Université de Strasbourg. Master M1 de Mathématiques.

Actuariat

Cours de Calcul Stochastique

J. FRANCHI

Deuxième semestre 2013-2014. 36 heures de cours intégrés.

CONTENU

0. Fondements du calcul des probabilités page 2

1) Probabilité ; conditionnement page 2
2) Variables aléatoires et leur lois page 9
3) Lois usuelles page 13
4) Variables aléatoires indépendantes page 17
5) Convergences des variables aléatoires page 19

I. Espérance conditionnelle page 23

6) Espérance conditionnelle discrète page 23
7) Autres propriétés de l'espérance conditionnelle page 28
8) Projection orthogonale dans L^2 page 29
9) Extension à L^1 page 33

II. Martingales à temps discret page 33

10) Surmartingales, sousmartingales page 34
11) Temps et théorème d'arrêt page 36
12) Inégalités et Convergence page 38
13) Décomposition de Doob et variations quadratiques page 39
14) Propriétés de l'intégrale stochastique discrète page 40
15) Changement de probabilité sur Ω fini page 43
16) Application aux marchés : arbitrage page 46
17) Modèle de Cox-Ross-Rubinstein page 48
18) Arrêt optimal page 51

III. Mouvement Brownien page 53
19) Mouvement Brownien réel page 53
20) Propriétés trajectorielles du mouvement brownien page 57
21) Martingales à temps continu page 58
IV. Intégrale stochastique page 61
22) Intégrale de Stieltjes page 61
23) Intégrale d'Itô page 62
24) Formule d'Itô page 66
25) Théorème de Girsanov page 69
26) Équations différentielles stochastiques (É.D.S.) page 71
27) Modèle de Black-Scholes page 74
Bibliographie page 76

0. Fondements de la théorie des probabilités

1 Probabilité ; conditionnement

Une probabilité est d'abord une *fonction* qui à un *événement* associe un nombre réel compris entre 0 et 1. Cela implique de préciser ce qu'est un événement. Or cela n'a de sens que dans le cadre d'un ensemble d'*épreuves aléatoires* ou *tirages*, qu'on note généralement Ω. Il peut s'agir par exemple de lancers de dés ou de pièces de monnaie, de tirages d'urne, de durées de vie (d'atomes ou d'individus), de tirs sur une cible, etc... Ces premiers exemples familiers montrent déjà que l'ensemble Ω peut être fini, dénombrable (ce qui signifie : infini indexé par $\mathbb{N}$ ou $\mathbb{N}^*$), ou continu. Ce sera donc a priori un ensemble non vide quelconque.

Lorsque Ω est fini ou dénombrable, toutes ses parties seront des événements. Mais en général il est nécessaire de se restreindre à un sous-ensemble de parties de Ω : $\mathcal{T} \subset \mathcal{P}(\Omega)$, qu'on nomme *tribu* (ou *σ-algèbre*). On a naturellement besoin de pouvoir considérer la réunion et la conjonction ($\equiv$ intersection) de 2 événements, de même que le complémentaire ($\equiv$ contraire) d'un événement ; en outre il faut aussi pouvoir considérer une réunion dénombrable d'événements. Enfin il est naturel de considérer l'événement impossible ($\equiv$ vide : $\emptyset$) et l'événement certain Ω. D'où la définition suivante.

Définition 1.1 *Une tribu (ou σ-algèbre) est une partie $\mathcal{T}$ de $\mathcal{P}(\Omega)$ stable par réunion dénombrable et par passage au complémentaire, et contenant l'ensemble vide $\emptyset$.*

Le couple $(\Omega, \mathcal{T})$ est appelé espace probabilisable. Un événement est un élément de $\mathcal{T}$.

$\mathcal{P}(\Omega)$ désigne l'ensemble de toutes les parties de Ω ; $\mathcal{T}$ est donc un ensemble de parties de Ω. La stabilité par réunion dénombrable s'écrit formellement : pour toute suite $\{E_n \,|\, n \in \mathbb{N}\} \subset \mathcal{T}$ d'événements, leur réunion est aussi un événement : $\bigcup_{n\in\mathbb{N}} E_n \in \mathcal{T}$.

La stabilité par passage au complémentaire s'écrit formellement : le complémentaire $E^{\mathbf{c}} := \Omega \setminus E = \{\omega \in \Omega \,|\, \omega \notin E\}$ de tout événement E est aussi un événement : $E^{\mathbf{c}} \in \mathcal{T}$.

Nota Bene Sur Ω fini ou dénombrable, on choisira toujours par défaut la tribu $\mathcal{P}(\Omega)$.

Rappel : $\Big[\bigcap_n A_n\Big]^{\mathbf{c}} = \bigcup_n A_n^{\mathbf{c}}$; $\Big[\bigcup_n A_n\Big]^{\mathbf{c}} = \bigcap_n A_n^{\mathbf{c}}$. On vérifie aussitôt les propriétés suivantes :

Proposition 1.2 *Ω est un événement (certain). La réunion et l'intersection d'un nombre fini d'événements sont des événements. La différence $A \setminus B := A \cap B^c$ et la différence symétrique $A\Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$ de deux événements A et B sont des événements. Toute intersection dénombrable d'événements est un événement.*

Nous pouvons maintenant définir rigoureusement ce qu'est une probabilité ; dans la pratique, ce sera le plus souvent soit une somme (soit finie, soit infinie, c'est-à-dire une série) ou une intégrale (soit propre, dite de Riemann, soit impropre). (Pour rassembler toutes ces possibilités, et d'autres encore, on parle de *mesure*.)

Définition 1.3 *Une probabilité sur l'espace probabilisable $(\Omega, \mathcal{T})$ est une fonction $\mathbb{P}$ de $\mathcal{T}$ dans $[0, 1]$ qui vérifie : $\mathbb{P}(\Omega) = 1$, et la propriété d'additivité dénombrable :*

$\mathbb{P}\Big(\bigsqcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ pour tout suite $\{A_n \,|n \in \mathbb{N}\}$ d'événements deux à deux disjoints. Le triplet $(\Omega, \mathcal{T}, \mathbb{P})$ est appelé espace probabilisé ou espace de probabilité. Les événements de probabilité nulle sont dits négligeables. Les événements de probabilité 1 sont dits presque sûrs.

C'est toujours dans le cadre d'un espace de probabilité, plus ou moins bien précisé, que peut avoir lieu un calcul de probabilité. Il est généralement préférable de bien le préciser. En effet c'est la non-précision de l'espace considéré qui est à l'origine de paradoxes ou d'erreurs courantes sur les probabilités.

On vérifie aisément les propriétés suivantes :

Proposition 1.4 *(i) L'événement impossible est négligeable : $\mathbb{P}(\emptyset) = 0$.*

(ii) $\mathbb{P}$ est croissante : $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, pour tous événements A et B.

(iii) $\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ pour tout suite $\{A_n \,|n \in \mathbb{N}\}$ d'événements.

(iv) $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$, lorsque A et B sont deux événements tels que $B \subset A$. En particulier, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ pour tout événement A.

(v) Toute intersection dénombrable d'événements presque sûrs est presque sûre.

(vi) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, pour tous événements A et B.

(vii) $\mathbb{P}$ *est continue :* $\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n)$, *pour toute suite* <u>*croissante*</u> *d'événements* $\{A_n \,|\, n\in\mathbb{N}\}$ *(id est* $A_n \subset A_{n+1}$ $(\forall\, n)$*) ; et de même* $\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} B_n\Big) = \lim_{n\to\infty} \mathbb{P}(B_n)$, *pour toute suite* <u>*décroissante*</u> *d'événements* $\{B_n \,|\, n\in\mathbb{N}\}$ *(id est* $B_n \supset B_{n+1}$ $(\forall\, n)$*).*

1.5 Exemples

1. Probabilité discrète sur un ensemble $\Omega = \{\omega_1, .., \omega_N\}$ fini :
elle est clairement définie par la liste des probabilités des singletons : $p_j := \mathbb{P}(\{\omega_j\})$.
Nous avons en effet $\mathbb{P}(A) = \sum_{\omega_j\in A} p_j$ pour toute partie $A \subset \Omega$.

Réciproquement, toute liste $\{p_1, .., p_N\}$ de réels $p_j \geq 0$ tels que $\sum_{j=1}^{N} p_j = 1$ définit bien (par la même formule) une probabilité unique sur Ω.

Exemples concrets : lancers de dés, de pièces de monnaie, tirages de cartes à jouer, tirages de boules dans des urnes, loteries, etc...

2. Probabilité discrète sur $\mathbb{N}$ (ou sur n'importe quel autre ensemble dénombrable) :
elle est encore définie par la liste des probabilités des singletons : $p_j := \mathbb{P}(\{\omega_j\})$.
Nous avons en effet $\mathbb{P}(A) = \sum_{\omega_j\in A} p_j$ pour toute partie $A \subset \mathbb{N}$. La seule différence avec le cas précédent est que la somme peut être une série ($\equiv$ comporter une infinité de termes).
Réciproquement, toute suite $\{p_j \,|\, j\in\mathbb{N}\}$ de réels $p_j \geq 0$ tels que $\sum_{j\geq 1} p_j = 1$ définit bien (par la même formule, forcément convergente) une probabilité unique sur $\mathbb{N}$.

3. *Cordes.* On tire une corde au hasard dans un disque de rayon R. Quelle est la probabilité que la longueur ℓ de la corde soit $\geq R$?

a. ℓ varie continûment dans $[0, 2R]$, de sorte que la probabilité cherchée vaut $1/2$.

b. ℓ est déterminée par la distance d de la corde au centre du disque ; d varie continûment dans $[0, R]$, et $\ell = 2\sqrt{R^2 - d^2} \geq R \Leftrightarrow d \leq R\sqrt{3}/2$, de sorte que la probabilité cherchée vaut $\sqrt{3}/2$.

c. ℓ est déterminée par le milieu M de la corde, qui varie continûment dans le disque ; et $\ell \geq R$ a lieu ssi M est dans le disque concentrique de rayon $\sqrt{3}/2$, de sorte que la probabilité cherchée vaut $3/4$.

Explication : la probabilité choisie est très insuffisamment précisée par l'expression "tirage au hasard". Ici on a considéré successivement la probabilité uniforme sur l'ensemble : des longueurs, des distances au centre, des milieux. Ce sont trois probabilités différentes !

4. Jeu de pile ou façe illimité ; première apparition de "pile", ou d'une séquence donnée.

Exercice n^o 1.1 Est-il plus probable d'obtenir au moins une fois 6 en lançant 4 dés usuels, ou bien d'obtenir au moins une fois un double 6 en lançant 24 fois 2 dés usuels ?

Exercice n^o 1.2 On lance n fois de suite 3 dés normaux. Pour quelles valeurs de n la probabilité d'avoir obtenu au moins un 421 dépasse-t-elle $\frac{1}{2}$?

Exercice n^o 1.3 On lance 5 pièces de monnaie. Calculer les probabilités des événements suivant : "la 1ère pièce donne face" ; "face sort exactement 2 fois" ; "face sort au plus 3 fois".

Exercice n^o 1.4 On lance 10 dés usuels. Calculer les probabilités des événements suivant : "6 ne sort pas" ; "6 sort 1 fois exactement" ; "6 sort 3 fois exactement" ; "6 sort 2 fois au moins" ; "6 sort 3 fois au moins".

Exercice n^o 1.5 Une armoire contient 10 paires de chaussures, parmi lesquelles on prélève au hasard 8 chaussures. Quelle est la probabilité d'avoir ainsi k paires de chaussures exactement ?

Exercice n^o 1.6 Une urne contient n boules noires et b boules blanches. Deux joueurs X et Y tirent avec remise une boule dans l'urne, tour à tour, X tirant le premier. Quelle est la probabilité que X soit le premier à tirer une boule noire ? Même question sans remise.

Exercice n^o 1.7 Une loterie comporte 100 billets, dont les seuls billets gagnants suivant : 1 billet gagne 50 euros, 5 billets gagnent chacun 30 euros, 10 billets gagnent chacun 10 euros. Quelle est la probabilité qu'un acheteur de 3 billets gagne 30 euros (au moins, puis exactement) ?

Exercice n^o 1.8 Un joueur X lance 2 dés usuels, et obtient ainsi la somme S.

a) Calculer la probabilité que $S > n$, en fonction des différentes valeurs de l'entier n.

b) Un joueur Y relance les 2 dés et obtient une somme T. Quelles sont les probabilités que $S = T$, que $S > T$, que $S \geq T$?

Proposition 1.6 *(Formule du crible, de Poincaré) Pour tout espace probabilisé* $(\Omega, \mathcal{T}, I\!\!P)$, *tout entier* $n \in I\!\!N^*$, *et tous événements* $A_1, \ldots, A_n$, *on a :*

$$I\!\!P\Big[A_1 \cup \ldots \cup A_n\Big] = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} I\!\!P\Big[A_{i_1} \cap \ldots \cap A_{i_k}\Big].$$

Preuve Par récurrence sur n : exercice.

Exercice n^o 1.9 Un sac contient n jetons numérotés de 1 à n, qu'on tire tous 1 à 1, sans remise. $i)$ Calculer $p_n := I\!\!P$(au moins un jeton sorte au rang indiqué par son numéro), sa limite p_∞, et majorer $|p_n - p_\infty|$.

$ii)$ Soit $p_n(k) := I\!\!P$(exactement k jetons sortent au rang indiqué par leur numéros), pour $k \in \{0, .., n\}$. Déduire de (i) une formule pour $p_n(k)$, puis $\lim\limits_{n\to\infty} p_n(k)$ (pour tout $k \in I\!\!N$).

1.7 Probabilités conditionnelles

Définition 1.8 (Probabilité conditionnelle) *Fixons un espace de probabilité* $(\Omega, \mathcal{T}, \mathbb{P})$*, et un événement* $C \in \mathcal{T}$*, non négligeable. La probabilité conditionnelle relative à* C *(ou* $\ll$ *sachant* C $\gg$*) est définie par :* $\mathbb{P}(A/C) := \mathbb{P}(A \cap C)/\mathbb{P}(C)$.

On vérifie immédiatement qu'il s'agit encore d'une probabilité sur $(\Omega, \mathcal{T})$, et que conditionner successivement par deux événements C et C' d'intersection non négligeable revient à conditionner par leur intersection : $\mathbb{P}\big(\,\cdot\,/C/C'\big) = \mathbb{P}(\cdot/C \cap C')$.

Exercice n^o 1.2.1 Lancer de 2 dés usuels : $\Omega = \{1, .., 6\}^2$. $\mathbb{P}$ uniforme. Soient X_1 le chiffre indiqué par le premier dé, S la somme des chiffres indiqués par les 2 dés, et $C = \{S = 5\}$. Dresser le tableau des valeurs de $\mathbb{P}(\cdot/C)$, puis de $\mathbb{P}(X_1 = \cdot/C)$.

Exercice n^o 1.2.2 Vous attendez un ami de Vancouver, qui voyage jusqu'à Strasbourg avec changement d'avion à New York, Londres et Francfort. La probabilité d'attentat est estimée à p pour chacun des 4 vols, avec indépendance entre les 4 vols. Votre ami n'arrivant pas, quelle est la probabilité que l'attentat ait eu lieu : a) dans le 1er avion ? b) dans le 2ème avion ? c) dans le 3ème avion ? c) dans le 4ème avion ?

Pourquoi la somme ne vaut-elle pas 1 ?

Pour effectuer un calcul, il est très souvent indispensable de pouvoir $\ll$ distinguer des cas $\gg$. Cela s'exprime par la formule suivante, très élémentaire et très utile à la fois.

Proposition 1.9 (Formule des probabilités totales) *Fixons un espace probabilisé* $(\Omega, \mathcal{T}, \mathbb{P})$ *et une partition de* Ω *en événements non négligeables :* $\Omega = \bigsqcup_{j=1}^{N} C_j$*. Alors nous avons* $\mathbb{P}(A) = \sum_{j=1}^{N} \mathbb{P}(A/C_j)\mathbb{P}(C_j)$*, pour tout événement* A*.*

Exercice n^o 1.2.3 Une urne contient b boules blanches et n boules noires. Quand une boule est tirée, on le remet dans l'urne, avec ℓ boules de la même couleur. On effectue ainsi 3 tirages au hasard. a) Quelle est la probabilité que la 1ère boule tirée soit noire sachant que la seconde est blanche ? b) Quelle est la probabilité que la 3ème boule soit noire ?

On a souvent à inverser un conditionnement. Cela se fait simplement, au moyen de la formule élémentaire suivante, très utile aussi, quoiqu'également de preuve immédiate.

Proposition 1.10 (Formule de Bayes) *Fixons un espace de probabilité* $(\Omega, \mathcal{T}, \mathbb{P})$*, et une partition de* Ω *en événements non négligeables :* $\Omega = \bigsqcup_{j=1}^{N} C_j$*. Alors nous avons pour tout événement non négligeable* A *et tout* $k \in \{1, .., N\}$ *:*

$$\mathbb{P}(C_k/A) = \mathbb{P}(A/C_k)\mathbb{P}(C_k) \Big/ \sum_{j=1}^{N} \mathbb{P}(A/C_j)\mathbb{P}(C_j)\,.$$

Exemple : Les candidats à un examen proviennent de 4 lycées K,L,V,W, à raison de 20% pour K, de 25% pour L, de 15% pour V, et de 40% pour W. K enregistre 35% de succès, L 30%, V 50%, W 45%. Alors la probabilité qu'un candidat reçu provienne de K est $\frac{7}{40}$, de L est $\frac{3}{16}$, de V est $\frac{3}{16}$, de W est $\frac{9}{20}$.

<u>Exercice n^o 1.2.4</u> Trois machines U, V, W produisent une même pièce dans une usine. U assure 40% de la production, V 35%, et W le reste. U produit 20% de pièces défectueuses, V 15%, et W 10%.

a) Quelle est la probabilité qu'une pièce prise au hasard soit défectueuse ?

b) Quelle est la probabilité qu'une pièce défectueuse prise au hasard provienne de U ?

<u>Exercice n^o 1.2.5</u> Vous allez chez des gens dont vous savez qu'ils ont 2 enfants, dont au moins une fille. a) Quelle est la probabilité que l'autre enfant soit aussi une fille ?

b) Une fille vous ouvre la porte ; quelle est la probabilité que l'autre enfant soit aussi une fille ? c) En l'absence de l'information préalable qu'ils ont au moins une fille, mais en voyant une fille ouvrir la porte, quelle est la probabilité que l'autre enfant soit aussi une fille ?

<u>Exercice n^o 1.2.6</u> Trois condamnés X, Y, Z sont informés que l'un d'eux, choisi au hasard, va être exécuté, et que les 2 autres vont être libérés. Mais ils ne doivent pas encore savoir qui le hasard a désigné. X demande au geôlier de lui nommer l'un de ses 2 codétenus devant être libéré, arguant que cette information serait innocente, puisqu'il sait que l'un des 2 au moins doit l'être. Le geôlier refuse, arguant que cette information modifierait réellement l'estimation que X peut faire de ses chances. Qui a raison ?

<u>Exercice n^o 1.2.7</u> 12% des individus d'une population sont porteurs d'une certaine maladie. Un test relatif à cette maladie est fiable à 95%, dans le cas d'un malade comme dans le cas d'un sujet sain. a) Quelle est la probabilité qu'un individu présentant un test positif soit effectivement malade ? b) Quelle est la probabilité qu'un individu présentant un test négatif soit effectivement sain ?

<u>Exercice n^o 1.2.8</u> Émile possède 5 pièces de monnaie, dont 2 sont normales, 2 ont 2 côtés "face", et une a 2 côtés "pile".

a) Il prend une pièce au hasard et la lance ; quelle est la probabilité qu'il voie "face" ?

b) Il voit "face"; quelle est la probabilité que l'autre côté de la pièce soit aussi "face" ?

Il relance la même pièce.

c) Quelle est la probabilité que le côté caché de la pièce soit "face" ?

d) Il voit de nouveau "face"; quelle est la probabilité que l'autre côté de la pièce soit aussi "face" ? Il choisit ensuite au hasard une des autres pièces et la lance.

e) Quelle est la probabilité de voir de nouveau "face" (pour la troisième fois) ?

<u>Exercice n^o 1.2.9</u> Un livre a une probabilité $p > 0$ de se trouver dans une commode comportant k tiroirs, et des chances égales de se trouver dans chacun des tiroirs.

i) On ouvre les $(k-1)$ premiers tiroirs, sans le trouver ; quelle est la probabilité de le trouver dans le dernier tiroir ?

ii) Soit $j \in \{2, .., k-1\}$. On ouvre les $(k-j)$ premiers tiroirs, sans le trouver ; quelle est la probabilité de le trouver dans le dernier tiroir ? dans l'un des j derniers tiroirs ?

