-1}}{(1+f(\langle M\rangle_k))^2} \\
&\leq \sum_{k\geq 1} \int_{\langle M\rangle_{k-1}}^{\langle M\rangle_k} \frac{dt}{[1+f(t)]^2}.
\end{aligned}$$

La martingale N est donc de carré intégrable. Elle est donc presque sûrement convergente.

La démonstration sera donc achevée dès qu'on aura démontré le lemme suivant. ∎

Lemme 12.4. (**Lemme de Kronecker**) *Soit u_n une suite croissante de réels positifs qui converge vers l'infini, et δ_n une suite de réels. Alors, si la série $\sum_n \delta_n/u_n$ converge, la suite*

$$\frac{1}{u_n}\sum_0^n \delta_k$$

converge vers 0.

Démonstration. — (Exercice.)

∎

Corollaire 12.5. *Si la martingale satisfait à $\mathbf{E}[(M_{n+1} - M_n)^2/\mathcal{F}_n] \leq C$, alors $M_n/n^\alpha \to 0$, $\forall\alpha > 1/2$.*

Démonstration. — Dans ce cas, nous avons $\langle M\rangle_n \le Cn$, et on applique le résultat précédent. Sur $\langle M\rangle_\infty < \infty$, la martingale converge, et donc M_n/n converge vers 0, tandis que sur $\langle M\rangle_\infty = \infty$, on applique la proposition 12.3. ∎

Il ne faudrait pas déduire de ce qui précède que M_n diverge presque sûrement sur l'ensemble $\{\langle M\rangle_\infty = \infty\}$. C'est vrai pour les martingales à accroissements bornés (voir la proposition 12.6 ci-dessous) mais faux en général, comme le montre l'exemple ci-dessous :

Exemple. — Soit Z_n une suite de variables aléatoires indépendantes, prenant la valeur $\alpha_n > 0$ avec probabilité p_n $(0 < p_n < 1)$ et $\beta_n = -\alpha_n \frac{p_n}{1-p_n}$ avec probabilité $1 - p_n$. On pose $M_n = \sum_{i=0}^n Z_i$. Alors M_n est une martingale.

On suppose que $\sum_n p_n < \infty$, que $\sum_n \alpha_n p_n < \infty$, et que $\sum_n \alpha_n^2 p_n = \infty$. Alors, M_n converge presque sûrement, tandis que $\langle M\rangle_\infty = \infty$.

Pour s'en convaincre, on remarque que le lemme de Borel-Cantelli nous dit que $Z_n = \beta_n$ à partir d'un certain rang. Puisque $\beta_n \simeq -\alpha_n p_n$, on en déduit que la série $\sum_n Z_n$ est convergente. Par ailleurs, on voit que $\langle M\rangle_n - \langle M\rangle_{n-1} \simeq \alpha_n^2 p_n$, et donc $\langle M\rangle_\infty = \infty$ par divergence de la troisième série.

Remarquons que ces hypothèses sont par exemple vérifiées avec $p_n = n^{-2}$ et $\alpha_n = \sqrt{n}$, pour $n \ge 1$.

Proposition 12.6. *Soit M_n une martingale à accroissements bornés par une constante K. Alors,* $\limsup_n |M_n| = \infty$ *sur l'ensemble* $\{\langle M\rangle_\infty = \infty\}$.

Démonstration. — On peut bien sûr supposer que $M_0 = 0$. Soit K un majorant de $|\Delta M|_n$. Fixons une constante N et considérons le *temps d'arrêt*

$$T_N = \inf\{n \mid |M_n| \ge N\}.$$

Alors, $\left|M^{T_N}\right| \le N + K$, puisque les accroissements de M sont bornés par K, et que $\left|M_n^{T_M}\right| \le M$ tant que $n \le T_M - 1$. On en déduit que M^{T_N} est presque sûrement convergente, et que

$$\mathbf{E}(\langle M\rangle_{T_N}) = \mathbf{E}((M_\infty^{T_N})^2) \le (K+N)^2.$$