Exercice n^o 1.2.10 Le quart d'une population est vacciné contre le choléra. Au cours d'une épidémie, on constate qu'il y a parmi les malades un vacciné pour 4 non-vaccinés, et qu'il y a un malade sur 12 parmi les vaccinés. Quelle est la probabilité qu'un non-vacciné tombe malade ? Le vaccin est-il efficace ?

1.11 Événements indépendants

L'indépendance est au cœur de la théorie des probabilités. On l'aborde en douceur, par le cas des événements. Le cas général des variables aléatoires viendra ensuite (section 4).

Définition 1.12 *Fixons un espace de probabilité $(\Omega, \mathcal{T}, I\!\!P)$. Deux événements A et B sont dits indépendants lorsque $I\!\!P(A \cap B) = I\!\!P(A) I\!\!P(B)$.*

Exemples : 1) "tirer un roi" et "tirer un trèfle", dans un jeu de bridge, ou de belote.

2) "obtenir un chiffre pair avec le 1er dé" et "obtenir un 6 avec le 2ème dé", lors du lancer de 2 dés.

3) "obtenir une somme paire" et "obtenir un 5 avec le 2ème dé", lors du lancer de 2 dés.

4) "obtenir une somme égale à 7" et "obtenir un produit égal à 12", lors du lancer de 2 dés (non indépendants !)

5) "obtenir un produit pair" et "obtenir un 5 avec le 2ème dé", lors du lancer de 2 dés.

Exercice n^o 1.11.1 Montrer que 2 événements A et B sont indépendants ssi A et B^c le sont, et ssi A^c et B^c le sont, ou bien encore (lorsqu'ils sont non négligeables) ssi $I\!\!P(A/B) = I\!\!P(A)$, et ssi $I\!\!P(B/A) = I\!\!P(B)$.

Exercice n^o 1.11.2 Une urne contient des jetons numérotés, rouges ou noirs. Lors du tirage d'un jeton la probabilité d'en tirer un rouge est $3/5$; d'en tirer un de numéro impair est $2/3$; d'en tirer un rouge et pair est p. Que vaut la probabilité d'en tirer un noir impair ? Pour quelles valeurs de p les événements "noir" et "impair" sont-ils indépendants ?

Définition 1.13 *Fixons un espace de probabilité $(\Omega, \mathcal{T}, I\!\!P)$. Des événements $A_1, \ldots, A_n$ sont dits indépendants lorsque $I\!\!P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = I\!\!P(A_{i_1}) I\!\!P(A_{i_2}) \ldots I\!\!P(A_{i_k})$ pour tous $1 \leq i_1 < \ldots . < i_k \leq n$. Les événements d'une suite sont dits indépendants lorsque toute sous-suite finie est constituée d'événements indépendants.*

Proposition 1.14 *Les événements $A_1, \ldots, A_n$ sont indépendants ssi*

$I\!\!P(A_1^{\varepsilon_1} \cap A_2^{\varepsilon_2} \cap \ldots \cap A_n^{\varepsilon_n}) = I\!\!P(A_1^{\varepsilon_1}) I\!\!P(A_2^{\varepsilon_2}) \ldots I\!\!P(A_n^{\varepsilon_n})$ *pour tout choix de* $\varepsilon_1, \ldots, \varepsilon_n$, *avec soit* $A_j^{\varepsilon_j} = A_j$ *soit* $A_j^{\varepsilon_j} = A_j^{\mathbf{c}}$.

Preuve Pour le sens direct, on vérifie par récurrence sur $k \in \{0,\ldots,n\}$ que pour tout choix de $\varepsilon_1,\ldots,\varepsilon_k$, tout $\ell \in \{k+1,\ldots,n\}$ et tous $i_{k+1},\ldots,i_\ell \in \{k+1\ldots,n\}$, on a :

$$P\Big[A_1^{\varepsilon_1} \cap \ldots \cap A_k^{\varepsilon_k} \cap A_{i_{k+1}} \cap \ldots \cap A_{i_\ell}\Big] = P[A_1^{\varepsilon_1}] \times \ldots \times \mathbb{P}[A_k^{\varepsilon_k}] \times \mathbb{P}[A_{i_{k+1}}] \times \ldots \times \mathbb{P}[A_{i_\ell}] \,.$$

Pour la réciproque, si $i_1,\ldots,i_k$ sont fixés dans $\{1,\ldots,n\}$, on note $\{j_1,\ldots,j_{n-k}\} := \{1,\ldots,n\} \setminus \{i_1,\ldots,i_k\}$, et on applique la formule des probabilités totales avec la partition $\{A_{j_1}^{\varepsilon_{j_1}} \cap \ldots \cap A_{j_{n-k}}^{\varepsilon_{j_{n-k}}} \,|\, \varepsilon_{j_1},\ldots,\varepsilon_{j_{n-k}} =$ ou $c\}$ (c'est bien une partition de Ω, car l'ensemble des intersections des éléments de 2 partitions en forme encore une). $\diamond$

Exercice n^o 1.11.3 On jette un dé normal n (≥ 3) fois de suite. Montrer que les événements {les lancers j et k donnent le même résultat} sont deux à deux indépendants, mais non indépendants.

Exercice n^o 1.11.4 Sur $\Omega := \{a,b,c,d\}$ on définit $\mathbb{P}$ par : $\mathbb{P}(\{a\}) = \alpha$, $\mathbb{P}(\{b\}) = \beta$, $\mathbb{P}(\{c\}) = \gamma$, $\mathbb{P}(\{d\}) = \delta$. Trouver les valeurs de $\alpha, \beta, \gamma, \delta$ telles que les événements $A := \{b,c\}$, $B := \{c,a\}$, $C := \{a,b\}$ soient 2 à 2 indépendants mais non indépendants.

Exercice n^o 1.11.5 Pour ridiculiser un astrologue (ou un ≪ voyant ≫, ou n'importe quelle autre sorte de charlatan), on le défie de prédire le résultat de 12 lancers successifs d'une pièce de monnaie usuelle. Quelle est la probabilité (en fonction de $n \in \{0,1,..,5\}$) qu'il se trompe au plus n fois ?

2 Variables aléatoires et leur lois

Définition 2.1 *Une variable aléatoire (≪ v.a. ≫) est une fonction V définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^d$, telle que $\{V \in E\} = V^{-1}(E) \in \mathcal{T}$, pour tout E pavé (ou ouvert ou fermé) de $\mathbb{R}^d$. Sa loi est la probabilité $\mathbb{P}_V := \mathbb{P} \circ V^{-1}$ sur $\mathbb{R}^d$, définie par : $\mathbb{P}_V(E) := \mathbb{P}(V^{-1}(E)) = \mathbb{P}(V \in E)$. Lorsque $d = 1$, on parle de variable aléatoire réelle, ≪ v.a.r. ≫.*

On vérifie immédiatement que la loi d'une variable aléatoire est bien une probabilité (sur $V(\Omega) \subset \mathbb{R}^d$, ou directement sur $\mathbb{R}^d$, la tribu étant celle des “boréliens”, engendrée par les pavés ou par les ouverts ou par les fermés de $\mathbb{R}^d$). Notons que dans le cas où Ω est discret (fini ou dénombrable), dans la définition ci-dessus la condition de mesurabilité sur V est vide, c'est-à-dire qu'elle est forcément vérifiée par n'importe quelle fonction V.

On vérifie les propriétés suivantes, qui autorisent toutes les opérations usuelles sur les variables aléatoires.

Proposition 2.2 *Une fonction de Ω dans $\mathbb{R}^d$ est une v.a. ssi ses coordonnées (dans n'importe quelle base de $\mathbb{R}^d$) sont des v.a.r.. Une combinaison linéaire de v.a. est encore une v.a.. Un produit de v.a.r. est une v.a.r.. La composée d'une v.a. par une fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}^{d'}$ est encore une v.a..*

Exemples: La somme et le produit des chiffres indiqués par 2 dés ; les durées de vie de particules fissiles ; le nombre de fois qu'une suite de N lancers d'une pièce de monnaie donne pile ; les temps qu'il faut attendre pour tirer dans une urne, lors de tirages successifs, une boule rouge, une boule verte ; ou bien lors de lancers illimités d'une pièce, pour obtenir une séquence donnée ; etc...

La notion d'*espérance* est fondamentale dans la théorie des probabilités. Il s'agit en fait simplement d'une somme ou d'une série ou d'une intégrale, qui est aussi une moyenne.

Définition 2.3 *L'espérance d'une variable aléatoire V définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ est son intégrale par rapport à la probabilité $\mathbb{P}$:*

$$\mathbb{E}(V) := \int_\Omega V \, d\mathbb{P} \, .$$

Elle existe lorsque (la norme de) V est intégrable, c'est-à-dire lorsque $\mathbb{E}(\|V\|) < \infty$.

Exemple Lorsque $\Omega = \{\omega_1, \ldots, \omega_N\}$ est fini, $\mathbb{P}$ étant donnée par la liste des probabilités des singletons : $p_j := \mathbb{P}(\{\omega_j\})$, on a $\mathbb{E}(V) = \sum_{j=1}^{N} p_j \, V(\omega_j)$: c'est une moyenne pondérée.

Lorsque $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$ est dénombrable, on a $\mathbb{E}(V) = \sum_{j \geq 1} p_j \, V(\omega_j)$.

Corollaire 2.4 *Soient V une v.a. et f une fonction continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}^{d'}$. Alors*

$$\mathbb{E}(f \circ V) = \int_\Omega f \circ V \, d\mathbb{P} = \int_{\mathbb{R}^d} f \, d\mathbb{P}_V \, .$$

Lorsque la loi de V est discrète, c'est-à-dire lorsque presque sûrement V ne prend qu'un ensemble fini ou dénombrable de valeurs, alors on a

$$\mathbb{E}(f \circ V) = \sum_v f(v) \, \mathbb{P}(V = v) \, ,$$

cette série convergeant dès que $f \circ V$ est intégrable (ce qui a lieu par exemple dès que V ne prend qu'un nombre fini de valeurs, ou encore, dès que f est bornée).

Justification Pour toute v.a. discrète $V = \sum_j v_j \, 1_{A_j}$ (avec les v_j distincts 2 à 2), on a :

$$\int_\Omega f \circ V \, d\mathbb{P} = \sum_j f(v_j) \int 1_{A_j} d\mathbb{P} = \sum_j f(v_j) \mathbb{P}(A_j) = \sum_j f(v_j) \mathbb{P}_V(\{v_j\}) = \int_{\mathbb{R}^d} f \, d\mathbb{P}_V \, .$$

Exercice n^o 2.0 Lors du lancer de 2 dés usuels, calculer :

(i) l'espérance du résultat du 2ème dé, puis du carré de ce résultat ;

(ii) l'espérance de la somme des 2 dés.

Proposition 2.5 *L'espérance est linéaire :* $\mathbb{E}\Big(\sum_{j=1}^{N}\lambda_j\,V_j\Big) = \sum_{j=1}^{N}\lambda_j\,\mathbb{E}(V_j)$ *, pour toutes v.a. intégrables* $V_1,..,V_N$ *et tous réels* $\lambda_1,..,\lambda_N$ *(c'est une propriété générale, fondamentale, de toute intégrale). En particulier, l'espérance d'une v.a. vectorielle intégrable, dans n'importe quelle base (déterministe), a pour coordonnées les espérances des coordonnées.*

Exercice n^o 2.1 On tire au hasard et sans remise toutes les boules d'une urne remplie de r boules rouges et n boules noires.

1) Montrer que que la probabilité de tirer une boule rouge au k-ième tirage ne dépend pas de k. (Il y a au moins 2 solutions bien différentes : soit par récurrence, soit en considérant directement l'ensemble de tous les tirages exhaustifs possibles.)

2) Quelle est l'espérance du rang de la k-ième boule rouge ? (Considérer le vecteur $(X_1, X_2 - X_1, \ldots, X_r - X_{r-1}, r+n+1-X_r)$, X_j désignant le rang de la j-ième boule rouge).

Exercice n^o 2.2 Une urne U contient initialement 2 boules blanches, et une urne U' 2 boules noires. A chaque instant, on tire au hasard une boule par urne, et on les interchange.

Notons X_k le nombre de boules blanches dans U au k-ième instant, et V_k le vecteur-colonne donnant la loi de X_k.

a) Quelle est la relation entre V_{k+1} et V_k ? b) Calculer $\lim\limits_{k\to\infty} \mathbb{P}(X_k = 2)$.

Soient T le premier instant où U contient deux boules noires, $p_k := \mathbb{P}(T \geq k, X_k = 1)$, et $q_k := \mathbb{P}(T \geq k, X_k = 2)$.

c) Exprimer (p_{k+1}, q_{k+1}) en fonction de (p_k, q_k), puis p_{k+1} en fonction de (p_k, p_{k-1}).

d) Déduire la valeur de p_k, puis la loi de T. Que vaut $\mathbb{P}(T = \infty)$?

Exercice n^o 2.3 Un marchand de journaux a X clients par jour, X étant une v.a. entière intégrable, de loi supposée connue. Il gagne a par journal vendu, perd b par journal invendu, et perd c par client insatisfait. Quel est le nombre n de journaux qu'il doit commander par jour pour optimiser son gain moyen ?

Exercice n^o 2.4 Montrer que pour toute v.a.r. $Z \geq 0$, on a $\mathbb{E}(Z) = \displaystyle\int_0^\infty \mathbb{P}(Z > s)\,ds$. Particulariser au cas où Z prend ses valeurs dans $\mathbb{N}$.

Définition 2.6 *Une variable aléatoire* V *définie sur un espace de probabilité* $(\Omega, \mathcal{T}, \mathbb{P})$ *(et à valeurs dans* $\mathbb{R}^d$*) admet la* <u>*densité*</u> h *lorsque sa loi est donnée par*

$$\mathbb{E}(f\circ V) := \int_{\mathbb{R}^d} f(v)\,h(v)\,dv\,, \quad \textit{pour toute fonction continue bornée } f : \mathbb{R}^d \to \mathbb{R}\,.$$

La densité h est forcément une fonction intégrable de $\mathbb{R}^d$ dans $\mathbb{R}_+$, d'intégrale égale à 1.

Attention, la plupart des variables aléatoires ne sont ni discrètes ni à densité ! Il suffit de songer à une v.a. V valant une v.a. discrète X avec probabilité $1/2$ et une v.a. à densité Y avec probabilité $1/2$: $I\!P_V = (I\!P_X + I\!P_Y)/2$. Imaginer par exemple un tir sur cible, avec probabilité $1/2$ de rater la cible, auquel cas la v.a. prend la valeur disons -1, et probabilité $1/2$ d'atteindre la cible, auquel cas la v.a. prend ses valeurs dans le disque formé par la cible, avec par exemple la loi uniforme.

Cela dit, les lois usuelles, celles qu'on rencontre le plus souvent, sont discrètes ou à densité (on dit aussi : absolument continues).

Définition 2.7 *Une v.a. V (définie sur un espace de probabilité $(\Omega, \mathcal{T}, I\!P)$) est dite de carré intégrable ou dans L^2 lorsque $I\!E(\|V\|^2) < \infty$. Elle est alors nécessairement intégrable.*

La variance d'une v.a.r. V de carré intégrable est :

$$\mathrm{Var}(V) := I\!E\Big[|V - I\!E(V)|^2\Big] = I\!E(V^2) - I\!E(V)^2 .$$

La covariance de deux v.a.r. V, V' de carré intégrable est :

$$\mathrm{Cov}(V, V') := I\!E\Big[(V - I\!E(V)) \times (V' - I\!E(V'))\Big] = I\!E(VV') - I\!E(V)I\!E(V') .$$

La matrice de covariance (ou de dispersion) d'une v.a. $V = (V_1, \ldots, V_d) \in I\!R^d$ de carré intégrable est :

$$K_V := I\!E\Big[{}^t(V - I\!E(V)) \times (V - I\!E(V))\Big] = I\!E({}^tV\, V) - I\!E({}^tV)I\!E(V) = \Big(\Big(\mathrm{Cov}(V_i, V_j)\Big)\Big)_{1 \le i,j \le d} .$$

Proposition 2.8 *La covariance est bilinéaire, symétrique, positive, et* $\mathrm{Var}(V)$ *est nulle ssi V est p.s. constante. En outre* $\mathrm{Var}(\lambda V + V') = \lambda^2 \mathrm{Var}(V) + 2\,\lambda\,\mathrm{Cov}(V, V') + \mathrm{Var}(V')$.

Preuve : La symétrie : $\mathrm{Cov}(V, V') = \mathrm{Cov}(V', V)$ est évidente. La linéarité à gauche : $\mathrm{Cov}(a_1V_1 + a_2V_2, V') = a_1\mathrm{Cov}(V_1, V') + a_2\mathrm{Cov}(V_2, V')$ découle immédiatement de la linéarité de l'espérance. Par symétrie elle entraîne la linéarité à droite, et donc la bilinéarité. La positivité : $\mathrm{Cov}(V, V) = \mathrm{Var}(V) \ge 0$ est claire. Ensuite, $\mathrm{Var}(\textit{constante}) = 0$ est clair, et réciproquement, si $I\!P[V \ne I\!E(V)] > 0$, alors il existe $\varepsilon > 0$ tel que $I\!P(|V - I\!E(V)| > \varepsilon) > \varepsilon$, de sorte qu'alors $\mathrm{Var}(V) = I\!E(|V - I\!E(V)|^2) > \varepsilon \times \varepsilon^2 > 0$. Enfin, la dernière formule découle très simplement du développement du carré d'une somme et de la linéarité de l'espérance (détails à rédiger en exercice). ⋄

Constatant que le trinôme en λ de cette proposition doit toujours être positif ou nul, et donc doit avoir son discriminant négatif ou nul, on obtient immédiatement l'inégalité de Schwarz (dite aussi de Cauchy-Schwarz) :

Corollaire 2.9 *(Inégalité de Schwarz) La covariance est majorée par le produit des écart-types :* $|\mathrm{Cov}(V,V')| \le \sigma(V)\sigma(V')$ *, où l'<u>écart-type</u> de la v.a.r.* V *est* $\sigma(V) := \sqrt{\mathrm{Var}(V)}$ *.*
De sorte que le <u>coefficient de corrélation linéaire</u> $\varrho(V,V') := \dfrac{\mathrm{Cov}(V,V')}{\sigma(V)\sigma(V')}$
(défini pour V *et* V' *v.a.r. de carré intégrable et non p.s. constantes) appartient à* $[-1,1]$*. Il vaut* ± 1 *ssi* $V = aV' + b$ *(presque sûrement, pour* a, b *réels fixes).*

<u>Exercice n° 2.6</u> a) Vérifier que $K_V = \mathbb{E}({}^tV\,V) - {}^t\mathbb{E}(V)\mathbb{E}(V) = \Big(\big(\mathrm{Cov}(V_i,V_j)\big)\Big)_{1\le i,j\le d}$.

b) Vérifier que pour tout $u \in \mathbb{R}^d$ on a $\mathrm{Var}(u\,{}^tV) = uK_V\,{}^tu = \sum_{1\le i,j\le d} u_i u_j \mathrm{Cov}(V_i,V_j)$.

c) Montrer que K_V est une matrice symétrique, et positive : $vK_V\,{}^tv \ge 0$ $(\forall\, v \in \mathbb{R}^d)$.

d) Montrer que K_V est inversible ssi il n'existe pas d'hyperplan de $\mathbb{R}^d$ contenant p.s. V.

e) Soit M une matrice de $\mathcal{M}_{n,d}(\mathbb{R})$. Calculer $\mathbb{E}(MV)$ et K_{MV}. Même question pour AV, si A est une application affine de $\mathbb{R}^d$ dans $\mathbb{R}^n$.

<u>Exercice n° 2.7</u> <u>Inégalité de Markov, ou de Bienaymé-Tchebitchev</u> :

Vérifier que $\mathbb{P}[|V| \ge v] \le \mathbb{E}[|V|^k]/v^k$, pour tous $v > 0$ et $k \ge 0$.

<u>Exercice n° 2.8</u> Vous écrivez chaque jour avec probabilité 1 si vous n'avez pas écrit la veille, et avec probabilité $1/2$ sinon. Montrez que vous écrivez ainsi en moyenne 243 lettres par an. (Considérer pour chaque jour la variable indicatrice de l'événement "écrire".)

L'inégalité suivante est une importante inégalité de convexité.

Proposition 2.10 *(Inégalité de Jensen) Pour toute v.a.r. intégrable* V *et toute fonction <u>convexe</u>* ψ *de* $\mathbb{R}$ *dans* $\mathbb{R}$*, telle que* $\psi \circ V$ *soit positive ou intégrable, on a*

$$\mathbb{E}[\psi \circ V] \ge \psi(\mathbb{E}[V]) .$$

<u>Exemples</u> : On a $|x_1 + \ldots + x_n|^p \le n^{p-1}(|x_1|^p + \ldots + |x_n|^p)$ pour tous $n \in \mathbb{N}^*$, $p \in [1,\infty[$, et $x_1, \ldots, x_n \in \mathbb{R}$. Et $(\mathbb{E}[|V|^p])^{1/p} \ge (\mathbb{E}[|V|^q])^{1/q}$ si $p \ge q > 0$.

3 Lois usuelles

3.1 Lois usuelles discrètes

Définition 3.2 *La <u>loi uniforme</u> sur un espace de probabilité fini est celle qui attribue la même valeur à tous les singletons.*

Corollaire 3.3 *Si* $\mathbb{P}$ *est uniforme sur* Ω *fini, alors* $\mathbb{P}(A) = \mathrm{Card}(A)/\mathrm{Card}(\Omega)$ *pour toute partie* A *de* Ω*.*

Exercice n^o 3.1.1 Calculer l'espérance et la variance d'une v.a. uniforme sur un intervalle de $\mathbb{Z}$.

Définition 3.4 *La loi de Bernoulli de paramètre p, notée $\mathcal{B}(p)$, est sur l'espace $\{0,1\}$ la loi (de probabilité) qui attribue la valeur p au singleton $\{1\}$. Ici $0 \leq p \leq 1$.*

Définition 3.5 *La loi binômiale de paramètres n et p, notée $\mathcal{B}(n,p)$, est sur l'espace $\{0,\dots,n\}$ la loi (de probabilité) qui attribue la valeur $C_n^k\, p^k(1-p)^{n-k}$ au singleton $\{k\}$. Ici $n \in \mathbb{N}^*$ et $0 \leq p \leq 1$.*

Remarque 3.6 C'est la loi de la somme de n v.a. indépendantes de même loi $\mathcal{B}(p)$. La formule du binôme est exactement équivalente au fait que la somme de ces probabilités $C_n^k\, p^k(1-p)^{n-k}$ vaut bien 1.

Exercice n^o 3.1.2 Montrer que l'espérance et la variance d'une v.a. de loi $\mathcal{B}(n,p)$ valent respectivement np et $np(1-p)$.

Définition 3.7 *La loi hypergéométrique de paramètre N,n,p, notée $\mathcal{H}(N,n,p)$, est sur l'espace $\{0,\dots,n\}$ la loi (de probabilité) qui attribue la valeur $\dfrac{C_{Np}^k\, C_{N(1-p)}^{n-k}}{C_N^n}$ au singleton $\{k\}$. Ici $N, Np, n \in \mathbb{N}$, $n \leq N$, et $0 \leq p \leq 1$.*

Remarque 3.8 C'est la loi du nombre d'éléments possédant un caractère donné K dans un échantillon de taille n, prélevé au hasard, uniformément parmi les parties de cardinal n, dans un ensemble de cardinal N, dont une proportion p possède le caractère K.

Cela s'applique aussi bien à des pièces défectueuses dans une production industrielle, qu'aux individus malades dans une population, etc...

Exercice n^o 3.1.3 a) Montrer que l'espérance d'une v.a. de loi $\mathcal{H}(N,n,p)$ vaut np.

b) Calculer sa variance. c) Vérifier que $\lim\limits_{N\to\infty} \mathcal{H}(N,n,p)(k) = \mathcal{B}(n,p)(k)$, pour n,p,k fixés.

Définition 3.9 *La loi géométrique de paramètre p, notée $\mathcal{G}(p)$, est sur l'espace $\mathbb{N}^*$ la loi (de probabilité) qui attribue la valeur $(1-p)^{n-1}\, p$ au singleton $\{n\}$. Ici $0 < p < 1$.*

Remarque 3.10 C'est la loi du nombre N de tentatives nécessaires pour obtenir un résultat positif (succès), lors d'une suite de tentatives identiques et indépendantes, p étant la probabilité de succès à chaque tentative. On a $\mathbb{P}(N > n) = (1-p)^n$ pour tout $n \in \mathbb{N}$.

Exercice n^o 3.1.4 a) Les lois géométriques vérifient la propriété de non-vieillissement :

$$\mathbb{P}(N > n+m / N > m) = \mathbb{P}(N > n) \quad \text{pour tous } n,m \in \mathbb{N}.$$

b) Y a-t-il d'autres lois sur $\mathbb{N}^*$ qui vérifient cette propriété ?

Exercice n^o 3.1.5 Montrer que l'esp. et la variance d'une v.a. de loi $\mathcal{G}(p)$ valent $\frac{1}{p}$ et $\frac{1-p}{p^2}$.

Exercice n^o 3.1.6 Au cours d'un jeu illimité de pile ou face avec $\mathbb{P}(\text{pile}) = p$, on note X_k le rang de la k-ième apparition de "pile". Calculer la loi de X_k, son espérance et sa variance. Pour $n \in \mathbb{N}^*$, calculer $\mathbb{P}\Big((\exists\, k \in \mathbb{N}^*)\ X_k = n\Big)$.

Définition 3.11 *La loi de Poisson de paramètre λ, notée $\mathcal{P}(\lambda)$, est sur l'espace $\mathbb{N}$ la loi (probabilité) qui attribue la valeur $e^{-\lambda}\,\dfrac{\lambda^n}{n!}$ au singleton $\{n\}$. Ici $\lambda > 0$.*

Exercice n^o 3.1.7 a) Montrer que l'esp. et la variance d'une v.a. de loi $\mathcal{P}(\lambda)$ valent λ.

b) Vérifier que $\lim\limits_{np\to\lambda} \mathcal{B}(n,p)(\{k\}) = \mathcal{P}(\lambda)(\{k\})$, pour $\lambda > 0$, $k \in \mathbb{N}$ fixés et $n \to \infty$.

Exercice n^o 3.1.8 Quelle est la valeur la plus probable pour une variable aléatoire poissonnienne de paramètre λ ?

Exercice n^o 3.1.9 Un trousseau de n clefs contient une seule clef ouvrant une serrure donnée. On les essaie l'une après l'autre au hasard. Calculer la loi, l'espérance et la variance du nombre d'essais nécessaires. Même question si on réessaie à chaque fois une clef au hasard sans avoir écarté la précédente.

3.12 Lois usuelles à densité

Définition 3.13 *La loi uniforme sur un ouvert (ou un fermé) O de $\mathbb{R}^d$, notée $\mathcal{U}(O)$, est la loi admettant une densité constante (par rapport à la mesure de Lebesgue) sur O.*

Exercice n^o 3.12.1 Calculer l'esp. et la variance d'une v.a. uniforme sur un intervalle de $\mathbb{R}$.

Exercice n^o 3.12.2 (Aiguille de Buffon) Sur un sol plat sont tracées 2 droites parallèles D et D', distantes de $L \geq 1$. On laisse tomber entre D et D' une aiguille de longueur 1, puis on note x la distance du centre de l'aiguille à D, et θ l'angle que fait l'aiguille avec D. On suppose que la v.a. (x,θ) est uniforme sur $[0,L] \times [0,\pi/2]$. Vérifier que la probabilité que l'aiguille intersecte D ou D' vaut $2/(\pi L)$. Qu'en est-il si $0 < L < 1$?

Définition 3.14 *La loi exponentielle de paramètre λ, notée $\mathcal{E}(\lambda)$, est la loi admettant sur $\mathbb{R}_+$ la densité $t \mapsto \lambda\, e^{-\lambda t}$. Ici $\lambda > 0$.*

Exercice n^o 3.12.3 a) Les lois exponentielles vérifient la propriété de non-vieillissement : si la loi de Y est $\mathcal{E}(\lambda)$, alors $\mathbb{P}(Y > s + t / Y > s) = \mathbb{P}(Y > t)$ pour tous $s, t > 0$.

b) Y a-t-il d'autres lois à densité sur $\mathbb{R}_+$ qui vérifient cette propriété ?

c) En déduire la loi de la durée de vie d'un atome fissile, en fonction de sa demi-vie.

Exercice n^o 3.12.4 Montrer que l'espérance et l'écart-type d'une v.a. de loi $\mathcal{E}(\lambda)$ valent $\frac{1}{\lambda}$.

Définition 3.15 *Une variable aléatoire réelle est dite* <u>gaussienne centrée réduite</u> *ou* <u>normale centrée réduite</u> *ou* <u>gaussienne standard</u> *lorsqu'elle admet (sur $\mathbb{R}$) la densité :* $t \longmapsto \exp(-t^2/2)/\sqrt{2\pi}$. *Une variable aléatoire réelle X est dite* <u>gaussienne</u> *ou* <u>normale</u> *lorsqu'il existe $m \in \mathbb{R}$ et $\sigma > 0$ tels que $(X-m)/\sigma$ soit normale centrée réduite. On dit alors que la loi de X est $\mathcal{N}(m, \sigma^2)$.*

<u>Exercice n^o 3.12.5</u> a) Vérifier que $m = \mathbb{E}(X)$ et $\sigma^2 = \mathrm{Var}(X)$, et que $X \in \bigcap_{p<\infty} L^p$.

b) Vérifier que $\frac{A^2\, e^{-x^2/2}}{(A^2+1)\, x} \leq \int_x^\infty e^{-t^2/2} dt \leq \frac{e^{-x^2/2}}{x}$ sur $[A, \infty[$, pour tout $A > 0$. Déduire un équivalent de $\mathbb{P}(X > x)$ quand $x \to +\infty$.

c) Montrer que la densité de X est $t \longmapsto \exp[-(t-m)^2/2\sigma^2]/\sigma\sqrt{2\pi}$.

<u>Nota Bene</u> (i) La courbe de la densité gaussienne est la célèbre “courbe en cloche”.

(ii) Valeurs numériques à connaître :

$$\mathcal{N}(0,1)\Big(]-\infty\,, 1{,}65]\Big) \approx 0,95\,, \qquad \text{et} \qquad \mathcal{N}(0,1)\Big(]-\infty\,, 1{,}96]\Big) \approx 0,975\,.$$

<u>Exercice n^o 3.12.6</u> Pour être en cours à 8h, un étudiant en voiture a le choix entre un trajet sur petite route, dont la durée (en minutes) X suit la loi normale de moyenne 35,2 et de variance 25, et un trajet sur autoroute, dont la durée Y suit la loi normale de moyenne 40 et de variance 4. Il désire arriver à l'heure. Quel trajet doit-il préférer s'il part à 7h15 ? Et s'il part à 7h30 ?

Définition 3.16 *Une variable aléatoire V à valeurs dans $\mathbb{R}^d$ est dite* <u>gaussienne</u> *ou* <u>normale</u> *lorsque $u \cdot V = {}^t\!u\, V$ est une v.a.r. normale ou p.s. constante pour tout $u \in \mathbb{R}^d$. On note $\mathcal{N}(m, K)$ la loi d'un vecteur gaussien d'espérance m et de matrice de covariance K. Une probabilité (resp. une densité) est dite gaussienne lorsqu'elle est la loi (resp. la densité) d'un vecteur gaussien (non p.s. constant).*

<u>Exercice n^o 3.12.7</u> a) Vérifier que si V est un vecteur gaussien de $\mathbb{R}^d$ et si A est une application affine de $\mathbb{R}^d$ dans $\mathbb{R}^n$, alors AV est un vecteur gaussien.

b) Montrer que si V est un vecteur gaussien, alors ses coordonnées dans n'importe quelle base sont gaussiennes (ou p.s. constantes).

Proposition 3.17 *Soit K une matrice réelle symétrique positive, de format $d \times d$ et de rang r, et soit M une matrice réelle de format $d \times r$ telle que $K = M \times {}^t\!M$. Soit $m \in \mathbb{R}^d$. Soit V un vecteur de $\mathbb{R}^r$ formé de coordonnées gaussiennes standard indépendantes. Alors $m + MV$ est un vecteur gaussien de loi $\mathcal{N}(m, K)$.*

Corollaire 3.18 *Si la matrice de covariance d'un vecteur gaussien d-dimensionnel de loi $\mathcal{N}(m, K)$ est inversible, alors ce vecteur admet la densité sur $\mathbb{R}^d$:*

$$v \longmapsto (2\pi)^{-d/2}\, (\det K)^{-1/2} \exp\Big(-\tfrac{1}{2}\, {}^t(v-m) K^{-1} (v-m)\Big).$$

Preuve Soit V un vecteur de $\mathbb{R}^d$ formé de coordonnées gaussiennes standard indépendantes. Soit M une matrice réelle de format $d \times d$ telle que $K = M \times {}^tM$. Alors M est inversible, et nous avons pour toute fonction-test f, par changement de variable :

$$\mathbb{E}(f(m+MV)) = \int_{\mathbb{R}^d} f(m+Mx) \prod_{j=1}^{d} (2\pi)^{-1/2}\, e^{-x_j^2/2} dx = \int_{\mathbb{R}^d} f(m+Mx)\,(2\pi)^{-d/2} e^{-\,{}^t xx/2} dx$$

$$= \int_{\mathbb{R}^d} f(v)\,(2\pi)^{-d/2}\, e^{-\,{}^t(M^{-1}(v-m))\times(M^{-1}(v-m))/2}\, |\det(M^{-1})|\, dv$$

$$= \int_{\mathbb{R}^d} f(v)\,(2\pi)^{-d/2}\,(\det K)^{-1/2}\, \exp\Big(-\tfrac{1}{2}\,{}^t(v-m)K^{-1}(v-m)\Big)\, dv\ .\ \diamond$$

Exercice n^o 3.12.8 : Simulation Soit F une fonction de répartition sur $\mathbb{R}$. Pour tout $p \in [0,1]$, posons $G(p) := \inf\{x \in \mathbb{R} \,|\, F(x) \geq p\}$. (Nota Bene : $\inf \mathbb{R} = -\infty$ et $\inf \emptyset = +\infty$.)

a) Justifier l'existence dans $\mathbb{R}$ de $G(p)$ si $p \in]0,1[$. b) Montrer que $(\forall x \in \mathbb{R})\ G(F(x)) \leq x$.

c) Montrer que si $G(p) \in \mathbb{R}$, alors $F(G(p)) \geq p$. d) Montrer que $G(p) \leq x \Leftrightarrow F(x) \geq p$.

e) Montrer que si U est une variable aléatoire de loi uniforme sur $[0,1]$, alors $G \circ U$ admet F pour fonction de répartition. Nota Bene : Ceci est utilisé pour simuler des variables aléatoires.

f) Que vaut G lorsque F est bijective de $\mathbb{R}$ sur $]0,1[$? Comment simuler la loi $\mathcal{E}(\lambda)$?

4 Variables aléatoires indépendantes

Définition 4.1 *Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$. Des v.a. (définies sur $(\Omega, \mathcal{T})$) $V_1, \dots, V_n, \dots$ sont dites indépendantes lorsque pour tous $B_1, \dots, B_n, \dots$, les événements $\{V_1 \in B_1\}, \dots, \{V_n \in B_n\}, \dots$ sont indépendants (définition 1.13, section 1.11).*

Proposition 4.2 *Les v.a. $V_1, \dots, V_n, \dots$ sont indépendantes ssi*

$$\mathbb{E}\Big(f_1 \circ V_1 \times \dots \times f_n \circ V_n\Big) = \mathbb{E}\Big(f_1 \circ V_1\Big) \times \dots \times \mathbb{E}\Big(f_n \circ V_n\Big)$$

pour tous $n \in \mathbb{N}^*$ et toutes fonctions (mesurables) positives $f_1, \dots, f_n$.

Cela se dit aussi sous la forme : la loi de $(V_1, \dots, V_n, \dots)$ est le produit des lois des V_n.

En particulier, lorsque les v.a.r. indépendantes V_n admettent des densités h_n sur $\mathbb{R}$, alors la v.a. $(V_1, \dots, V_d)$ admet la densité $h_1 \otimes \cdots \otimes h_d$ sur $\mathbb{R}^d$

$\Big($*définie par :* $h_1 \otimes \cdots \otimes h_d\,(x_1, \dots, x_d) := h_1(x_1) \times \dots \times h_d(x_d).\Big)$

Remarque 4.3 1) Les événements $A_1, \ldots, A_n, \ldots$ sont indépendants ssi les v.a. $1_{A_1}, \ldots, 1_{A_n}, \ldots$ sont indépendantes (revoir la proposition 1.14, section 1.11).

2) Si des v.a.r. $V_1, \ldots, V_n, \ldots \in L^2$ sont indépendantes, alors leurs covariances sont nulles. Mais la réciproque est fausse. Contrexemples : a) Si G ($\mathcal{N}(0,1)$) et ε ($\mathcal{B}(\pm 1, \frac{1}{2})$) sont indépendantes, alors $\mathrm{Cov}(G, \varepsilon G) = 0$, mais $\mathbb{P}(G + \varepsilon G = 0) = \frac{1}{2}$ contredit l'éventuelle indépendance de G et de εG. b) Cas de $V_1, \ldots, V_n, \ldots$ indépendantes 2 à 2, mais non indépendantes. c) Cas de $V_1, \ldots, V_n, \ldots$ indépendantes mais de carré non intégrable.

3) Les v.a. discrètes $V_1, \ldots, V_n, \ldots$ sont indépendantes ssi pour tous $v_1, \ldots, v_n, \ldots$ les événements $\{V_1 = v_1\}, \ldots, \{V_n = v_n\}, \ldots$ sont indépendants.

<u>Exemples</u> : les résultats de différents dés, de lancers successifs d'une pièce, de tirages successifs (roulette, ...), les durées de vie de différents atomes fissiles (en l'absence de toute réaction en chaîne), etc...

<u>Exercice n^o 4.1</u> a) Soient U_1, U_2, U_3 trois v.a. indépendantes, uniformes sur $[0,1]$. Montrer qu'elles sont p.s. 2 à 2 distinctes. On les réordonne en $V_1 < V_2 < V_3$. Montrer que (V_1, V_2, V_3) admet la densité $6 \times 1_{\{0<v_1<v_2<v_3<1\}}$, et en déduire les densités de V_1, V_2, V_3.

b) Soient $U_1, \ldots, U_n$ des v.a.i.i.d., uniformes sur $[0,1]$, et $J_n := \min\{U_1, \ldots, U_n\}$, $M_n := \max\{U_1, \ldots, U_n\}$. Calculer $\mathbb{P}(x < J_n, M_n < y)$, et en déduire la loi (conjointe) de (J_n, M_n).
c) Que vaut $\lim\limits_{n\to\infty} \mathbb{P}(J_n < x < y < M_n)$?

<u>Exercice n^o 4.2</u> Soient $Y_1, \ldots, Y_n$ des v.a. indépendantes, exponentielles. Quelle est la loi de $\min\{Y_1, \ldots, Y_n\}$?

<u>Exercice n^o 4.3</u> Soient $X_1, .., X_n$ n v.a. indépendantes, X_j étant poissonnienne de paramètre λ_j. a) Quelle est la loi de $X_1 + X_2$, puis de $X_1 + \ldots + X_n$?

b) Calculer la loi, l'espérance et la variance de X_j sachant $X_1 + \ldots + X_n$.

c) Soient Y_j, $j \in \mathbb{N}$, des v.a. indépendantes à valeurs dans $\{1, .., n\}$, de même loi donnée par $\mathbb{P}(Y_j = k) = \alpha_k$, pour $1 \leq k \leq n$. Soient N une variable de Poisson de paramètre λ, indépendante des Y_j, et $N_k := \sum\limits_{j=1}^{N} 1_{\{Y_j = k\}}$. Montrer que les variables $N_1, \ldots, N_k$ sont poissonniennes indépendantes.

<u>Exercice n^o 4.4</u> Soient X_1, X_2 deux v.a.r. indépendantes de lois exponentielles de paramètres λ_1, λ_2. a) Calculer les lois de $J := \min\{X_1, X_2\}$, $M := \max\{X_1, X_2\}$.

b) Supposant que $\lambda_1 = \lambda_2$, montrer que J et $M - J$ sont des variables indépendantes.

<u>Exercice n^o 4.5</u> Soient X, Y deux variables aléatoires indépendantes, de densité respectivement f, g, avec $n \in \mathbb{N}^*$, $f(x) = 1_{\{x>0\}} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$, et $g(y) = 1_{\{y>0\}} \lambda\, e^{-\lambda y}$.

Calculer la densité de $X + Y$, $\mathbb{E}(X+Y)$, et $\mathrm{Var}(X+Y)$.

<u>Exercice n^o 4.6</u> Calculer la densité du carré d'une v.a.r. normale standard, puis de la somme de 2 tels carrés, puis du quotient de deux v.a.r. normales standards indépendantes.

Exercice n^o 4.7 Soient X et Y 2 v.a.r. indépendantes, admettant toutes 2 la densité $t \mapsto t^{-2}\, 1_{[1,\infty[}(t)$. On pose $U := XY$ et $V := X/Y$.

a) Calculer les lois de (U, V), de U, et de V. U et V sont-elles indépendantes ?

b) Calculer $\mathbb{E}(U^{-1/2}\, V^{-1})$.