Donc, aux ensemble de mesure nulle près,

$$\{T_N = \infty\} \subset \{\langle M\rangle_\infty < \infty\}.$$

D'où l'on tire

$$\{\langle M\rangle_\infty = \infty\} \subset \bigcap_{N\in\mathbb{N}} \{T_N < \infty\} = \{\limsup_n |M|_n = \infty\}.$$

D'où le résultat. ∎

La loi des grands nombres du théorème 12.3 donne des résultats assez frappants. Ainsi le résultat suivant

Corollaire 12.7. *(***Lemme de Borel Cantelli conditionnel.***) Soit (A_n) une suite d'événements $\mathcal{F}_n$ mesurables. Alors*

$$\limsup_n A_n = \{\omega \ | \ \sum_n \mathbf{P}(A_n/\mathcal{F}_{n-1})(\omega) = \infty\}.$$

Démonstration. — Rappelons que la limite sup des événements A_n est par définition l'ensemble des ω qui appartiennent à une infinité de A_n. C'est donc aussi l'ensemble où la série $\sum_n \mathbf{1}_{A_n}$ diverge.

Soit alors $X_n = \sum_{i=1}^n \mathbf{1}_{A_i}$ (avec $X_0 = 0$)et $B_n = \sum_{i=1}^n \mathbf{P}(A_i/\mathcal{F}_{i-1})$. Par construction, $M_n = X_n - B_n$ est une martingale, évidemment de carré intégrable. En posant $U_n = \mathbf{P}(A_n/\mathcal{F}_{n-1})$, on a

$$\langle M\rangle_n = \sum_{i=1}^n \mathbf{P}(\mathbf{1}_{A_i}^2/\mathcal{F}_{i-1}) - \mathbf{P}(\mathbf{1}_{A_i}/\mathcal{F}_{i-1})^2 = \sum_{i=1}^n U_i(1-U_i) \leq B_n.$$

Sur l'ensemble où B_n converge, alors $\langle M\rangle_\infty < \infty$, et donc M_n converge. Il en va de même par conséquent de X_n, puisque B_n converge aussi.

Sur l'ensemble où $\langle M\rangle_\infty = \infty$ diverge, alors B_n diverge et $M_n/\langle M\rangle_n^{3/4} \to 0$. Par conséquent

$$|M_n| \leq K\langle M\rangle_n^{3/4} \leq KB_n^{3/4},$$

pour un certain K aléatoire et n assez grand. Alors, $X_n = M_n + B_n$ converge aussi vers l'infini.

Reste à traiter l'ensemble $\{B_\infty = \infty\} \cap \{\langle M\rangle_\infty < \infty\}$. Sur cet ensemble, M_n converge et B_n diverge, donc X_n diverge.

Remarquons par ailleurs que si $\sum_n \mathbf{P}(A_n) < \infty$, alors $\sum_n \mathbf{P}(A_n/\mathcal{F}_{n-1})$ est intégrable, et donc fini presque sûrement. ∎

Remarque. — La démonstration du corollaire 12.7 nous montre un peu mieux. Si $X_n = X_0 + \sum_{p=1}^n Z_p$, avec $Z_p \geq 0$, adaptée, si $A_n = \sum_{p=1}^n \mathbf{E}(Z_p/\mathcal{F}_{p-1})$ et si $B_n = \sum_{p=1}^n \sigma^2(Z_p/\mathcal{F}_{p-1})$, alors les ensembles où X_n et A_n convergent vers une limite finie sont identiques (aux ensembles de mesure nulle près) dès que $B_n \leq KA_n^\alpha$, pour un $\alpha < 2$.

Enfin, un corollaire qui est utile dans l'étude des chaînes de Markov dénombrable (voir par exemple le problème **??**).