Exercice n^o 4.8 Trois clients A, B, C arrivent au même temps 0 à la poste, où 2 guichets sont ouverts, qu'occupent A et B tout de suite. C remplace le premier des 2 qui a terminé. On admet que les temps de service X, Y, Z requis par ces 3 clients sont des v.a.r.i.i.d. de même loi $\mathcal{E}(\lambda)$.

a) Quelle est la loi du temps d'attente T de C ? b) Calculer la probabilité que C termine (et parte) le dernier. c) Calculer la loi du temps du dernier départ.

Exercice n^o 4.9 Justifier la remarque 3.6 (somme de v.a.i.i.d. de Bernoulli). Quand la somme de 2 variables aléatoires binômiales indépendantes est-elle binômiale ?

Exercice n^o 4.10 On effectue n tirages indépendants avec remise dans une urne contenant une proportion p_j de boules marquées j, pour $1 \le j \le r$, $r > 1$ étant fixé. On note N_j le nombre de boules marquées j qu'on tire ainsi. Préciser la loi du vecteur $N := (N_1, \dots, N_r)$, et calculer l'espérance et la variance de N_j , la covariance de N_j et N_k , et le nombre moyen de j tels que $N_j = 0$.

Exercice n^o 4.11 Montrer que si $X_1, \dots, X_n$ sont des v.a.i.i.d. exponentielles, alors $\max\{X_1, \dots, X_n\}$ et $X_1 + \frac{X_2}{2} + \cdots + \frac{X_n}{n}$ ont la même loi (pour tout $n \in \mathbb{N}^*$).

5 Convergences des variables aléatoires

Soient X et X_n , $n \in \mathbb{N}$, des v.a.r., définies sur un même espace de probabilité.

Définition 5.1 *On dit que la suite* $\{X_n\, , n \in \mathbb{N}\}$ *converge vers* X

i) *dans L^p* *lorsque* $\mathbb{E}(|X_n - X|^p) \to 0$ *(pour $1 \le p < \infty$) ; la convergence dans L^1 est aussi appelée convergence en moyenne, et la convergence dans L^2 convergence en moyenne quadratique ; la convergence dans L^∞ est la convergence uniforme presque sûre ;*

ii) *presque sûrement* *lorsque* $\mathbb{P}(X_n \text{ converge vers } X) = 1$ *;*

iii) *en probabilité* *lorsque pour tout* $\varepsilon > 0$ $\mathbb{P}(|X_n - X| > \varepsilon)$ *tend vers 0.*

Théorème 5.2 *1)* *Convergence monotone :* *Si (V_n) est une suite p.s. croissante de v.a.r. ≥ 0 , alors* $\lim_n\uparrow \mathbb{E}(V_n) = \mathbb{E}(\lim_n\uparrow V_n)$.

2) *Convergence dominée (de Lebesgue) :* *Si (X_n) est une suite de v.a. qui converge p.s. vers une v.a. X, et si elle est dominée par une v.a.r. fixe intégrable Y : $|X_n| \le Y \in L^1$, alors $X \in L^1$ et X_n converge vers X dans L^1.*

Exercice n^o 5.1 a) Montrer que la convergence presque sûre et la convergence dans L^p entraînent (chacune) la convergence en probabilité.
b) Montrer que la convergence dans L^p entraîne la convergence dans L^q si $p > q$.
c) Supposons que X_n converge en probabilité vers X et que Y_n converge en probabilité vers Y. Soient a et b réels. Montrer que $aX_n + bY_n$ et $X_n Y_n$ convergent en probabilité.

Exercice n^o 5.2 Trouver (sur $(\Omega = [0,1],\ \mathbb{P} \equiv dx)$)

a) une suite de v.a.r. qui converge dans L^q mais pas dans L^p (pour $p > q$ fixés) ;

b) une suite de v.a.r. qui converge dans L^p mais pas presque sûrement (pour $p < \infty$ fixé) ;

c) une suite de v.a.r. qui converge presque sûrement mais dans aucun L^p.

Exercice n^o 5.3 Soit $\{X_n \,|\, n \in \mathbb{N}^*\}$ une suite de v.a.r. indépendantes, telles que $\mathbb{P}(X_n = n) = 1/n$ et $\mathbb{P}(X_n = 0) = 1 - 1/n$ pour chaque $n \in \mathbb{N}^*$. Montrer que cette suite converge en probabilité, mais ni presque sûrement ni dans aucun L^p.

Exercice n^o 5.4 a) Montrer que $\{X_n \,|\, n \in \mathbb{N}\}$ converge en probabilité vers X ssi $\mathbb{E}(\min\{|X_n - X|, 1\})$ tend vers 0.
b) Vérifier que $(X,Y) \longmapsto \mathbb{E}(\min\{|X - Y|, 1\})$ est une distance sur l'ensemble noté $L^0(\Omega, \mathcal{T}, \mathbb{P}; \mathbb{R}^d)$ $\big($des v.a. sur $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d\big)$.

Définition 5.3 *On dit que la suite $\{X_n \,|\, n \in \mathbb{N}\}$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers X (à valeurs dans $\mathbb{R}^d$) lorsque la loi de X_n converge vaguement (ou étroitement) vers celle de X. Id est: lorsque $\mathbb{E}[f(X_n)]$ tend vers $\mathbb{E}[(f(X)]$, pour toute fonction f continue à support compact (ou continue bornée) de $\mathbb{R}^d$ dans $\mathbb{R}$.*

Proposition 5.4 *La convergence en probabilité entraîne la convergence en loi.*

Exercice n^o 5.6 a) Prouver que lorsque X est p.s. constante, les convergences vers X en probabilité et en loi sont équivalentes.
b) Montrer que ces deux convergences sont généralement non équivalentes.

Exercice n^o 5.7 Montrer que si X_n converge en loi vers X et si Y_n converge en probabilité vers 0, alors $X_n Y_n$ converge en probabilité vers 0. (Distinguer entre $\{|Y_n| > A\}$ et $\{|Y_n| \le A\}$).

Remarque 5.5 *a) Si les lois de X_n et de X ont un même support dénombrable discret S (en général $S \subset \mathbb{Z}$), alors la convergence en loi de X_n vers X équivaut à:*

$$\mathbb{P}(X_n = s) \text{ tend vers } \mathbb{P}(X = s) \quad \text{pour tout } s \in S\,.$$

b) Si X_n converge en loi vers X et si g est continue, alors $g \circ X_n$ converge en loi vers $g(X)$.

Proposition 5.6 *La convergence en loi équivaut à la convergence simple des transformées de Fourier:* $X_n \xrightarrow[n\to\infty]{\text{loi}} X \iff \Big(\mathbb{E}[e^{\sqrt{-1}\,t X_n}] \xrightarrow[n\to\infty]{} \mathbb{E}[e^{\sqrt{-1}\,t X}] \quad (\forall\, t \in \mathbb{R})\Big)$.

Proposition 5.7 *La convergence en loi d'une suite de v.a.r. $\{X_n \,|\, n \in \mathbb{N}^*\}$ vers une v.a.r. X équivaut à la convergence simple de F_n vers F en chaque point de continuité de F ; F_n désignant la fonction de répartition de X_n, et F celle de X.*

Exemples a) La loi $B(n,p_n)$ converge vers la loi $\mathcal{P}(\lambda)$ si $n \to \infty$ et $np_n \to \lambda > 0$.

b) La loi $H(N,n,p_N)$ converge vers la loi $B(n,p)$ si $N \to \infty$ et $p_N \to p \in]0,1[$.

c) Soient B une variable de Bernoulli ($\mathbb{P}(B=1) = \mathbb{P}(B=-1) = 1/2$), et pour tout $n \in \mathbb{N}$: $X_n = B = X'_{2n}$ et $X'_{2n+1} = -B$. Alors les 2 suites X et X' convergent en loi, mais pas la suite $X + X'$. Et la suite X' ne converge pas en probabilité.

Exercice n^o 5.8 Notons $\{X_n \,|\, n \in \mathbb{N}^*\}$ une suite de v.a.r. indépendantes gaussiennes standard. Étudier la convergence en loi de $\frac{nX_1+(n-1)X_2+\cdots+X_n}{n\sqrt{n}}$.

Exercice n^o 5.9 Soit $\{X_n \,|\, n \in \mathbb{N}\}$ une suite de v.a.r. de fonction de répartition commune F, telle que $\lim\limits_{x\to\infty} x\,(1-F(x)) = \lim\limits_{x\to\infty} x\,F(-x) = 0$. Posons $M_n := \max\{X_1,\ldots,X_n\}$, et $m_n := \min\{X_1,\ldots,X_n\}$. Montrer que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ convergent en probabilité vers 0.

Exercice n^o 5.10 Soit $\{X_j \,|\, j \in \mathbb{N}^*\}$ une suite de v.a.i.i.d. $\mathcal{N}(0,1)$. Soit $\{a_j \,|\, j \in \mathbb{N}\}$ une suite de réels tels que $a_j a_{j+1} = 0$ pour tout j et tels que $\sum\limits_j a_j^2 < \infty$. Soit $Y_n := \sum\limits_{j=1}^{n} a_{n-j} X_j$, pour tout $n \in \mathbb{N}^*$.

a) Étudier la convergence en loi de la suite (Y_n). b) Les v.a. Y_n et Y_{n+1} sont-elles indépendantes ? c) Étudier la convergence dans L^2 de la suite (Y_n).

Théorème 5.8 *(Loi des Grands Nombres) Soit $\{X_n \,|\, n \in \mathbb{N}^*\}$ une suite de v.a. indépendantes et intégrables, de même loi. Alors la suite des moyennes de Césaro $\dfrac{X_1+\cdots+X_n}{n}$ converge p.s. vers $\mathbb{E}(X_1)$.*

Remarque 5.9 0) La réciproque suivante de la loi des grands nombres est vraie :

Si $\{X_n \,|\, n \in \mathbb{N}^*\}$ est une suite de v.a.r indépendantes et de même loi telle que la suite des moyennes de Césaro $\frac{X_1+\cdots+X_n}{n}$ converge p.s., alors X_1 est intégrable.

1) Le théorème 5.8 ci-dessus constitue la loi forte des grands nombres. On appelle loi faible des grands nombres le résultat (strictement plus faible) qui énonce la convergence en probabilité des moyennes, au lieu de leur convergence presque sûre.

2) On montre que la convergence dans la loi forte des grands nombres a lieu également dans L^1.

Exercice n^o 5.11 Prouver la formulation suivante de la loi faible des grands nombres :

Si les v.a.r. X_n sont non-corrélées 2 à 2 et ont la même loi admettant un second moment, alors leurs moyennes de Césaro convergent dans L^2 (et donc en probabilité) vers $\mathbb{E}(X_1)$.

Exercice n^o 5.12 Soit $\{X_n \,|\, n \in \mathbb{N}^*\}$ une suite de v.a.r.i.i.d. telles que $\mathbb{E}(X_1^+) = \infty > \mathbb{E}(X_1^-)$. Montrer que p.s. $\frac{X_1+\cdots+X_n}{n} \longrightarrow +\infty$. (Utiliser $X_n^k := X_n \wedge k$).

Théorème 5.10 *(Théorème Central Limite) Soit* $\{X_n \,|\, n \in I\!\!N^*\}$ *une suite de v.a. à valeurs dans* $I\!\!R^d$*, indépendantes, et de même loi admettant un second moment. Alors* $\dfrac{X_1 + \cdots + X_n - nI\!\!E(X_1)}{\sqrt{n}}$ *converge en loi, vers la loi gaussienne* $\mathcal{N}(0, K_{X_1})$*, où* $K_{X_1} := I\!\!E({}^tX_1\, X_1) - {}^tI\!\!E(X_1)I\!\!E(X_1)$ *désigne la matrice de covariance de* X_1 *.*

Preuve L'énoncé étant clairement invariant par une translation sur la suite X, il suffit de considérer le cas où X_1 est centrée. Notons $\phi(v) := I\!\!E(e^{\sqrt{-1}\,v\cdot X_1})$ la transformée de Fourier de X_1, qui est de classe C^2, puisque X_1 est supposée de carré intégrable. La formule de Taylor-Young à l'origine s'écrit alors : $\phi(v) = 1 - \frac{1}{2}\, {}^tvK(X_1)v + o(\|v\|^2)$.
Par indépendance, la transformée de Fourier de $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ est égale à $\phi(v/\sqrt{n}\,)^n$, qui vaut $\Big[1 - \frac{{}^tvK(X_1)v + n\,o(\|v\|^2/n)}{2n}\Big]^n$, et donc qui converge vers $\exp\big[-\frac{1}{2}\, {}^tvK(X_1)v\,\big]$, du fait de la convergence de $(1+\frac{z}{n})^n$ vers e^z uniformément sur les parties bornées de $\mathbb{C}$.
En effet, si $|z| \leq A$, utilisant que $n^k \geq k!\, C_n^k$ et l'inégalité triangulaire nous avons :

$$\Big|e^z - \Big(1+\frac{z}{n}\Big)^n\Big| \leq \sum_{k=0}^{n}\Big(\frac{1}{k!} - \frac{C_n^k}{n^k}\Big)|z|^k + \sum_{k=n+1}^{\infty}\frac{|z|^k}{k!} \leq \sum_{k=0}^{n}\Big(\frac{1}{k!} - \frac{C_n^k}{n^k}\Big)A^k + \sum_{k=n+1}^{\infty}\frac{A^k}{k!}$$

$$= e^A - (1+\tfrac{A}{n})^n = e^A - e^{n\,\log(1+\frac{A}{n})} = e^A - e^{A-\mathcal{O}(1/n)} \longrightarrow 0\,.$$

Finalement, nous venons de montrer la convergence simple de la transformée de Fourier de $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ vers celle de $\mathcal{N}(0, K_{X_1})$, ce qui suffit d'après la proposition 5.6. $\diamond$

Nota Bene Le TCL évalue les fluctuations d'une somme de v.a.i.i.d. (de carré intégrable) autour du terme (déterministe) dominant, donné par la LGN ; en dimension 1 (pour simplifier) on peut en effet le réécrire sous forme d'égalité en loi :

$$X_1 + \cdots + X_n \overset{\text{loi}}{=} n\,I\!\!E(X_1) + \sqrt{n}\,\sigma(X_1)\,\mathcal{N}(0,1) + o(\sqrt{n}\,).$$

Exercice n^o 5.13 (Surlocation) Une agence de voyage dispose de 160 places à louer pour une destination donnée. Elle sait que les locations sont honorées par ses clients avec une probabilité fixe p. Elle vend $N = 160\alpha > 160$ places. Pour quelles valeurs de α la probabilité de ne pas louer trop de places vaut-elle $0,95$; $0,975$?

Exercice n^o 5.14 Le prix S_n d'une action au jour n est modélisé ainsi : $S_0 = s > 0$ est fixe, et $S_{n+1} = (1 + r + \sigma\varepsilon_{n+1})\, S_n$, où $r > 0$ est un taux fixe, $\sigma \in]0, 1+r[$ est une volatilité fixe, et $\{\varepsilon_n \,|\, n \in I\!\!N\}$ est une suite i.i.d. de loi de Bernoulli $\mathcal{B}(\pm 1, 1/2)$.
a) Étudier le comportement des suites $(\log S_n)/n$ et S_n .
b) Étudier le comportement de la suite $(\log S_n)/\sqrt{n}$.
c) Étudier le comportement de la suite $[(1+r)^2 - \sigma^2]^{(-\sqrt{n}\,/2)} \times S_n^{[1/\sqrt{n}\,]}$.

Exercice n^o 5.14 Notons $\{X_n, Y_n, Z_n \,|\, n \in I\!\!N^*\}$ une famille de v.a.i.i.d. de loi commune $\mathcal{B}(\pm 1, 1/2)$. Posons pour tout $n \in I\!\!N$:
$R_n := \sum_{k=1}^n X_k Y_k$, $S_n := \sum_{k=1}^n Y_k Z_k$, $T_n := \sum_{k=1}^n Z_k X_k$, puis $V_n := (R_n, S_n, T_n)$.

a) R_n, S_n, T_n sont-elles indépendantes 2 à 2 ? indépendantes ?
b) Quelle est la loi de R_n ? c) Calculer la transformée de Fourier de V_n .
d) Étudier la convergence de V_n/n , puis de $V_n/\sqrt{n}$.

Exercice n^o 5.15 3016 mathématiciens sont invités à un colloque ; en moyenne un sur 4 répondra favorablement, et les réponses sont supposées indépendantes les unes des autres. Combien les organisateurs doivent-ils prévoir de places, afin que la probabilité de ne pas en manquer soit $\geq 0,99$?

Exercice n^o 5.16 Une compagnie assure 10000 clients sur la vie, qui payent chacun une prime annuelle de A euros. On estime que chaque client a une probabilité de décès au cours d'une année égale à 6/1000, indépendamment les uns des autres. La prime de décès est de B euros.

a) Quelle est la loi du nombre annuel des décès ? Comment peut-on l'approcher ?

b) Si $B = 1000$, pour quels A la compagnie a-t-elle une probabilité $< 1/100$ d'être en déficit ? c) Si $A = 15$, pour quels B la compagnie a-t-elle une probabilité $> 0,7$ de faire un bénéfice annuel > 5000 euros ?

Exercice n^o 5.17 Soit $\{X_n \,|\, n \in \mathbb{N}\}$ une suite de v.a.i.i.d. de loi commune $\mathcal{B}(p)$. Posons $Y_n := X_n X_{n+1}$ et $S_n := (Y_1 + \cdots + Y_n)/n$, pour tout $n \in \mathbb{N}$.

Calculer la loi et l'espérance de Y_n , puis la covariance de Y_n et de Y_{n+k} , l'espérance de S_n^2 , et enfin montrer que S_n converge en probabilité vers p^2.

I. Espérance conditionnelle

Il s'agit ici d'introduire une notion essentielle, aussi bien pour définir les martingales que l'intégrale stochastique. Elle généralise la notion d'espérance (Définition 2.3), en tenant compte de la connaissance (conditionnement, probabilité conditionnelle : voir la section 1.7 du chapitre 0) d'une certaine information, représentée par une tribu (Définition 1.1). Le plus simple est de penser cette tribu comme engendrée par une certaine variable aléatoire X. Alors l'espérance conditionnelle (d'une variable aléatoire intégrable quelconque Y, donnée) par rapport à X dépendra de X, et sera ainsi elle-même une variable aléatoire, fonction déterministe de X.

6 Espérance conditionnelle discrète

Le cas le plus simple est celui l'espérance conditionnelle par rapport à une variable aléatoire X discrète, c'est-à-dire ne prenant (avec probabilité 1) qu'une quantité finie ou dénombrable de valeurs.

6.1 Cas d'une indicatrice d'événement : $X = 1_B$

Partons du cas particulier le plus élémentaire: lorsque la variable aléatoire discrète X ne prend que 2 valeurs, qu'on peut ramener (par une fonction affine évidente) à 0 et 1. Alors X est simplement la fonction indicatrice $X = 1_B$ de l'événement $B := \{X = 1\}$, et l'information que contient X se résume à celle de B, c'est-à-dire simplement au fait de savoir si le tirage ω tombe dans B ou dans son complémentaire $B^{\mathbf{c}}$. La tribu associée à cette information est la tribu engendrée par X, c'est-à-dire l'ensemble des événements de la forme $X^{-1}(A)$, pour A ensemble quelconque de valeurs que X peut prendre ; c'est aussi la plus petite tribu contenant $\{B\}$, c'est-à-dire: $\sigma(X) = \sigma(B) = \{\emptyset, \Omega, B, B^{\mathbf{c}}\}$.

Alors l'espérance conditionnelle par rapport à X (ou ici: par rapport à B) ne peut prendre que 2 valeurs, suivant que $\omega \in B$ ou que $\omega \in B^{\mathbf{c}}$. Et il est bien naturel que ces 2 valeurs possibles soient données par les probabilités conditionnelles de la définition 1.8: $\mathbb{P}(\cdot/B)$ et $\mathbb{P}(\cdot/B^{\mathbf{c}})$. Et puisque nous souhaitons aboutir à une espérance, cette considération conduit à la définition suivante.

Définition 6.2 *Soit V une v.a. intégrable définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ (cf la définition 2.3), et soit $B \in \mathcal{T}$. L'espérance conditionnelle de V par rapport à la variable aléatoire 1_B (ou bien : par rapport à la tribu $\sigma(B) \equiv \sigma(1_B)$) vaut :*

$$\mathbb{E}(V \mid 1_B) \equiv \mathbb{E}(V \mid \sigma(1_B)) \equiv \mathbb{E}(V \mid \sigma(B)) := \mathbb{E}(V/B)\, 1_B + \mathbb{E}(V/B^{\mathbf{c}})\, 1_{B^{\mathbf{c}}} .$$

Noter que si l'événement B est négligeable, alors $1_B = 0$ presque sûrement, de sorte qu'on conviendra que $\mathbb{E}(V/B)\, 1_B = 0$. De même pour le second terme si $B^{\mathbf{c}}$ est négligeable.

Cette définition donne bien pour l'espérance conditionnelle $\mathbb{E}(V \mid 1_B)$ une variable aléatoire ne prenant que deux valeurs, fonction déterministe de la variable aléatoire 1_B . On peut la réécrire : $\mathbb{E}(V \mid 1_B) = \mathbb{E}(V/B^{\mathbf{c}}) + [\mathbb{E}(V/B) - \mathbb{E}(V/B^{\mathbf{c}})]\, 1_B$.

Noter que si on intègre cette espérance conditionnelle, on obtient par la formule des probabilités totales (Proposition 1.9, écrite avec des indicatrices et des espérances, puis étendue à des combinaisons linéaires d'indicatrices et enfin à des v.a. intégrables quelconques) :

$$\mathbb{E}[\mathbb{E}(V \mid 1_B)] = \mathbb{E}(V).$$

6.3 Cas d'une v.a. prenant un nombre fini de valeurs

Plaçons nous maintenant dans le cas où la v.a. X prend (avec probabilité 1) ses valeurs dans une partie finie $\{x_1, \ldots, x_n\}$. La tribu engendrée par X (c'est-à-dire l'ensemble des événements de la forme $X^{-1}(A)$, pour A ensemble quelconque de valeurs que X peut prendre) est alors la plus petite tribu contenant les événements $\{X = x_1\}, \ldots, \{X = x_n\}$. Et il doit suffire de se placer sur chacun de ces événements élémentaires pour décider de la valeur de l'espérance conditionnelle par rapport à la variable aléatoire X .

Dans le cas précédent de la définition 6.2, on avait $n = 2$, $x_1 = 0$, $x_2 = 1$, et

$$\mathbb{E}(V\,|\,X) \equiv \mathbb{E}(V\,|\,\sigma(X)) = \mathbb{E}(V/\{X=x_1\})\,1_{\{X=x_1\}} + \mathbb{E}(V/\{X=x_2\})\,1_{\{X=x_2\}}.$$

La définition 6.2 est maintenant très naturellement généralisée comme suit.

Définition 6.4 *Soit V une v.a. intégrable définie sur un espace de probabilité $(\Omega,\mathcal{T},\mathbb{P})$, et soit X une v.a. sur $(\Omega,\mathcal{T})$, ne prenant qu'un nombre fini de valeurs $\{x_1,\ldots,x_n\}$. L'espérance conditionnelle par rapport à la variable aléatoire X (ou par rapport à la tribu finie $\sigma(X)$) vaut :*

$$\mathbb{E}(V\,|\,X) \equiv \mathbb{E}(V\,|\,\sigma(X)) := \sum_{j=1}^{n} \mathbb{E}(V/X=x_j)\,1_{\{X=x_j\}}\,.$$

Noter qu'il n'y a pas de raison de faire figurer une valeur x_j que la v.a. X ne prend pas (ou bien prend avec probabilité 0). Mais même dans ce cas, c'est-à-dire si $1_{\{X=x_j\}} = 0$ presque sûrement, on conviendra (comme en 6.1) que $\mathbb{E}(V/X=x_j)\,1_{\{X=x_j\}} = 0$, ce qui revient au même que de ne pas faire figurer x_j dans la liste des valeurs prises par X.

Comme dans la section 6.1 ci-dessus, si on intègre cette espérance conditionnelle, on obtient par la même formule des probabilités totales (Proposition 1.9) : $\mathbb{E}[\mathbb{E}(V\,|\,X)] = \mathbb{E}(V)$.

Nous allons maintenant largement généraliser cette égalité, en deux temps.

Par simple commodité, plaçons-nous dans le cas de valeurs x_j réelles, la variable V étant à valeurs dans $\mathbb{R}^d$. Alors le produit XV est une v.a. bien définie sur $(\Omega,\mathcal{T},\mathbb{P})$, et, utilisant successivement la définition 1.8, le corollaire 2.4, et la définition 6.4, nous avons :

$$\mathbb{E}(XV) = \mathbb{E}\Big[\sum_{j=1}^{n} 1_{\{X=x_j\}} x_j V\Big] = \sum_{j=1}^{n} x_j \mathbb{E}(V\,1_{\{X=x_j\}}) = \sum_{j=1}^{n} x_j \mathbb{E}(V/X=x_j)\,\mathbb{P}(X=x_j)$$

$$= \mathbb{E}\Big[\sum_{j=1}^{n} x_j \mathbb{E}(V/X=x_j) 1_{\{X=x_j\}}\Big] = \mathbb{E}\Big[\sum_{j=1}^{n} X\,\mathbb{E}(V/X=x_j) 1_{\{X=x_j\}}\Big] = \mathbb{E}\Big[X\,\mathbb{E}(V|X)\Big].$$

Considérons ensuite n'importe quelle fonction réelle f définie sur $\{x_1,\ldots,x_n\} = X(\Omega)$, et notons $f(X)$ pour la v.a.r. $f\circ X$. De façon très similaire à ce qui précède, nous avons :

$$\mathbb{E}[f(X)V] = \mathbb{E}\Big[\sum_{j=1}^{n} 1_{\{X=x_j\}} f(x_j) V\Big] = \sum_{j=1}^{n} f(x_j)\mathbb{E}(V 1_{\{X=x_j\}})$$

$$= \sum_{j=1}^{n} f(x_j)\mathbb{E}(V/X=x_j)\mathbb{P}(X=x_j) = \mathbb{E}\Big[\sum_{j=1}^{n} f(x_j)\mathbb{E}(V/X=x_j)\,1_{\{X=x_j\}}\Big]$$

$$= \mathbb{E}\Big[\sum_{j=1}^{n} f(X)\,\mathbb{E}(V/X=x_j)\,1_{\{X=x_j\}}\Big] = \mathbb{E}\Big[f(X)\,\mathbb{E}(V\,|\,X)\Big].$$

Nous avons ainsi démontré la très importante propriété suivante.

Proposition 6.5 *Soient V une v.a. intégrable vectorielle définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et X une v.a. sur $(\Omega, \mathcal{T})$, ne prenant qu'un nombre fini de valeurs. Alors pour toute fonction réelle f définie sur $X(\Omega)$ (par exemple l'application identique, ou bien la fonction constante égale à 1), nous avons*

$$\mathbb{E}[f(X)V] = \mathbb{E}\Big[f(X)\,\mathbb{E}(V\,|\,X)\Big].$$

Remarque 6.6 Si une tribu finie $\mathcal{F}$ est donnée sans référence à une variable aléatoire X, elle est tout de même nécessairement engendrée par une partition : $\Omega = \bigsqcup_{j=1}^{n} A_j$ p.s., avec $\{A_1, \ldots, A_n\}$ engendrant la tribu $\mathcal{F}$ (plus précisément, tout événement de $\mathcal{F}$ est alors réunion de certains des événements élémentaires A_j ; pour vérifier cela, il suffit de considérer les <u>atomes</u> de $\mathcal{F}$: $(\forall\ \omega \in \Omega)\ A(\omega) := \bigcap\{E \in \mathcal{F}\,|\,\omega \in E\}$: ils constituent automatiquement la partition cherchée $\{A_1, \ldots, A_n\}$).

Ce cas est en fait identique au cas envisagé ci-dessus : il suffit de considérer alors (par exemple) la variable aléatoire $X := \sum_{j=1}^{n} j\, 1_{A_j}$ valant $x_j = j$ sur l'atome A_j, de sorte que $A_j = \{X = x_j\}$. La définition 6.4 prend ainsi la forme :

$$\mathbb{E}(V\,|\,\mathcal{F}) \;=\; \sum_{j=1}^{n} \mathbb{E}(V/A_j)\, 1_{A_j}\,.$$

6.7 Cas général d'une v.a. discrète

Commençons par étendre directement le cas précédent de la section 6.3 au cas où la v.a. X prend (avec probabilité 1) ses valeurs dans une partie dénombrable $\{x_j\,|\,j \in \mathbb{N}^*\} = \{x_1, \ldots, x_n, \ldots\}$. La <u>tribu engendrée</u> par X est alors la plus petite tribu contenant la suite des événements $\{X = x_1\}, \ldots, \{X = x_n\}, \ldots$.

Définition 6.8 *Soit V une v.a. intégrable définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et soit X une v.a. sur $(\Omega, \mathcal{T})$, prenant une suite de valeurs $\{x_1, \ldots, x_n, \ldots\}$. L'espérance conditionnelle par rapport à la variable aléatoire X (ou par rapport à la tribu $\sigma(X)$) vaut*

$$\mathbb{E}(V\,|\,X) \equiv \mathbb{E}(V\,|\,\sigma(X)) \;:=\; \sum_{j\geq 1} \mathbb{E}(V/X = x_j)\, 1_{\{X=x_j\}}\,.$$

L'intégrabilité de V suffit à garantir dans cette définition (qui étend la définition 6.4) l'intégrabilité de la série. La proposition 6.5 s'étend aussi (avec la même preuve).

Proposition 6.9 *Soient V une v.a. vectorielle intégrable définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et X une v.a. discrète sur $(\Omega, \mathcal{T})$. Alors pour toute fonction réelle f bornée définie sur $X(\Omega)$ (par exemple l'application identique si X est réelle bornée, ou bien la fonction constante égale à 1), nous avons*

$$\mathbb{E}[f(X)V] = \mathbb{E}\Big[f(X)\,\mathbb{E}(V\,|\,X)\Big].$$

Notons que l'hypothèse de bornitude, automatique dans le cas de $X(\Omega)$ fini, n'est là que pour garantir l'intégrabilité, et peut être un peu affaiblie.

Observons que si la fonction f est bijective, alors $\{f(X) = f(x_j)\} = \{X = x_j\}$ pour tout j, de sorte que la variable $f(X) \equiv f \circ X$ contient exactement la même information que la variable X. C'est-à-dire que $f(X)$ et X engendrent la même tribu.

Au contraire, si f n'est pas bijective, c'est-à-dire dans le cas général, alors ceci devient faux ! Par exemple, avec $f(x) = x^2$, savoir que $X^2 = 1$ ne dit pas si $X = 1$ ou si $X = -1$...
Mais on a tout de même pour tout $y \in \mathbb{R}$: $\{f(X) = y\} = \bigcup\limits_{\{j\,|\,f(x_j)=y\}} \{X = x_j\} \in \sigma(X)$.
Ce qui signifie qu'on a l'inclusion immédiate suivante entre tribus : $\sigma(f \circ X) \subset \sigma(X)$, pour toute fonction f ; ce qu'on interprète en disant que l'information que contient $f(X)$ est contenue dans celle que contient X. Il se trouve que la réciproque suivante est vraie.

Proposition 6.10 *Soient X, Y deux v.a. définies sur un espace probabilisé $(\Omega, \mathcal{T})$. Alors on a l'inclusion des tribus engendrées : $\sigma(Y) \subset \sigma(X)$ si et seulement si il existe une fonction (déterministe) f définie sur $X(\Omega)$ telle que $Y = f \circ X$.*

Preuve Il reste à établir la réciproque. Considérons pour simplifier de nouveau le cas où X est discrète. Par hypothèse, pour toute valeur y prise par Y, il existe des valeurs $x_{j_1}, \dots x_{j_k}, \dots$ prises par X telles que $\{Y = y\} = \bigsqcup\limits_i \{X = x_{j_i}\}$. Et il est clair que les listes des valeurs x_{j_i} correspondant à des valeurs y distinctes sont disjointes deux à deux. On peut donc définir f sans ambigüité en posant : $f(x_{j_1}) = \dots = f(x_{j_k}) = \dots := y$, pour toutes les valeurs y prises par Y (qui sont nécessairement en quantité au plus égale au nombre des valeurs prises par X). Il n'y a plus qu'à constater qu'on a bien ainsi $f(X) = Y$ (avec probabilité 1). $\diamond$

Le sens direct de cet énoncé s'applique en particulier à la v.a. $\mathbb{E}(V\,|\,X)$. Cela et la proposition 6.9 montrent le sens direct de l'importante caractérisation suivante.

Proposition 6.11 *(caractérisation de l'espérance conditionnelle) Soient V, W deux v.a. vectorielles intégrables définies sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et X une v.a. discrète sur $(\Omega, \mathcal{T})$. Alors $W = \mathbb{E}(V\,|\,X)$ p.s. si et seulement si nous avons*

$$(i)\ \sigma(W) \subset \sigma(X) \quad \text{et} \qquad (ii)\ \mathbb{E}[V 1_A] = \mathbb{E}[W 1_A] \ \text{ pour tout } A \in \sigma(X).$$

Preuve Il reste à établir la réciproque. L'hypothèse (i) et la proposition 6.10 permettent d'écrire $W = g(X)$ et $\mathbb{E}(V \mid X) = h(X)$, pour deux fonctions déterministes g, h. Rappelons que tout $A \in \sigma(X)$ s'écrit $X^{-1}(B) \equiv \{X \in B\}$, pour $B = X(A) \subset X(\Omega)$, de sorte que $1_A = 1_{X^{-1}(B)} = 1_B(X)$. Appliquant la proposition 6.9 avec $f = 1_B$, nous obtenons : $\mathbb{E}[h(X)1_A] = \mathbb{E}[V 1_A] = \mathbb{E}[g(X)1_A]$ par l'hypothèse (ii), ceci pour tout $A \in \sigma(X)$. Donc en posant $\varphi := h - g$: $\mathbb{E}[\varphi(X)1_A] = 0$. Choisissant $A = \{\varphi(X) \geq \varepsilon\} \in \sigma(X)$, cela donne $0 \geq \varepsilon\, \mathbb{P}(\varphi(X) \geq \varepsilon)$, et donc $\mathbb{P}(\varphi(X) \geq \varepsilon) = 0$, id est $\varphi(X) \leq \varepsilon$ p.s., pour tout $\varepsilon > 0$. D'où $\varphi(X) \leq 0$ p.s.. Changeant φ en $-\varphi$, cela prouve que $\varphi(X) = 0$ p.s., id est $W = \mathbb{E}(V \mid X)$ p.s.. $\diamond$

Exercice n° 6.1 Soient X, Y deux v.a.r. indépendantes, et soit $S := X + Y$ leur somme. Calculer la loi conditionnelle $\mathbb{E}[f(X) \mid S]$, puis $\mathbb{E}[X - Y \mid S]$, dans chacun des cas suivant.

(i) $\mathbb{P}_X = \mathcal{B}(m, p)$ et $\mathbb{P}_Y = \mathcal{B}(n, p)$; (ii) $\mathbb{P}_X = \mathcal{P}(\lambda)$ et $\mathbb{P}_Y = \mathcal{P}(\mu)$;

(iii) $\mathbb{P}_X = \mathcal{G}(p)$ et $\mathbb{P}_Y = \mathcal{G}(q)$. Dans ce dernier cas, calculer aussi la loi conditionnelle de X sachant $\min\{X, Y\}$ et la loi conditionnelle de $\min\{X, Y\}$ sachant X.

7 Autres propriétés de l'espérance conditionnelle

L'espérance conditionnelle jouit des mêmes propriétés élémentaires que l'espérance.

Proposition 7.1 *L'espérance conditionnelle est linéaire : pour des v.a. $V_1, \ldots, V_n$ vectorielles intégrables définies sur le même espace $(\Omega, \mathcal{T}, \mathbb{P})$ que la v.a.r. discrète X, et pour tous réels $\lambda_1, \ldots, \lambda_n$, nous avons*

$$\mathbb{E}\left[\sum_{j=1}^{n} \lambda_j V_j \,\middle|\, X\right] = \sum_{j=1}^{n} \lambda_j\, \mathbb{E}(V_j \mid X).$$

Par ailleurs, nous avons les règles de calcul suivantes.

Proposition 7.2 *Soient V une v.a. vectorielle intégrable définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et X, Y des v.a. discrètes sur $(\Omega, \mathcal{T})$. Alors*

(o) *$V \geq 0$ p.s. $\Rightarrow \mathbb{E}(V \mid X) \geq 0$ p.s. $\big($d'où $V \geq V'$ p.s. $\Rightarrow \mathbb{E}(V \mid X) \geq \mathbb{E}(V' \mid X)$ p.s.$\big)$;*

(i) *$\mathbb{E}[\mathbb{E}(V \mid X)] = \mathbb{E}(V)$;*

(ii) *Si $\sigma(V) \subset \sigma(X)$, on a $\mathbb{E}(V \mid X) = V$ p.s. ;*

(iii) *Si V est indépendante de X, on a $\mathbb{E}(V \mid X) = \mathbb{E}(V)$ p.s. ;*

(iv) *$\mathbb{E}[f(X)V \mid X] = f(X)\, \mathbb{E}(V \mid X)$ p.s., pour toute fonction réelle f bornée sur $X(\Omega)$;*

(v) *Si $\sigma(X) \subset \sigma(Y)$, on a $\mathbb{E}\Big[\mathbb{E}(V \mid Y) \,\Big|\, X\Big] = \mathbb{E}(V \mid X) = \mathbb{E}\Big[\mathbb{E}(V \mid X) \,\Big|\, Y\Big]$ p.s..*

Preuve (i) a déjà été mentionnée (et découle de la proposition 6.9). Si on a $\sigma(V) \subset \sigma(X)$, alors la proposition 6.10 permet d'écrire $V = f(X)$, et la définition 6.8 entraîne :

$$\mathbb{E}(V \,|\, X) = \sum_j \mathbb{E}(f(X)/X = x_j) 1_{\{X=x_j\}} = \sum_j f(x_j) 1_{\{X=x_j\}} = f(X) = V\,.$$

Et si V est indépendante de X, elle l'est de tout événement $\{X = x_j\}$, de sorte que :

$$\mathbb{E}(V \,|\, X) = \sum_j \mathbb{E}(V/X = x_j) 1_{\{X=x_j\}} = \sum_j \mathbb{E}(V) 1_{\{X=x_j\}} = \mathbb{E}(V) \text{ p.s..}$$

Puis (iv) découle de la caractérisation (proposition 6.11). En effet, la tribu engendrée par $f(X)\,\mathbb{E}(V \,|\, X)$ est clairement une sous-tribu de $\sigma(X)$, et pour tout $A \in \sigma(X)$, écrivant $1_A = g(X)$ et posant $h := fg$, nous avons :

$$\begin{aligned}\mathbb{E}\big[f(X)\mathbb{E}(V \,|\, X)\, 1_A\big] &= \mathbb{E}\big[h(X)\, \mathbb{E}(V \,|\, X)\big] = \mathbb{E}[h(X)\, V] = \mathbb{E}\big[(f(X)\, V) 1_A\big] \\ &= \mathbb{E}\big[\mathbb{E}[f(X)\, V \,|\, X]\, 1_A\big].\end{aligned}$$

La seconde égalité de (v) découle directement de (ii). Enfin, la première égalité de (v) découle de la caractérisation (proposition 6.11) : en effet, $A \in \sigma(X) \Rightarrow A \in \sigma(Y)$, et donc

$$\mathbb{E}\left(\mathbb{E}\big[\mathbb{E}(V \,|\, Y) \,\big|\, X\big] 1_A\right) = \mathbb{E}[\mathbb{E}(V \,|\, Y)\, 1_A] = \mathbb{E}[V\, 1_A] = \mathbb{E}\big(\mathbb{E}(V \,|\, X)\, 1_A\big). \;\diamond$$

8 Projection orthogonale dans L^2

Nous considérons maintenant le cas général d'une sous-tribu $\mathcal{F}$ de la tribu de base $\mathcal{T}$. Non seulement elle pourra être engendrée par une (ou des) variable(s) aléatoire(s) pouvant désormais prendre un continuum de valeurs, mais nous allons en outre directement nous passer de mentionner une telle façon de la représenter : $\mathcal{F}$ sera seulement une sous-tribu de $\mathcal{T}$, sans plus nous soucier même de savoir s'il existe bien une v.a. qui l'engendre.

D'autre part nous allons exiger que les variables aléatoires intervenant dans cette section soient de carré intégrable , id est “dans L^2” (et non plus seulement intégrables, dans L^1). Par simple commodité (car ce n'est pas essentiel), nous ne considérerons que des variables aléatoires réelles.

Ce cas général reprend bien entendu l'essentiel du cas discret précédent (sections 6.7 et 7), en le généralisant. Les preuves, quand il y en aura, se limiteront au cas précédent de sous-tribus engendrées par une v.a. discrète.