Corollaire 12.8. *Soit X_n une suite de variables aléatoires intégrables, adaptées à la filtration $\mathcal{F}_n$. On pose $\Delta_n = X_{n+1} - X_n$ et on suppose qu'il existe une constant $K \in \mathbb{R}_+$ et une constante $\alpha > 0$ telles que, pour tout n, on ait*

$$|\Delta_n| \le K, \ \mathbf{E}(\Delta_n/\mathcal{F}_n) \ge \alpha.$$

Alors, $\liminf_n \frac{X_n}{n} \ge \alpha$, *(p.s.).*

Démonstration. — On peut bien sûr supposer que $X_0 = 0$. On pose alors $\alpha_n = \mathbf{E}(\Delta_n/\mathcal{F}_n)$ et $A_n = \sum_{k=0}^{n-1} \alpha_k$. Par hypothèse, nous savons que $A_n \ge \alpha n$, et par construction, $M_n = X_n - A_n$ est une martingale. Puisque $|\Delta_n| \le K$, alors $0 < \alpha \le \alpha_n \le K$, et $|M_{n+1} - M_n| \le 2K$.

Alors, $\langle M \rangle_n \le 4K^2 n$, et par suite $\frac{M_n}{n}$ converge vers 0 presque sûrement. Puisque $X_n = M_n + A_n$ et que A_n est minorée par αn, on obtient le résultat.

∎

Pour les martingales à accroissements bornés, nous énonçons sans démonstration un résultat plus fin sur le comportement asymptotique de la suite M_n, et nous renvoyons à [5], p. 153, pour plus de détails.

Proposition 12.9. *(***Loi du logarithme itéré***) Si M_n est une martingale de carré intégrable telle que $\| \sup_n |M_{n+1} - M_n| \|_\infty < \infty$, alors, presque sûrement sur l'événement $\{\langle M \rangle_n = \infty\}$,*

$$\begin{aligned} \limsup_n \frac{M_n}{\sqrt{2\langle M \rangle_n \log\log \langle M \rangle_n}} &= 1 \\ \liminf_n \frac{M_n}{\sqrt{2\langle M \rangle_n \log\log \langle M \rangle_n}} &= -1 \end{aligned}$$

Enfin, la proposition 12.3 sert de loi des grands nombres pour les martingales. Le résultat suivant sert quand à lui de théorème de la limite centrale. Nous ne l'énonçons pas dans toute sa généralité mais seulement dans la forme qui nous servira plus tard pour les chaînes de Markov.

Théorème 12.10. *(***Théorème de la limite centrale***) Soit M_n une martingale de carré intégrable. Pour $n \ge 0$, posons $\Delta_n = M_{n+1} - M_n$, et $\Sigma_n = \mathbf{E}\big(\Delta_n^2/\mathcal{F}_n\big)$. Nous supposons qu'il existe une constante K telle que, pour tout n, $\mathbf{E}(|\Delta_n|^3/\mathcal{F}_n) \le K$.*

Si la suite $(\langle M \rangle_n/n)$ converge en probabilité vers une constante σ^2, alors $M_n/\sqrt{n}$ converge en loi vers une variable gaussienne $N(0, \sigma^2)$.

Remarquons que la même propriété s'en déduit immédiatement s'il y a convergence en loi de $\langle M\rangle_n/n$ vers σ^2, puisque la convergence en loi vers une constante implique la convergence en probabilité.

Démonstration. — Si U_n converge en loi vers une mesure de probabilité μ, et si V_n converge en probabilité vers 0, alors U_n+V_n converge en loi vers la même limite. On peut donc se ramener au cas où $M_0=0$.

Ensuite, quitte à remplacer M_n par $M_n/K^{3/2}$, on peut se ramener au cas où $K=1$, et dans ce cas $\Sigma_n\leq 1$, par l'inégalité de Jensen conditionnelle.

En appliquant le théorème de P. Lévy, il nous suffit alors de démontrer que, pour tout réel t,

$$\lim_{n\to\infty}\mathbf{E}\big(\exp(it\frac{M_n}{\sqrt{n}}+\frac{t^2}{2}\sigma^2)\big)=1.$$

Pour cela, nous allons montrer qu'il existe une constante universelle K_1 telle que, pour tout $t\in[-1,1]$, on ait

$$|\mathbf{E}\big(\exp(itM_n+\frac{t^2}{2}\langle M\rangle_n)\big)-1|\leq nK_1t^3\exp(nt^2). \tag{12.7}$$