Théorème 8.1 *Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et une sous-tribu $\mathcal{F}$ de $\mathcal{T}$ (id est une tribu $\mathcal{F}$ sur Ω telle que $\mathcal{F} \subset \mathcal{T}$). Soit $V \in L^2(\Omega, \mathcal{T}, \mathbb{P})$.*

Alors il existe une unique variable aléatoire, définie p.s. et notée $\mathbb{E}(V \,|\, \mathcal{F})$, telle que :

- (i) $\mathbb{E}(V\,|\,\mathcal{F}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$;
- (ii) $V - \mathbb{E}(V\,|\,\mathcal{F}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})^{\perp}$.

En outre, nous avons le théorème de Pythagore :

$$\mathbb{E}(V^2) = \mathbb{E}\Big[\mathbb{E}(V\,|\,\mathcal{F})^2\Big] + \mathbb{E}\Big[\big[V - \mathbb{E}(V\,|\,\mathcal{F})\big]^2\Big],$$

ainsi que la propriété de minimalité suivante :

$$\mathbb{E}\Big[\big[V - \mathbb{E}(V\,|\,\mathcal{F})\big]^2\Big] = \min\Big\{\mathbb{E}\big[(V-Y)^2\big]\,\Big|\, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})\Big\}.$$

Noter l'analogie rigoureuse avec le théorème de projection orthogonale dans l'espace vectoriel euclidien $\mathbb{R}^d$; la seule différence est qu'ici la dimension est infinie. $L^2(\Omega, \mathcal{T}, \mathbb{P})$ est le gros espace vectoriel "euclidien" (on dit en fait : "hilbertien" lorsqu'on est en dimension infinie) ambiant dans lequel tout a lieu, et $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est le sous-espace vectoriel sur lequel on projette orthogonalement (le vecteur V).

Noter aussi que le sens de (i) est double : la variable aléatoire projetée $\mathbb{E}(V\,|\,\mathcal{F})$ est $\mathcal{F}$-mesurable, et de carré intégrable.

La condition d'orthogonalité (ii) a le même sens qu'en dimension finie, en n'oubliant toutefois pas la définition du produit scalaire de $L^2(\Omega, \mathcal{T}, \mathbb{P})$: $(V, Z) \mapsto \mathbb{E}[VZ]$. De sorte que (ii) signifie précisément : pour toute $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}\Big[\big(V - \mathbb{E}(V\,|\,\mathcal{F})\big) \times Y\Big] = 0$, ou encore :

$$\mathbb{E}\big[VY\big] = \mathbb{E}\big[\mathbb{E}(V\,|\,\mathcal{F}) \times Y\big], \qquad \text{pour toute } Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}). \tag{1}$$

Définition 8.2 *La variable aléatoire projetée $\mathbb{E}(V\,|\,\mathcal{F})$ du théorème 8.1 est appelée <u>espérance conditionnelle par rapport à la tribu $\mathcal{F}$</u> de la variable aléatoire $V \in L^2(\Omega, \mathcal{T}, \mathbb{P})$.*

<u>Preuve</u> (du théorème 8.1 de projection orthogonale, dans le cas $\mathcal{F} = \sigma(Z)$ avec Z discrète) Soit comme dans la définition 6.8 :

$$\mathbb{E}(V\,|\,\mathcal{F}) := \sum_j \mathbb{E}(V/Z = z_j)\, 1_{\{Z=z_j\}},$$

qui par définition est bien $\mathcal{F}$-mesurable. Et comme (par disjonction des $\{Z = z_j\}$) on a

$$\mathbb{E}\Big[\mathbb{E}(V\,|\,\mathcal{F})^2\Big] = \mathbb{E}\Big[\sum_j \mathbb{E}(V/Z = z_j)^2 1_{\{Z=z_j\}}\Big] \le \sum_j \mathbb{E}(V^2/Z = z_j)\mathbb{P}(Z = z_j) = \mathbb{E}[V^2],$$

on est assuré que $\mathbb{E}(V\,|\,\mathcal{F}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Soit ensuite $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors selon les propositions 6.10 et 6.9, Y s'écrit $Y = f(Z)$ et nous avons :

$$\mathbb{E}[VY] = \mathbb{E}\big[V\, f(Z)\big] = \mathbb{E}\big[\mathbb{E}(V|\mathcal{F})\, f(Z)\big] = \mathbb{E}\big[\mathbb{E}(V|\mathcal{F})\, Y\big].$$

Ceci montre la propriété d'orthogonalité (ii), sous sa forme équivalente (1).
Ensuite, si W était une autre solution, on aurait (par soustraction) à la fois $W - \mathbb{E}(V|\mathcal{F}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et $W - \mathbb{E}(V|\mathcal{F}) \in L^2(\Omega, \mathcal{F}, \mathbb{P})^\perp$, d'où $\mathbb{E}\Big[\big[W - \mathbb{E}(V|\mathcal{F})\big]^2\Big] = 0$, ce qui implique l'unicité : $W = \mathbb{E}(V|\mathcal{F})$ p.s..
Pour le théorème de Pythagore :

$$\mathbb{E}[V^2] = E\Big[\big[(V - \mathbb{E}(V|\mathcal{F})) + \mathbb{E}(V|\mathcal{F})\big]^2\Big] = E\Big[\big[V - \mathbb{E}(V|\mathcal{F})\big]^2\Big] + E\Big[\mathbb{E}(V|\mathcal{F})^2\Big],$$

puisque le terme rectangle s'annule, précisément selon (i) et l'orthogonalité (ii).
Enfin, pour la même Y que ci-dessus, nous avons dans le même esprit :

$$\mathbb{E}\big[(V - Y)^2\big] = E\Big[\big[(V - \mathbb{E}(V|\mathcal{F})) + (\mathbb{E}(V|\mathcal{F}) - Y)\big]^2\Big]$$

$$= E\Big[\big[V - \mathbb{E}(V|\mathcal{F})\big]^2\Big] + E\Big[\big[\mathbb{E}(V|\mathcal{F}) - Y\big]^2\Big] \geq E\Big[\big[V - \mathbb{E}(V|\mathcal{F})\big]^2\Big]. \ \diamond$$

Les propriétés vues aux sections 6.7 et 7 demeurent valables : la proposition 6.9 est remplacée par (1) ; la caractérisation (proposition 6.11) est maintenant $(i), (ii)$ du théorème 8.1, (ii) pouvant être remplacé par (1), ou bien encore par :

$$\mathbb{E}\Big[V 1_A\Big] = \mathbb{E}\Big[\mathbb{E}(V \,|\, \mathcal{F}) \times 1_A\Big], \qquad \text{pour tout } \ A \in \mathcal{F}\,. \tag{2}$$

$\big($Justification : si on a (2), on peut par linéarité remplacer 1_A par toute variable en escalier $\mathcal{F}$-mesurable, puis par limite croissante, par toute variable $\mathcal{F}$-mesurable positive bornée, et enfin par différence puis densité, par toute $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.$\big)$

De même pour les propositions 7.1 et 7.2, qui deviennent :

Proposition 8.3 (oo) *L'espérance conditionnelle est linéaire :*

$$\mathbb{E}\bigg[\sum_{j=1}^n \lambda_j V_j \,\bigg|\, \mathcal{F}\bigg] = \sum_{j=1}^n \lambda_j \, \mathbb{E}(V_j \,|\, \mathcal{F})\,, \quad \forall\, V_1, \ldots, V_n \in L^2(\Omega, \mathcal{T}, \mathbb{P}), \ \forall\, \lambda_1, \ldots, \lambda_n \in \mathbb{R}.$$

Soient $V \in L^2(\Omega, \mathcal{T}, \mathbb{P})$, et $\mathcal{F}, \mathcal{G}$ deux sous-tribus de $\mathcal{T}$. Alors
(o) $V \geq 0$ *p.s.* $\Rightarrow \mathbb{E}(V \,|\, X) \geq 0$ *p.s.* $\big($*d'où* $V \geq V'$ *p.s.* $\Rightarrow \mathbb{E}(V \,|\, X) \geq \mathbb{E}(V' | \, X)$ *p.s.*$\big)$;
(i) $\mathbb{E}\big[\mathbb{E}(V \,|\, \mathcal{F})\big] = \mathbb{E}(V)$;
(ii) *Si* $V \in L^2(\Omega, \mathcal{F}, \mathbb{P})$*, on a* $\mathbb{E}(V \,|\, \mathcal{F}) = V$ *p.s.* ;
(iii) *Si* V *est indépendante de* $\mathcal{F}$*, on a* $\mathbb{E}(V \,|\, \mathcal{F}) = \mathbb{E}(V)$ *p.s.* ;
(iv) $\mathbb{E}[ZV \,|\, \mathcal{F}] = Z\, \mathbb{E}(V|\mathcal{F})$ *p.s., pour toute variable* Z $\mathcal{F}$*-mesurable bornée* ;
(v) *Si* $\mathcal{G} \subset \mathcal{F} \subset \mathcal{T}$*, on a* $\mathbb{E}\Big[\mathbb{E}(V \,|\, \mathcal{F}) \,\Big|\, \mathcal{G}\Big] = \mathbb{E}(V \,|\, \mathcal{G}) = \mathbb{E}\Big[\mathbb{E}(V \,|\, \mathcal{G}) \,\Big|\, \mathcal{F}\Big]$ *p.s..*

Remarquons qu'en fait, (i) et (ii) sont des cas particuliers de (v) : on obtient (i) en prenant $\mathcal{G} = \{\emptyset, \Omega\}$, car alors on a trivialement $\mathbb{E}(V|\mathcal{G}) = \mathbb{E}(V)$; et (i) en prenant $\mathcal{G} = \mathcal{F}$.

Exemple Soit $(Y, X_1, \ldots, X_n)$ un vecteur gaussien, de matrice de covariance K. Alors

$$\mathbb{E}(Y \mid X_1, \ldots, X_n) = a_0 + \sum_{j=1}^{n} a_j X_j \,.$$

Pour établir cette formule, commençons par trouver la valeur des coefficients a_j, et tout d'abord de a_0 en l'intégrant : on aussitôt $\mathbb{E}(Y) = a_0 + \sum_{j=1}^{n} a_j \mathbb{E}(X_j)$. Il suffit donc de poser $a_0 = \mathbb{E}(Y) - \sum_{j=1}^{n} a_j \mathbb{E}(X_j)$, et (par différence) de traiter le cas de $(Y, X_1, \ldots, X_n)$ centré (pour lequel $a_0 = 0$), sur lequel nous nous concentrons maintenant.
On doit avoir pour chaque $k \in \{1, \ldots, n\}$:

$$K(0,k) = \mathbb{E}(Y X_k) = \mathbb{E}\Big(\sum_{j=1}^{n} a_j X_j X_k\Big) = \sum_{j=1}^{n} a_j K(j,k)\,.$$

Donc matriciellement, notant $a := (a_1, \ldots, a_n)$ et $K = \begin{pmatrix} \sigma^2 & V \\ {}^tV & K_X \end{pmatrix}$, où σ est l'écart-type de Y et K_X la matrice de covariance de $X := (X_1, \ldots, X_n)$, nous avons : $V = a \times K_X$.
Si K_X n'est pas inversible : il existe un vecteur (déterministe) $v = (v_1, \ldots, v_n) \neq 0$ dans le noyau de K_X, de sorte que $\mathbb{E}[|X\,{}^tv|^2] = \mathbb{E}[v\,{}^tXX\,{}^tv] = v\,K_X\,{}^tv = 0$, et donc p.s. : $0 = X\,{}^tv = \sum_{j=1}^{n} v_j X_j$: le vecteur aléatoire $X = (X_1, \ldots, X_n)$ est p.s. cantonné dans l'hyperplan déterministe $v^{\perp}$. L'un au moins des v_j doit être non nul ; quitte à réordonner les X_j, nous pouvons supposer $v_n \neq 0$, de sorte que $X_n = \sum_{j=1}^{n-1} v'_j X_j$. Cela entraîne aussitôt que $\mathbb{E}(Y \mid X_1, \ldots, X_n) = \mathbb{E}(Y \mid X_1, \ldots, X_{n-1})$. Donc par récurrence sur n, nous pouvons diminuer n jusqu'à avoir K_X inversible, ou alors $n = 0$, auquel cas $a = 0$ convient trivialement. Nous sommes ainsi ramenés au cas de K_X inversible.
Or si K_X est inversible, nous avons simplement $a = V\,K_X^{-1}$.
Il reste seulement à vérifier que réciproquement, le vecteur a ainsi obtenu convient bien.
Or pour chaque $k \in \{1, \ldots, n\}$:

$$\mathbb{E}[(Y - a\,{}^tX)X_k] = K(0,k) - \sum_{j=1}^{n} a_j K(j,k) = V_k - (V K_X^{-1} K_X)_k = 0\,.$$

C'est dire que la matrice de covariance du vecteur gaussien $(Y - a\,{}^tX, X_1, \ldots, X_n)$ est $\begin{pmatrix} \sigma^2 & 0 \\ 0 & K_X \end{pmatrix}$, ce qui implique que ses coordonnées $Y - a\,{}^tX$ et $(X_1, \ldots, X_n)$ sont indépendantes (cela se déduit par exemple du corollaire 3.18 et de la proposition 4.2 : la matrice de covariance détermine la loi d'un vecteur gaussien centré, et elle a cette forme en cas d'indépendance, de sorte que la réciproque est vraie). Donc $\mathbb{E}(Y - a^t X | X_1, \ldots, X_n) = 0$, ou bien :

$$\mathbb{E}(Y \,|\, X_1, \ldots, X_n) = \mathbb{E}(a^t X \,|\, X_1, \ldots, X_n) = a^t X = \sum_{j=1}^{n} a_j X_j \, . \;\diamond$$

Exercice n° 6.2 Soit $(X, Y) \in \mathbb{R}^k \times \mathbb{R}^\ell$ une variable aléatoire admettant une densité $h = h(x, y)$. Montrer que la loi conditionnelle de X sachant Y admet la densité $\frac{h(\cdot, Y)}{\int h(x, Y)\, dx}$. (Calculer $\mathbb{E}\big[f(X) \,\big|\, Y\big]$, pour toute fonction test f sur $\mathbb{R}^k$).

9 Extension à L^1

Ayant fait le détour par L^2 dans la section précédente, nous pouvons finalement revenir aux variables seulement intégrables, tout en conservant la notion d'espérance conditionnelle par rapport à une sous-tribu quelconque. (Cela se fait par densité de L^2 dans L^1 et passage à la limite croissante pour des variables intégrables positives, puis par différence pour des variables réelles intégrables quelconques, et enfin, pour des variables vectorielles intégrables, coordonnée par coordonnée.) On perd ainsi en route la notion de projection orthogonale, mais il reste l'essentiel.

Théorème 9.1 *Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et une sous-tribu $\mathcal{F}$ de $\mathcal{T}$. Soit $V \in L^1(\Omega, \mathcal{T}, \mathbb{P})$. Il existe une unique variable aléatoire, définie p.s., notée $\mathbb{E}(V \,|\, \mathcal{F})$, et et appelée l'espérance conditionnelle de V par rapport à $\mathcal{F}$, telle que :*

– *(i)* $\mathbb{E}(V \,|\, \mathcal{F}) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$;

– *(ii)* $\mathbb{E}[V 1_A] = \mathbb{E}\big[\mathbb{E}(V \,|\, \mathcal{F}) \times 1_A\big]$, *pour tout* $A \in \mathcal{F}$.

Proposition 9.2 *(Inégalité de convexité de Jensen) Soient $V \in L^1(\Omega, \mathcal{T}, \mathbb{P})$, et $\mathcal{F}$ une sous-tribu de $\mathcal{T}$. Pour toute fonction convexe réelle minorée φ , on a*

$$\mathbb{E}\big[\varphi(V) \,|\, \mathcal{F}\big] \;\geq\; \varphi\big(\mathbb{E}[V \,|\, \mathcal{F}]\big).$$

Exemple : pour tout $p \geq 1$, on a $\quad \mathbb{E}\Big[\|V\|^p \,\Big|\, \mathcal{F}\Big] \geq \Big\| \mathbb{E}[V \,|\, \mathcal{F}] \Big\|^p .$

Enfin la proposition 8.3 reste valable en remplaçant partout L^2 par L^1.

II. Martingales à temps discret

Il s'agit d'une notion cruciale, modélisant un jeu équilibré, et indispensable à la notion d'intégrale stochastique (discrète), qui sera introduite et étudiée dans ce même chapitre.

10 Surmartingales, sousmartingales

Donnons-nous un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration, i.e. d'une suite croissante $\{\mathcal{F}_n \,|\, n \in \mathbb{N}\}$ de sous-tribus de $\mathcal{F}$. $\mathcal{F}_n$ représente l'information connue à l'instant n. Dans la pratique, on pourra souvent se limiter à un "horizon" fini T, c'est-à-dire que les temps n considérés pourront ne varier qu'entre les instants 0 et T. Cela rentre toutefois sans difficulté dans le cas $n \in \mathbb{N}$, simplement en remplaçant n par $n \wedge T := \min\{n, T\}$.

Définition 10.1 *Une suite $\{X_n \,|\, n \in \mathbb{N}\}$ de v.a.r. intégrables est une surmartingale lorsque* (i) *elle est adapté, i.e. X_n est $\mathcal{F}_n$-mesurable, pour tout $n \in \mathbb{N}$;*

(ii) $\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \leq X_n$ *p.s. pour tout $n \in \mathbb{N}$.*

La suite (X_n) est une sousmartingale lorsque $(-X_n)$ est une surmartingale, et est une martingale lorsqu'elle est à la fois une surmartingale et une sousmartingale.

Remarque : Si (X_n) est une surmartingale, la suite de ses espérances est décroissante !

Exemples : 1) Si X est une v.a.r. intégrable, $X_n := \mathbb{E}(X \,|\, \mathcal{F}_n)$ définit une martingale.

2) Soit $\{Y_n \,|\, n \in \mathbb{N}\}$ une suite de v.a.r. intégrables indépendantes, et soient $X_n := \sum_{j=0}^{n} Y_j$, puis $\mathcal{F}_0 := \{\emptyset, \Omega\}$, et $\mathcal{F}_n := \sigma\{Y_j \,|\, 1 \leq j \leq n\}$. Alors (X_n) est une martingale ssi les Y_j sont centrées (et donc en particulier les marches aléatoires centrées sont à la fois des martingales et des chaînes de Markov), (X_n) est une surmartingale ssi les Y_j ont toutes une espérance ≤ 0, et c'est une sousmartingale ssi les Y_j ont toutes une espérance ≥ 0.

Exercice n^o 10.0 Si (X_n) est une sousmartingale, alors $(X_n - a)^+$ est une sousmartingale. Si (X_n) est une surmartingale, alors $(\min\{X_n\,, a\})$ est une surmartingale (pour tout a réel).

Exercice n^o 10.1 a) Si (X_n) est une surmartingale, alors on a $\mathbb{E}(X_m \,|\, \mathcal{F}_n) \leq X_n$ p.s. pour tous $m > n$ dans $\mathbb{N}$, et de plus (X_n) est une surmartingale dans sa propre filtration.

b) L'ensemble des sousmartingales forme un cône convexe, stable par $(X_n, Y_n) \mapsto X_n \vee Y_n$.

c) Si (X_n) est une sousmartingale et si ϕ est une fonction convexe croissante de $\mathbb{R}$ dans $\mathbb{R}$ telle que chaque $\phi \circ X_n$ est intégrable, alors $(\phi \circ X_n)$ est une sousmartingale. Par exemple, pour toute constante réelle a, $(X_n - a)^+$ et $\max\{X_n, a\}$ sont des sousmartingales.

d) Si (X_n) est une martingale et si ϕ est une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$ telle que chaque $\phi \circ X_n$ est intégrable, alors $(\phi \circ X_n)$ est une sousmartingale.

Exercice n^o 10.2 Notons (Z_n) un processus de Galton-Watson, id est un processus à valeurs dans $\mathbb{N}$ tel que $Z_{n+1} = \sum_{j=1}^{Z_n} V_{n+1}^j$, où $\{V_n^j \,|\, n, j \in \mathbb{N}^*\}$ constitue une suite de v.a.i.i.d. de loi donnée sur $\mathbb{N}$ (la loi dite "de reproduction"). Soit m la moyenne de la loi de reproduction. Montrer que (Z_n) est une surmartingale lorsque $m \leq 1$, une sousmartingale lorsque $m \geq 1$, et une martingale lorsque $m = 1$.

Exercice n^o 10.3 Soient $\{Y_n \,|\, n \in I\!N\}$ une chaîne de Markov sur un espace d'états discret E, de matrice de transition $P = ((P(i,j)))$, et f une fonction bornée sur E.

a) Montrer que si $Pf = e^{\lambda} f$, alors $(e^{-\lambda n} f \circ Y_n)$ est une martingale.

b) Montrer que si $f \leq Pf$, alors $(f \circ Y_n)$ est une sousmartingale.

c) Montrer que $\Big(f(Y_n) + \sum_{k=0}^{n-1} [f(Y_k) - Pf(Y_k)]\Big)$ est une martingale.