Pour le voir, posons

$$\mathbf{E}\big(\exp(itM_n+\frac{t^2}{2}\langle M\rangle_n)\big)=1+v_n,$$

et remarquons tout d'abord qu'il existe une constante universelle K_2 telle que, pour tout réel $y\in[0,1]$, tout réel $t\in[-1,1]$,

$$\exp(itx+t^2y)=1+itx-t^2\frac{x^2}{2}+t^2y+R(t,x,y),$$

avec $|R(t,x,y)|\leq K_2(1+|x|^3)t^3\exp(t^2)$: écrire la formule de Taylor avec reste intégral pour la fonction $f(t)=\exp(itx+t^2y)$:

$$f(t)=1+tx+\frac{t^2}{2}(y-x^2)+\int_0^t f^{(3)}(s)(t-s)^2\,ds,$$

et de remarquer que la dérivée troisième de f, qui est égale à

$$\big((2ty+ix)^2+2y)\big)(ix+2ty)f$$

est majorée par $K_2(1+|x|^3)$ sur le domaine considéré.

Ensuite, pour démontrer la majoration 12.7, nous posons

$$Y_n = \exp(itM_n + \frac{t^2}{2}\langle M\rangle_n),$$

et écrivons $Y_{n+1} = Y_n Z_n$, avec

$$Z_n = \exp(it\Delta_n + \frac{t^2}{2}\Sigma_n).$$

Commençons par calculer $\mathbf{E}(Y_{n+1}/\mathcal{F}_n)$. La variable Y_n étant mesurable par rapport à $\mathcal{F}_n$, cette dernière quantité est égale à $Y_n\mathbf{E}(Z_n/\mathcal{F}_n)$. Utilisons alors le développement limité précédent, avec $x = \Delta_n$, $y = \Sigma_n = \mathbf{E}(\Delta_n^2/\mathcal{F}_n)$, développement justifié puisque $|\Sigma_n| \leq 1$. On obtient ainsi

$$\begin{aligned}\mathbf{E}(Z_n/\mathcal{F}_n) = \quad & 1 + it\mathbf{E}(\Delta_n/\mathcal{F}_n) - \frac{t^2}{2}\mathbf{E}(\Delta_n^2/\mathcal{F}_n) + \frac{t^2}{2}\mathbf{E}(\Sigma_n/\mathcal{F}_n) \\ & + \mathbf{E}\big(R(t,\Delta_n,\Sigma_n)/\mathcal{F}_n\big).\end{aligned}$$

Tout est fait pour que ce dévelopement se résume à

$$\mathbf{E}(Z_n/\mathcal{F}_n) = 1 + \mathbf{E}\big(R(t,\Delta_n,\Sigma_n) \mid \mathcal{F}_n\big) = 1 + V_n,$$

où la variable aléatoire $|V_n|$ est majorée par

$$K_2 t^3 \exp(t^2)\mathbf{E}(1 + |\Delta_n|^3/\mathcal{F}_n) \leq 2K_2 t^3 \exp(t^2).$$

Posons alors $K_1 = 2K_2$, et établissons la majoration cherchée sur v_n par récurrence sur n. En prenant les espérances des espérances conditionnelles, nous obtenons

$$1 + v_{n+1} = \mathbf{E}\big(Y_n(1 + V_n)\big) = 1 + v_n + \mathbf{E}(Y_n V_n).$$

Nous voyons d'autre part que $|Y_n| \leq \exp(nt^2/2)$, et donc, si par hypothèse $|v_n| \leq nK_1 t^3 \exp(nt^2)$, nous obtenons

$$|v_{n+1}| \leq nK_1 t^3 \exp(nt^2) + K_1 t^3 \exp(t^2)\exp(n\frac{t^2}{2}) \leq (n+1)K_1 t^3 \exp\big(((n+1)t^2\big).$$

Cette majoration étant établie, il ne reste plus qu'à changer t et $t/\sqrt{n}$ pour obtenir

$$\lim_{n\to\infty} \mathbf{E}\Big(\exp\big(it\frac{M_n}{\sqrt{n}} + t^2\frac{\langle M\rangle_n}{2n}\big)\Big) = 1.$$