Exercice n^o 10.4 Soient $p \in]0,1[$, G une v.a. géométrique de paramètre p: $I\!P(G = n) = (1-p)p^n$, et $\mathcal{F}_n$ la tribu engendrée par la v.a. $n \wedge G$, pour $n \in I\!N$.

a) Vérifier que $\{\mathcal{F}_n \,|\, n \in I\!N\}$ définit une filtration.

b) Montrer que $I\!E\Big(1_{\{G>n\}} \,\Big|\, \mathcal{F}_n\Big) = p\, 1_{\{G \geq n\}}$, pour tout $n \in I\!N$.

c) Déduire que $I\!E(G \wedge (n+1) \,|\, \mathcal{F}_n) = G \wedge n + p\, 1_{\{G \geq n\}}$, pour tout $n \in I\!N$.

d) Trouver $a \in I\!R$ tel que $X_n := a \times (G \wedge n) + 1_{\{G \geq n\}}$ définisse une $(\mathcal{F}_n)$-martingale.

e) Calculer $I\!E\Big((X_{n+1} - X_n)^2 \,\Big|\, \mathcal{F}_n\Big)$, et déduire que $Y_n := X_n^2 - a \times (G \wedge (n-1))$ définit une $(\mathcal{F}_n)$-martingale.

Définition 10.2 *Un processus prévisible est une suite $\{H_n \,|\, n \in I\!N^*\}$ de v.a.r. telles que H_n est $\mathcal{F}_{n-1}$-mesurable, pour tout $n \in I\!N^*$.*

Il faut penser à cela comme à la somme qu'un parieur peut miser au n-ième tirage d'un jeu, n'ayant connaissance que du résultat des $(n-1)$ tirages précédents. Traduit en termes de marché des valeurs (bourse), un processus prévisible, ou plutôt une famille finie de processus prévisibles, représente une stratégie financière (ou de gestion de portefeuille) donnant à chaque instant n la quantité des valeurs (ou actifs disponibles sur le marché) dont est constitué le portefeuille considéré. La nature prévisible signifie alors précisément l'absence de délit d'initié !

Proposition 10.3 *Soient H un processus prévisible borné, (X_n) un processus adapté, et $(H\cdot X)_0 := 0$, puis $(H\cdot X)_n := \sum_{k=1}^{n} H_k(X_k - X_{k-1})$, pour tout $n \in I\!N^*$. Alors*

(i) $H\cdot X$ est une martingale si X en est une.

(ii) Si $H \geq 0$ et si X est une surmartingale (respectivement : sousmartingale), alors il en est de même pour $H\cdot X$.

Preuve Exercice (immédiat d'après les définitions).

Le processus $H\cdot X$ de la proposition 10.3 est appelé l'intégrale stochastique de H par rapport à X.

Dans l'exemple d'un portefeuille de valeurs, X_n représentera le cours (à l'instant n) d'un titre boursier (obligation, option, ou action), H_n la quantité qu'on en détient (ou stratégie de gestion), et $(H{\cdot}X)_n$ la valeur de cette partie du portefeuille (on suppose implicitement la stratégie "autofinancée" : $(H\cdot X)_n - (H\cdot X)_{n-1} = H_n(X_n - X_{n-1})$ signifie que les variations de valeur du portefeuille proviennent uniquement des fluctuations des cours).

Proposition 10.4 *Un processus adapté intégrable (X_n) est une martingale si et seulement si pour tout processus prévisible borné (H_n) on a $\mathbb{E}[(H\cdot X)_n] = 0$ pour tout $n \in \mathbb{N}^*$.*

Preuve Le sens direct découle aussitôt de la proposition 10.3 : $\mathbb{E}[(H{\cdot}X)_n] = \mathbb{E}[(H{\cdot}X)_0] = 0$. Pour la réciproque, à tout $A \in \mathcal{F}_k$ on peut associer $H_n := 1_{A\cap\{n=k+1\}}$, ce qui conduit à $0 = \mathbb{E}[(H\cdot X)_{k+1}] = \mathbb{E}[1_A(X_{k+1} - X_k)]$, d'où $\mathbb{E}[X_{k+1} - X_k \,|\, \mathcal{F}_k] = 0$. $\diamond$

11 Temps et théorème d'arrêt

Voici une notion très importante, qui permet par exemple de modéliser des stratégies. Il s'agit de généraliser les temps constants, en introduisant une famille de temps aléatoires (qui peuvent donc varier d'une trajectoire à l'autre), mais pas n'importe lesquels : on doit pouvoir décider à tout instant si l'événement (généralement non explicite) que représente un tel temps s'est produit ou non, cela uniquement en fonction de l'information connue au moment considéré. Formulons cela rigoureusement, c'est-à-dire mathématiquement.

Définition 11.1 *On appelle temps d'arrêt toute v.a. N à valeurs dans $\overline{\mathbb{N}}$ telle que $\{N \le n\} \in \mathcal{F}_n, \forall\, n \in \mathbb{N}$. On pose alors $\mathcal{F}_N := \{A \in \mathcal{F}_\infty \,|\, (\forall n \in \mathbb{N})\ A \cap \{N \le n\} \in \mathcal{F}_n\}$.*

L'exemple fondamental de temps d'arrêt est le temps d'atteinte par un processus adapté (Z_n) (marche aléatoire, chaîne de Markov) d'une certaine partie de son espace d'états E : $\min\{n \in \mathbb{N} \,|\, Z_n \in E\}$ est bien un temps d'arrêt.

La tribu $\mathcal{F}_N$ représente l'information connue au temps aléatoire N.

Exercice n^o 11.0 Montrer que N est un temps d'arrêt ssi $\{N = n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$, et que $\mathcal{F}_N := \{A \in \mathcal{F}_\infty \,|\, (\forall n \in \mathbb{N})\ A \cap \{N = n\} \in \mathcal{F}_n\}$.

Exercice n^o 11.1 Soient N et N' deux temps d'arrêt tels que $N \le N'$.
a) Vérifier que $\mathcal{F}_N$ est une tribu. b) Montrer que $\mathcal{F}_N \subset \mathcal{F}_{N'}$.
c) Soit $A \in \mathcal{F}_N$. Montrer que $N1_A + N'1_{A^c}$ est un temps d'arrêt.
d) Montrer que si Z est un processus adapté (i.e. : $Z_n \in \mathcal{F}_n$ pour tout n), alors $Z_N \in \mathcal{F}_N$.

Exercice n^o 11.2 Soient N et N' deux temps d'arrêt.
a) Vérifier que $\min\{N, N'\}$ et $\max\{N, N'\}$ sont aussi deux temps d'arrêt ; et $N + N'$?
b) Montrer que les événements $\{N < N'\}$, $\{N \le N'\}$, $\{N = N'\}$ appartiennent à $\mathcal{F}_{\min\{N,N'\}}$, puis que $\mathcal{F}_{\min\{N,N'\}} = \mathcal{F}_N \cap \mathcal{F}_{N'}$.

Exercice n^o 11.3 Soient $\{X_j | j \in I\!N\}$ une suite de v.a.i.i.d. dans $I\!R^d$, $S_n := \sum_{j=1}^{n} X_j$ la marche aléatoire associée, $\mathcal{F}_n := \sigma\{X_j | j \leq n\}$, et N un temps d'arrêt. Supposons $I\!E(\|X_1\|)$ et $I\!E(N)$ finies. Montrer que $I\!E(S_N) = I\!E(X_1) \times I\!E(N)$. (Écrire $S_N = \sum_j X_j 1_{\{N<j\}^c}$).

Exercice n^o 11.4 Si N est un temps d'arrêt, alors $(H_n \equiv 1_{\{N \geq n\}})$ est un processus prévisible, et $(H \cdot X)_n = X_{n \wedge N} - X_0$, $((1 - H) \cdot X)_n = X_n - X_{n \wedge N}$ (de sorte que ce sont des sousmartingales si X en est une).

Proposition 11.2 *Si (X_n) est une surmartingale et si T est un temps d'arrêt (relativement à la filtration $(\mathcal{F}_n)$), alors la "surmartingale arrêtée" $(X_{n \wedge T})$ est encore une surmartingale.*

Preuve $(X_{n \wedge T})$ est intégrable, puisque sa valeur absolue est majorée par $|X_0| \vee .. \vee |X_n|$. L'adaptation découle aussitôt des exercices 11.2 et 11.1. Enfin on applique l'exercice 11.4, ou on écrit directement :

$$I\!E\big(X_{(n+1)\wedge T} - X_{n \wedge T} \,\big|\, \mathcal{F}_n\big) = I\!E\big((X_{(n+1)} - X_n) 1_{\{T \geq n+1\}} \,\big|\, \mathcal{F}_n\big)$$
$$= 1_{\{T \geq n+1\}} I\!E\big(X_{(n+1)} - X_n \,|\, \mathcal{F}_n\big) \leq 0 \,. \;\diamond$$

Exemple Dans un jeu de pile ou face : $(X_n)_{n \geq 1}$ v.a.i.i.d. de loi de Bernoulli $\mathcal{B}(\pm 1, \frac{1}{2})$, soit $\tau := \min\{n \in I\!N^* | \, X_n = 1\}$. Parier deux fois plus à chaque lancer jusqu'à gagner revient mathématiquement à poser $H_n := 2^{n-1} 1_{\{\tau \geq n\}}$, $S_n := \sum_{k=1}^{n} X_k$, et à considérer $(H \cdot S)$. On a $(H \cdot S)_\tau = 1$, c'est-à-dire qu'on gagne 1 au temps τ, qui est p.s. fini. Mais attention, τ n'est pas borné, et dans la pratique on ne peut pas jouer indéfiniment ; de sorte qu'il y a un risque significatif de perdre gros !

Nous avons maintenant l'important "théorème d'arrêt" de Doob :

Théorème 11.3 *Soient S et T deux temps d'arrêt p.s. bornés, tels que p.s. $S \leq T$. Soit (X_n) une surmartingale. Alors on a p.s. $I\!E(X_T \,|\, \mathcal{F}_S) \leq X_S$.*

Preuve Soit M un entier fixe majorant p.s. le temps T. X_T est intégrable, sa valeur absolue étant majorée par $|X_0| + \ldots + |X_M|$. Pour tout $A \in \mathcal{F}_S$, la proposition 11.2 ci-dessus entraîne :

$$I\!E[(X_T - X_S) 1_A] = \sum_{m=0}^{M} I\!E[(X_{M \wedge T} - X_{m \wedge T}) 1_{A \cap \{S=m\}}] \leq 0 \,. \;\diamond$$

En particulier, pour toute martingale (X_n) et tout temps d'arrêt borné T, on a $I\!E(X_T) = I\!E(X_0)$.

Si T n'est pas borné : le temps d'atteinte de 1 pour la marche simple sur $\mathbb{Z}$ (c'est-à-dire (S_n) dans l'exemple ci-dessus) fournit un contrexemple, qui prouve du même coup que T ne peut pas être intégrable ; il suffit en effet d'appliquer l'exercice 11.3.

Exercice n^o 11.5 Soient N un entier ≥ 2, un réel $A > 0$, une filtration $\{\mathcal{F}_n \,|\, n \in \mathbb{N}\}$, (M_n) une $(\mathcal{F}_n)$-martingale, (H_n) un processus $(\mathcal{F}_n)$-adapté,

$T := N \wedge \min\{n \in \mathbb{N}^* \,|\, |H_n| > A\}$, et $X_n := \sum_{j=1}^{T\wedge n} H_{j-1}(M_j - M_{j-1})$ pour $n \in \mathbb{N}^*$.
Montrer que T est un $(\mathcal{F}_n)$-temps d'arrêt, et que (X_n) est une $(\mathcal{F}_n)$-martingale.

12 Inégalités et Convergence

Le théorème suivant énonce l'importante "inégalité de Doob".

Théorème 12.1 *Soient (X_n) une martingale, et $N \in \mathbb{N}$. Nous avons :*

$$\Big\| \max\{|X_n| \,|\, 0 \leq n \leq N\} \Big\|_2 \leq 2\, \|X_N\|_2 \,, \quad \textit{et} \quad \Big\| \sup\{|X_n| \,|\, n \in \mathbb{N}\} \Big\|_2 \leq 2 \sup_n \|X_n\|_2 \,.$$

Théorème 12.2 *Soit (X_n) une surmartingale, telle que $\sup_n \mathbb{E}(X_n^-) < \infty$. Alors (X_n) est bornée dans L^1, et converge presque sûrement vers une variable intégrable X_∞.*

Preuve partielle La convergence L^2 est facile à obtenir dans le cas d'une martingale (ou bien sousmartingale positive) bornée dans L^2, car le carré de la norme L^2 est alors contrôlé par la somme des carrés des normes des accroissements : pour tous entiers $0 \leq n < N$,

$$\sum_{k=n+1}^{N} \|M_k - M_{k-1}\|_2^2 = \sum_{k=n+1}^{N} \left(\|M_k\|_2^2 - \|M_{k-1}\|_2^2 \right) = \|M_N\|_2^2 - \|M_n\|_2^2 \leq C < \infty \,;$$

d'où (comme queue de série convergente, pour tout $\varepsilon > 0$) pour $N > n \geq n_\varepsilon$:

$$\|M_N - M_n\|_2^2 = \mathbb{E}\big[(M_N - M_n)^2 + 2(M_N - M_n)M_n\big] = \|M_N\|_2^2 - \|M_n\|_2^2 \leq \varepsilon \,;$$

de sorte que le critère de Cauchy s'applique. Faisant tendre m vers l'infini dans l'égalité $\mathbb{E}(M_{n+m} \,|\, \mathcal{F}_n) = M_n$, on déduit alors en outre que $\mathbb{E}(M_\infty \,|\, \mathcal{F}_n) = M_n$ pour tout n, presque sûrement.

Corollaire 12.3 *Toute surmartingale positive $(X_n) \geq 0$ converge presque sûrement vers une variable intégrable $X_\infty \geq 0$, et on a pour tout $n \in \mathbb{N}$:* $\mathbb{E}\Big(X_\infty \,\Big|\, \mathcal{F}_n\Big) \leq X_n$.

Exercice n^o 12.1 Une urne contient initialement n boules noires et b boules blanches. On tire une boule dans l'urne, selon la loi uniforme, puis on remet la boule tirée avec a autres boules de la même couleur. On itère indéfiniment cette procédure. Notons X_ℓ la proportion de boules noires dans l'urne après le ℓ-ième tirage (de sorte que $X_0 = n/(n+b)$).

a) Montrer que (X_ℓ) est une chaîne de Markov inhomogène sur $\mathbb{Q} \cap [0,1]$.

b) Montrer que (X_ℓ) est l'image par une application à expliciter d'une chaîne de Markov homogène sur $\mathbb{N}^2$, dont on précisera le noyau de transition.

c) Montrer que (X_ℓ) est une martingale, qui converge p.s. et dans tous les L^p.

d) Dans le cas $n = b = a = 1$, calculer la loi de (X_ℓ), et en déduire la loi de X_∞.

Exercice n^o 12.2 Soient $N \in \mathbb{N}^*$, et $\{B_n \,|\, n \in \mathbb{N}^*\}$ une suite i.i.d. de v.a. de Bernoulli prenant les valeurs $1, -1$ avec probabilité $p, 1-p$ respectivement, p étant fixé dans $]\frac{1}{2}, 1[$. Posons $\alpha := p \log p + (1-p)\log(1-p) + \log 2$.

Considérons un jeu se déroulant en N parties, et tel qu'à la n-ième partie on gagne $B_n Y_n$ si on a misé Y_n. Notons $\mathcal{F}_n$ la tribu engendrée par $\{B_1, .., B_n\}$, et G_n la fortune du joueur après la n-ième partie, pour tout $n \in \mathbb{N}$. G_0 est une constante positive donnée.

Le processus (Y_n) représente la stratégie du joueur, et doit donc être "prévisible" : Y_n doit être $\mathcal{F}_{n-1}$-mesurable, pour tout $n \in \mathbb{N}$. De plus, Y_n doit appartenir à $[0, G_{n-1}[$.

a) Montrer que, pour toute stratégie, $(\log G_n - \alpha n)$ est une surmartingale.

b) Montrer que, pour une certaine stratégie, $(\log G_n - \alpha n)$ est une martingale.

c) Quelle est la stratégie maximisant $\mathbb{E}[\log(G_N/G_0)]$?

13 Décomposition de Doob et variations quadratiques

Proposition 13.1 *Tout processus adapté $X = \{X_n \,|\, n \in \mathbb{N}\}$ de v.a.r. intégrables se décompose de façon unique en $X = X_0 + M + A$, où M est une martingale nulle en 0 et A est un processus nul en 0, intégrable et prévisible (cf définition 10.2). En outre, X est une surmartingale si et seulement si A est p.s. décroissant (et une sousmartingale si et seulement si A est p.s. croissant).*

Preuve Tout d'abord, on a nécessairement $A_n = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1} | \mathcal{F}_{k-1})$ p.s. si cette décomposition existe, ce qui établit l'unicité. Réciproquement, on prend cette formule comme définition de (A_n), et on vérifie qu'alors $M_n := X_n - X_0 - A_n$ définit bien une martingale, ce qui est immédiat, et établit l'existence de la décomposition. La dernière assertion est évidente. $\diamond$

Définition 13.2 *Dans la décomposition de Doob de la proposition 13.1, le processus $X_0 + A$ est appelé la projection duale prévisible du processus X. Sa projection prévisible est le processus prévisible nul en 0 et égal à $\mathbb{E}[X_n \,|\, \mathcal{F}_{n-1}]$ en tout temps $n \geq 1$.*

Définition 13.3 *La variation quadratique optionnelle d'un processus (non nécessairement adapté) X est le processus croissant noté $[X, X]$ défini par :*

$$[X, X]_n := X_0^2 + \sum_{k=1}^{n} (X_k - X_{k-1})^2 , \quad \textit{pour tout } n \geq 0 .$$

La <u>*variation quadratique prévisible*</u> *(relativement à l'espace probabilisé fitré* $(\Omega, \mathcal{F}_n, \mathbb{P})$*) de* X *est le processus croissant* $\langle X, X\rangle$ *défini comme la projection duale prévisible de* $[X, X]$*. (Cela sous-entend que le processus* X *est de carré intégrable :* $X_n \in L^2$ *pour tout* n*.)*

Cela s'étend aussitôt par polarisation à la <u>*covariation quadratique optionnelle*</u> *et à la* <u>*covariation quadratique prévisible*</u> *de deux processus* X *et* Y *:*

$$[X, Y] = \tfrac{1}{2}([X+Y, X+Y]-[X, X]-[Y, Y])\,; \quad \langle X, Y\rangle = \tfrac{1}{2}(\langle X+Y, X+Y\rangle-\langle X, X\rangle-\langle Y, Y\rangle)\,.$$

On a ainsi : $\quad [X, Y]_n = X_0Y_0 + \sum_{k=1}^{n}(X_k - X_{k-1})(Y_k - Y_{k-1})\,,\quad$ et :

$\langle X, Y\rangle_n = X_0Y_0 + \sum_{k=1}^{n} \mathbb{E}\Big((X_k - X_{k-1})(Y_k - Y_{k-1})\,\Big|\,\mathcal{F}_{k-1}\Big)$ (pour X, Y de carré intégrable).