Ensuite, nous écrivons

$$\begin{aligned} & \mathbf{E}\big(|\exp(it\frac{M_n}{\sqrt{n}} + t^2\frac{\sigma^2}{2}) - \exp(it\frac{M_n}{\sqrt{n}} + t^2\frac{\langle M\rangle_n)}{2n}|\big) \\ = \quad & \mathbf{E}\big(|\exp(t^2\frac{\sigma^2}{2}) - \exp(t^2\frac{\langle M\rangle_n}{2n})|\big), \end{aligned}$$

et cette dernière quantité converge vers 0, car les variables $\exp\left(t^2\frac{\langle M\rangle_n}{2n}\right)$ sont bornées, convergent en probabilité vers $\exp(t^2\frac{\sigma^2}{2})$. Elles y convergent donc dans L^1.

∎

13 Le théorème de Radon-Nikodym.

Nous terminons ce chapitre par une application classique du théorème de convergence des martingales au théorème de Radon-Nikodym. Nous l'énonçons dans le cadre des tribus séparables (engendrées par un nombre dénombrable d'événements), mais la démonstration présentée ici reste vraie pour des tribus plus générales, par exemple les tribus engendrées aux ensembles de mesure nulle près par un nombre dénombrable d'événements, ce qui englobe une classe infiniment plus vaste.

Rappelons qu'on dit qu'une probabilité $\mathbf{Q}$ admet une densité par rapport à $\mathbf{P}$ s'il existe une variable aléatoire X, intégrable par rapport à $\mathbf{P}$, d'intégrale 1, positive, telle que, pour tout élément A de la tribu $\mathcal{A}$, on ait

$$\mathbf{Q}(A) = \int_A X\, d\mathbf{P}.$$

On la notera alors $\mathbf{Q} = X.\mathbf{P}$. Une condition nécessaire et suffisante pour que Q ait une densité par rapport à $\mathbf{P}$ est la suivante :

Théorème 13.1. *(***De Radon-Nikodym***) Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace de probabilité, et on suppose la tribu $\mathcal{A}$ séparable. Soit $\mathbf{Q}$ une autre probabilité sur $(\Omega, \mathcal{A})$. Une CNS pour que $\mathbf{Q}$ admette une densité par rapport à $\mathbf{P}$ est que,*

$$\forall A \in \mathcal{A},\ \mathbf{P}(A) = 0 \Longrightarrow \mathbf{Q}(A) = 0.$$

Démonstration. — La condition est évidemment nécessaire, et tout le problème est de montrer qu'elle est suffisante. L'idée de la preuve est de construire la variable aléatoire X grâce au théorème de convergence des martingales.

Dans ce qui suit, nous appellerons (A_n) une suite d'éléments qui engendrent la tribu $\mathcal{A}$, et la tribu $\sigma(A_0, \cdots, A_n)$ sera dénotée $\mathcal{F}_n$. La suite $(\mathcal{F}_{n)}$ est évidement croissante, et la tribu $\mathcal{F}_n$, engendrée par une famille finie de variables aléatoires, est en fait engendrée par une partition finie $\{B_1^n, B_2^n, \cdots, B_{p_n}^n\}$. Il est facile d'exhiber une densité de $\mathbf{Q}$ par rapport à $\mathbf{P}$ sur la tribu $\mathcal{F}_n$, c'est à dire une variable aléatoire X_n positive telle que, pour tous les B de $\mathcal{F}_n$, on ait $\mathbf{Q}(B) = \int_B X_n \, d\mathbf{P}$. Il suffit de poser

$$X_n = \sum_{i=1}^{p_n} \frac{\mathbf{Q}(B_i^n)}{\mathbf{P}(B_i^n)} \mathbf{1}_{B_i^n}.$$

A cause de l'hypothèse, si dans la somme précédente l'un des numérateurs est nul, il en va de même du dénominateur, et on n'a qu'à poser alors $0/0 = 1$. On voit que X_n joue bien le rôle d'une densité car tout élément de $\mathcal{F}_n$ est une réunion finie de B_i^n.

Bien sûr, la suite (X_n) est adaptée, positive et d'espérance 1 (sous $\mathbf{P}$) puisque c'est une densité de lois de probabilité. La remarque fondamentale dans ce qui suit est la suivante : (X_n) est une martingale sous la probabilité $\mathbf{P}$. Ceci provient d'une remarque plus générale : si $\mathcal{B}_1 \subset \mathcal{B}_2$ sont deux sous-tribus de $\mathcal{A}$, et si X_1 et X_2 sont deux variables qui jouent le rôle des densités de $\mathbf{Q}$ par rapport à $\mathbf{P}$ respectivement sur $\mathcal{B}_1$ et $\mathcal{B}_2$, alors

$$\mathbf{E}_{\mathbf{P}}(X_2 \mid \mathcal{B}_1) = X_1,$$

où $\mathbf{E}_{\mathbf{P}}$ désigne ici l'espérance (conditionnelle) par rapport à $\mathbf{P}$. En effet, on voit que X_1 est $\mathcal{B}_1$ mesurable par construction, et, pour tout B de $\mathcal{B}_1$, on a

$$\mathbf{E}_{\mathbf{P}}(\mathbf{1}_B X_1) = \mathbf{Q}(B) = \mathbf{E}_{\mathbf{P}}(\mathbf{1}_B X_2),$$

car $B \in \mathcal{B}_2$.

En appliquant ceci à la tribu $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, on voit que (X_n) est une martingale sous $\mathbf{P}$. Elle est positive d'espérance 1, donc converge presque sûrement pour $\mathbf{P}$. Nous allons nous attacher à montrer qu'elle est uniformément intégrable, et ceci grâce à l'hypothèse. En effet, supposons qu'elle ne le soit pas ; alors, il existe $\varepsilon > 0$ et une sous-suite n_k tendant vers l'infini telle que

$$\mathbf{E}_{\mathbf{P}}(X_{n_k} \mathbf{1}_{\{X_{n_k} \geq k\}}) \geq \varepsilon.$$

Ceci s'écrit encore $\mathbf{Q}(X_{n_k} \geq k) \geq \varepsilon$, d'où encore

$$\mathbf{Q}(\limsup_k(\{X_{n_k} \geq k\}) \geq \limsup \mathbf{Q}(X_{n_k} \geq k) \geq \varepsilon.$$

Si l'on appelle B l'événement $\{\limsup_n X_n = \infty\}$, nous voyons donc que $\mathbf{Q}(B) \geq \varepsilon$, alors que $\mathbf{P}(B) = 0$, d'où la contradiction.

La suite (X_n) étant sous $\mathbf{P}$ uniformément intégrable, elle converge sous $\mathbf{P}$ et dans L^1 vers une variable aléatoire X positive intégrable et d'intégrale 1. Il nous reste à monter que X est la densité cherchée de $\mathbf{Q}$ par rapport à $\mathbf{P}$. Pour ce faire, nous commençons à remarquer que, si B est dans l'une des tribus $\mathcal{F}_n$, alors

$$\mathbf{Q}(B) = \int_B X_n \, d\mathbf{P} = \int_B X \, d\mathbf{P},$$

car $X_n = \mathbf{E}_{\mathbf{P}}(X/\mathcal{F}_n)$. Les probabilités $\mathbf{Q}$ et $X.\mathbf{P}$ coïncident alors sur l'algèbre $\cup_n \mathcal{F}_n$, et donc sur la σ-algèbre $\mathcal{A}$ engendrée. C'est ce que nous voulions démontrer. ∎

Bibliographie

[1] P. Barbe and M. Ledoux. *Probabilité, De la licence à l'agrégation.* Espace 34, Belin, Montpellier, 1998.

[2] C. Dellacherie and P. A. Meyer. *Probabilités et potentiel.* Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[3] C. Dellacherie and P.A. Meyer. *Probabilités et potentiel. Chapitres V à VIII*, volume 1385 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics].* Hermann, Paris, revised edition, 1980. Théorie des martingales. [Martingale theory].

[4] G. Fayolle, V. A. Malyshev, and M. V. Menshikov. *Topics in the constructive theory of countable Markov chains.* Cambridge University Press, 1995.

[5] J. Neveu. *Martingales à temps discret.* Masson, Paris, 1972.

[6] P. Toulouse. *Thèmes de Probabilités et Statistique, Agrégation de mathématiques.* Dunod, Paris, 1999.