Proposition 13.4 *Si* X *et* Y *sont deux martingales de carré intégrable, alors les deux processus* $XY - [X, Y]$ *et* $XY - \langle X, Y\rangle$ *sont aussi des martingales.*

<u>Preuve</u> On a en effet pour tout $n \geq 1$:

$$\mathbb{E}(X_nY_n - [X, Y]_n \,|\, \mathcal{F}_{n-1}) - (X_{n-1}Y_{n-1} - [X, Y]_{n-1})$$
$$= \mathbb{E}\Big(X_nY_n - X_{n-1}Y_{n-1} - (X_n - X_{n-1})(Y_n - Y_{n-1})\,\Big|\,\mathcal{F}_{n-1}\Big)$$
$$= \mathbb{E}(X_nY_{n-1} + X_{n-1}Y_n \,|\, \mathcal{F}_{n-1}) - 2\,X_{n-1}Y_{n-1}$$
$$= \Big(\mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1}\Big)Y_{n-1} + \Big(\mathbb{E}(Y_n|\mathcal{F}_{n-1}) - Y_{n-1}\Big)X_{n-1} = 0\,,$$

et d'après les définitions 13.2 et 13.3, $[X, Y] - \langle X, Y\rangle$ est une martingale ; ou bien, de même :

$$\mathbb{E}(X_nY_n - \langle X, Y\rangle_n \,|\, \mathcal{F}_{n-1}) - (X_{n-1}Y_{n-1} - \langle X, Y\rangle_{n-1})$$
$$= \mathbb{E}\Big(X_nY_n - X_{n-1}Y_{n-1} - (X_n - X_{n-1})(Y_n - Y_{n-1})\,\Big|\,\mathcal{F}_{n-1}\Big) = 0 \quad \diamond$$

14 Propriétés de l'intégrale stochastique discrète

Nous passons ici en revue les propriétés (autre que celles de la proposition 10.3) de l'intégrale stochastique discrète $H\cdot X$, de H prévisible (borné) par rapport à X adapté, introduite dans la proposition 10.3 (section 10), par ce qui peut se résumer en :

$$(H\cdot X)_0 = 0\,, \quad \text{et} \quad (H\cdot X)_n - (H\cdot X)_{n-1} = H_n(X_n - X_{n-1})\,, \text{ pour tout } n \in \mathbb{N}^* \qquad (3)$$

Proposition 14.1 *Si X et Y sont deux processus adaptés de carré intégrable et si H et K sont deux processus prévisibles bornés, alors nous avons :*

$$K\cdot(H\cdot X) = (HK)\cdot X \ ; \quad [H\cdot X, Y] = H\cdot[X,Y] \ ; \quad \langle H\cdot X, Y\rangle = H\cdot\langle X,Y\rangle \, .$$

<u>Preuve</u> Utilisant (3), nous avons

$$(K\cdot(H\cdot X))_n - (K\cdot(H\cdot X))_{n-1} = K_n((H\cdot X)_n - (H\cdot X)_{n-1}) = K_n(H_n(X_n - X_{n-1}))$$
$$= (HK)_n(X_n - X_{n-1}) = ((HK)\cdot X)_n - ((HK)\cdot X)_{n-1} \, ,$$

ce qui prouve la première formule, ses deux membres étant nuls en 0. Ensuite, nous avons

$$[H\cdot X, Y]_n = \sum_{k=1}^{n} H_n(X_n - X_{n-1})(Y_n - Y_{n-1}) = (H\cdot[X,Y])_n \, .$$

Enfin,

$$\langle H\cdot X, Y\rangle_n - \langle H\cdot X, Y\rangle_{n-1} = I\!E\Big(H_n(X_n - X_{n-1})(Y_n - Y_{n-1})\Big|\mathcal{F}_{n-1}\Big)$$
$$= H_n\, I\!E\Big((X_n - X_{n-1})(Y_n - Y_{n-1})\Big|\mathcal{F}_{n-1}\Big) = H_n\Big(\langle X,Y\rangle_n - \langle X,Y\rangle_{n-1}\Big)$$
$$= (H\cdot\langle X,Y\rangle)_n - (H\cdot\langle X,Y\rangle)_{n-1} \, ,$$

en utilisant (3) et la prévisibilité de H. Cela prouve la dernière formule, ses deux membres étant nuls en 0. $\diamond$

Notation Pour tous processus adapté X et prévisible H, notons $\displaystyle\int H dX$ pour $H\cdot X$, $\displaystyle\int_0^n H dX$ pour $(H\cdot X)_n$, et X_- le processus prévisible défini par : $(X_-)_n := X_{n-1}$, pour tout $n \geq 1$.

Proposition 14.2 *Soient X et Y deux processus adaptés. Nous avons*

$$X^2 = [X,X] + 2\int X_-\, dX \, ,$$

et la formule d'intégration par parties :

$$XY = [X,Y] + \int X_-\, dY + \int Y_-\, dX \, .$$

<u>Preuve</u> La première formule, dont les deux membres valent X_0^2 en 0, résulte aussitôt de l'identité élémentaire suivante, qu'il suffit de sommer de $k=1$ à $k=n$:

$$X_k^2 - X_{k-1}^2 = (X_k - X_{k-1})^2 + 2\, X_{k-1}(X_k - X_{k-1}) \, .$$

La seconde formule s'en déduit aussitôt par polarisation. $\diamond$

Proposition 14.3 *Soient X un processus adapté, et $Y_0 \in \mathbb{R}$. Il existe un unique processus Y solution de l'équation*

$$Y = Y_0 + \int Y_- \, dX \,.$$

On a $Y_n = Y_0 \prod_{k=1}^{n} (1 + X_k - X_{k-1})$, *et Y est une martingale si X en est une.*

Preuve Par récurrence sur n, on voit sans difficulté que la formule de l'énoncé est nécessairement la bonne, et qu'elle définit bien une solution. La dernière assertion résulte aussitôt de la proposition 10.3. $\diamond$

Définition 14.4 *La solution unique Y de la proposition 14.3, pour $Y_0 = 1$, s'appelle l'exponentielle stochastique discrète de X, et est notée $\mathcal{E}(X)$.*

Proposition 14.5 *Soient X et Y deux processus adaptés. Nous avons*

$$\mathcal{E}(X)\,\mathcal{E}(Y) = \mathcal{E}(X + Y + [X, Y]) \,.$$

Preuve Les 2 membres valent 1 en 0. Par unicité dans la proposition 14.3, il suffit donc de vérifier que

$$\mathcal{E}(X)_n \times \mathcal{E}(Y)_n - \mathcal{E}(X)_{n-1} \times \mathcal{E}(Y)_{n-1} = \int_{n-1}^{n} \mathcal{E}(X)_- \, \mathcal{E}(Y)_- \, d(X + Y + [X, Y]) \,.$$

Or, utilisant la formule d'intégration par parties de la proposition 14.2, la définition 13.3 de la variation quadratique optionnelle, et les équations que résolvent $\mathcal{E}(X)$ et $\mathcal{E}(Y)$, nous obtenons :

$$\begin{aligned}
&\mathcal{E}(X)_n \times \mathcal{E}(Y)_n - \mathcal{E}(X)_{n-1} \times \mathcal{E}(Y)_{n-1} \\
&= [\mathcal{E}(X), \mathcal{E}(Y)]_n - [\mathcal{E}(X), \mathcal{E}(Y)]_{n-1} + \int_{n-1}^{n} \mathcal{E}(X)_- \, d\mathcal{E}(Y) + \int_{n-1}^{n} \mathcal{E}(Y)_- \, d\mathcal{E}(X) \\
&= (\mathcal{E}(X)_n - \mathcal{E}(X)_{n-1})(\mathcal{E}(Y)_n - \mathcal{E}(Y)_{n-1}) + \int_{n-1}^{n} \mathcal{E}(X)_- \, \mathcal{E}(Y)_- \, dY + \int_{n-1}^{n} \mathcal{E}(Y)_- \, \mathcal{E}(X)_- \, dX \\
&= \mathcal{E}(X)_{n-1}(X_n - X_{n-1}) \, \mathcal{E}(Y)_{n-1}(Y_n - Y_{n-1}) + \int_{n-1}^{n} \mathcal{E}(X)_- \, \mathcal{E}(Y)_- \, d(X + Y) \\
&= \int_{n-1}^{n} \mathcal{E}(X)_- \, \mathcal{E}(Y)_- \, d[X, Y] + \int_{n-1}^{n} \mathcal{E}(X)_- \, \mathcal{E}(Y)_- \, d(X + Y) \,. \diamond
\end{aligned}$$

15 Changement de probabilité sur Ω fini

Supposons pour cette section que $\Omega = \{\omega_1, \ldots, \omega_N\}$ est fini, et considérons sur $\mathcal{T} = \mathcal{P}(\Omega)$ deux probabilités $\mathbb{P}$ et Q de support égal à Ω entier : $\mathbb{P}(\{\omega\}) \times Q(\{\omega\}) > 0$ pour tout $\omega \in \Omega$.

Fixons une filtration $(\mathcal{F}_n)$ sur Ω, et posons $f(\omega) := Q(\{\omega\})/\mathbb{P}(\{\omega\})$ pour tout $\omega \in \Omega$, ce qui définit une fonction (ou variable aléatoire) f de Ω dans $\mathbb{R}_+^*$, qui est la densité de Q par rapport à $\mathbb{P}$. On a $\mathbb{E}^Q[h] := \int h\, dQ = \int h\, f\, d\mathbb{P} =: \mathbb{E}^{\mathbb{P}}[h\, f]$, pour toute v.a. h sur $(\Omega, \mathcal{F}_N)$.

Considérons la $\mathbb{P}$-martingale (ce qui signifie qu'elle est une martingale par rapport à $(\Omega, \mathcal{F}_n, \mathbb{P})$) strictement positive définie par : $M_n := \mathbb{E}^{\mathbb{P}}(f \mid \mathcal{F}_n)$, pour tout n ; la notation $\mathbb{E}^{\mathbb{P}}$ signifiant que l'espérance conditionnelle est prise relativement à la probabilité $\mathbb{P}$.

Le lien entre les espérances conditionnelles relatives à $\mathbb{P}$ et à Q est facile à établir.

Proposition 15.1 *Soit X un processus adapté. Pour tous temps $0 \le k \le n$, nous avons :*

$$\mathbb{E}^{\mathbb{P}}(X_n\, M_n \mid \mathcal{F}_k) = M_k \times \mathbb{E}^Q(X_n \mid \mathcal{F}_k)\,.$$

De sorte que X est une Q-martingale si et seulement si MX est une $\mathbb{P}$-martingale.

<u>Preuve</u> Utilisant la remarque 6.6, et notant $\{A_1, \ldots, A_{n_k}\}$ la partition engendrant $\mathcal{F}_k$, nous avons :

$$\mathbb{E}^{\mathbb{P}}(X_n\, M_n \mid \mathcal{F}_k) = \sum_{j=1}^{n_k} \mathbb{E}^{\mathbb{P}}(X_n\, M_n/A_j)\, 1_{A_j} = \sum_{j=1}^{n_k} \frac{\mathbb{E}^{\mathbb{P}}(X_n\, M_n\, 1_{A_j})}{\mathbb{P}(A_j)}\, 1_{A_j}$$

$$= \sum_{j=1}^{n_k} \frac{\mathbb{E}^{\mathbb{P}}(X_n\, f\, 1_{A_j})}{\mathbb{P}(A_j)}\, 1_{A_j} = \sum_{j=1}^{n_k} \frac{\mathbb{E}^Q(X_n\, 1_{A_j})}{\mathbb{P}(A_j)}\, 1_{A_j} = \sum_{j=1}^{n_k} \frac{\mathbb{E}^Q(X_n\, 1_{A_j})}{Q(A_j)}\, \frac{\mathbb{E}^{\mathbb{P}}(f\, 1_{A_j})}{\mathbb{P}(A_j)}\, 1_{A_j}$$

$$= \sum_{j=1}^{n_k} \frac{\mathbb{E}^Q(X_n\, 1_{A_j})}{Q(A_j)}\, \frac{\mathbb{E}^{\mathbb{P}}(M_k\, 1_{A_j})}{\mathbb{P}(A_j)}\, 1_{A_j} = \mathbb{E}^Q(X_n \mid \mathcal{F}_k)\, \mathbb{E}^{\mathbb{P}}(M_k \mid \mathcal{F}_k) = \mathbb{E}^Q(X_n \mid \mathcal{F}_k)\, M_k\,. \diamond$$

Théorème 15.2 *(Théorème de Girsanov discret) Soit X une $\mathbb{P}$-martingale. Considérons les 3 intégrales stochastiques discrètes : $\int \frac{dM}{M_-}$, $\int \frac{d[M,X]}{M}$, $\int \frac{d\langle M,X\rangle}{M_-}$. $\int \frac{dM}{M_-}$ est une $\mathbb{P}$-martingale, et $X - \int \frac{d[M,X]}{M}$ et $X - \int \frac{d\langle M,X\rangle}{M_-}$ sont des Q-martingales.*
(N.B. La variation quadratique prévisible $\langle M,X\rangle$ est ici relative à $(\mathcal{F}_n, \mathbb{P})$ (et non à Q).)
(N.B. L'intégrande M^{-1} de la seconde intégrale est adapté, mais non prévisible.)

Preuve Pour ce qui est de la première intégrale, c'est seulement dû à la proposition 10.3. Considérons la seconde, et un temps $n \geq 1$. Utilisant la proposition 15.1, nous avons :

$$\mathbb{E}^Q\Big(X_n - \int_0^n \frac{d[M,X]}{M}\,\Big|\,\mathcal{F}_{n-1}\Big) - \Big(X_{n-1} - \int_0^{n-1} \frac{d[M,X]}{M}\Big)$$

$$= \mathbb{E}^Q(X_n\,|\,\mathcal{F}_{n-1}) - X_{n-1} - \mathbb{E}^Q\Big((X_n - X_{n-1})(M_n - M_{n-1})\Big/M_n\,\Big|\,\mathcal{F}_{n-1}\Big)$$

$$= \frac{\mathbb{E}^{\mathbb{P}}\Big(X_n\,M_n\,\Big|\,\mathcal{F}_{n-1}\Big) - \mathbb{E}^{\mathbb{P}}\Big((X_n - X_{n-1})(M_n - M_{n-1})\,\Big|\,\mathcal{F}_{n-1}\Big)}{M_{n-1}} - X_{n-1}$$

$$= \mathbb{E}^{\mathbb{P}}\Big(X_{n-1}\,\tfrac{M_n}{M_{n-1}} + X_n - X_{n-1}\,\Big|\,\mathcal{F}_{n-1}\Big) - X_{n-1} = \tfrac{X_{n-1}}{M_{n-1}}\,\mathbb{E}^{\mathbb{P}}(M_n\,|\,\mathcal{F}_{n-1}) - X_{n-1} = 0\,.$$

Et de façon analogue :

$$\mathbb{E}^Q\Big(X_n - \int_0^n \frac{d\langle M,X\rangle}{M_-}\,\Big|\,\mathcal{F}_{n-1}\Big) - \Big(X_{n-1} - \int_0^{n-1} \frac{d\langle M,X\rangle}{M_-}\Big)$$

$$= \mathbb{E}^Q(X_n\,|\,\mathcal{F}_{n-1}) - X_{n-1} - \frac{\langle M,X\rangle_n - \langle M,X\rangle_{n-1}}{M_{n-1}}$$

$$= \frac{\mathbb{E}^{\mathbb{P}}\Big(X_n\,M_n\,\Big|\,\mathcal{F}_{n-1}\Big)}{M_{n-1}} - X_{n-1} - \frac{\mathbb{E}^{\mathbb{P}}\Big((X_n - X_{n-1})(M_n - M_{n-1})\,\Big|\,\mathcal{F}_{n-1}\Big)}{M_{n-1}} = 0\,. \quad \diamond$$

15.3 Exemple numérique (tiré du livre de Dothan, pages 105-109)

Prenons $N = 6$, $\mathbb{P}$ uniforme sur Ω, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_3 = \mathcal{T} = \mathcal{P}(\Omega)$,

$$Q(\omega_1) = Q(\omega_6) = 1/4\,,\quad Q(\omega_2) = Q(\omega_3) = Q(\omega_4) = Q(\omega_5) = 1/8\,,$$

puis $\mathcal{F}_1 := \big\{\emptyset, \{\omega_1,\omega_2\}, \{\omega_3,\omega_4,\omega_5,\omega_6\}, \Omega\big\}$, $\mathcal{F}_2 := \sigma\big(\{\omega_1,\omega_2\}, \{\omega_3,\omega_4\}, \{\omega_5,\omega_6\}\big)$.

Alors $f = \frac{3}{2}\,1_{\{\omega_1,\omega_6\}} + \frac{3}{4}\,1_{\{\omega_2,\omega_3,\omega_4,\omega_5\}}$, et la $\mathbb{P}$-martingale positive (M_n) est donnée par :

$M_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n = 0$	1	1	1	1	1	1
$n = 1$	$\frac{9}{8}$	$\frac{9}{8}$	$\frac{15}{16}$	$\frac{15}{16}$	$\frac{15}{16}$	$\frac{15}{16}$
$n = 2$	$\frac{9}{8}$	$\frac{9}{8}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{9}{8}$	$\frac{9}{8}$
$n = 3$	$\frac{3}{2}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{2}$

La $\mathbb{P}$-martingale $\displaystyle\int_0^n \frac{dM}{M_-} = \sum_{k=1}^n \frac{M_k - M_{k-1}}{M_{k-1}}$ est donnée par :

$\int_0^n \frac{dM}{M_-}(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{-1}{16}$	$\frac{-1}{16}$	$\frac{-1}{16}$	$\frac{-1}{16}$
$n=2$	$\frac{1}{8}$	$\frac{1}{8}$	$-\frac{21}{80}$	$-\frac{21}{80}$	$\frac{11}{80}$	$\frac{11}{80}$
$n=3$	$\frac{11}{24}$	$\frac{-5}{24}$	$-\frac{21}{80}$	$-\frac{21}{80}$	$\frac{-47}{240}$	$\frac{113}{240}$

Considérons la $I\!\!P$-martingale X donné par :

$X_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	10	10	-5	-5	-5	-5
$n=2$	10	10	-21	-21	11	11
$n=3$	$\frac{110}{3}$	$-\frac{50}{3}$	-21	-21	$-\frac{47}{3}$	$\frac{113}{3}$

Les covariations quadratiques $[M,X]_n = M_0X_0 + \sum_{k=1}^{n}(M_k - M_{k-1})(X_k - X_{k-1})$ et

$\langle M,X\rangle_n = M_0X_0 + \sum_{k=1}^{n} I\!\!E\Big((M_k - M_{k-1})(X_k - X_{k-1})\,\Big|\,\mathcal{F}_{k-1}\Big)$ sont données par :

$[M,X]_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{5}{4}$	$\frac{5}{4}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{16}$
$n=2$	$\frac{5}{4}$	$\frac{5}{4}$	$\frac{53}{16}$	$\frac{53}{16}$	$\frac{53}{16}$	$\frac{53}{16}$
$n=3$	$\frac{45}{4}$	$\frac{45}{4}$	$\frac{53}{16}$	$\frac{53}{16}$	$\frac{213}{16}$	$\frac{213}{16}$

$\langle M,X\rangle_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$
$n=2$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{29}{8}$	$\frac{29}{8}$	$\frac{29}{8}$	$\frac{29}{8}$
$n=3$	$\frac{85}{8}$	$\frac{85}{8}$	$\frac{29}{8}$	$\frac{29}{8}$	$\frac{109}{8}$	$\frac{109}{8}$

$$A_n := \int_0^n \frac{d[M,X]}{M} = \sum_{k=1}^{n} \frac{[M,X]_k - [M,X]_{k-1}}{M_k} \quad \text{et}$$

$$B_n := \int_0^n \frac{d\langle M,X\rangle}{M_-} = \sum_{k=1}^{n} \frac{\langle M,X\rangle_k - \langle M,X\rangle_{k-1}}{M_{k-1}} \quad \text{sont données par :}$$

$A_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$n=2$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{13}{3}$	$\frac{13}{3}$	3	3
$n=3$	$\frac{70}{9}$	$\frac{130}{9}$	$\frac{13}{3}$	$\frac{13}{3}$	$\frac{49}{3}$	$\frac{29}{3}$

$B_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{5}{8}$
$n=2$	$\frac{5}{8}$	$\frac{5}{8}$	$\frac{153}{40}$	$\frac{153}{40}$	$\frac{153}{40}$	$\frac{153}{40}$
$n=3$	$\frac{685}{72}$	$\frac{685}{72}$	$\frac{153}{40}$	$\frac{153}{40}$	$\frac{4577}{360}$	$\frac{4577}{360}$

Donc les Q-martingales $(X-A)$ et $(X-B)$ du théorème 15.2 sont données par

$(X-A)_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{80}{9}$	$\frac{80}{9}$	$-\frac{16}{3}$	$-\frac{16}{3}$	$-\frac{16}{3}$	$-\frac{16}{3}$
$n=2$	$\frac{80}{9}$	$\frac{80}{9}$	$-\frac{76}{3}$	$-\frac{76}{3}$	8	8
$n=3$	$\frac{260}{9}$	$-\frac{280}{9}$	$-\frac{76}{3}$	$-\frac{76}{3}$	-32	28

$(X-B)_n(\omega)$	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
$n=0$	0	0	0	0	0	0
$n=1$	$\frac{75}{8}$	$\frac{75}{8}$	$-\frac{45}{8}$	$-\frac{45}{8}$	$-\frac{45}{8}$	$-\frac{45}{8}$
$n=2$	$\frac{75}{8}$	$\frac{75}{8}$	$-\frac{993}{40}$	$-\frac{993}{40}$	$\frac{287}{40}$	$\frac{287}{40}$
$n=3$	$\frac{1955}{72}$	$-\frac{1885}{72}$	$-\frac{993}{40}$	$-\frac{993}{40}$	$-\frac{10217}{360}$	$\frac{8983}{360}$

Il est maintenant facile de contrôler sur cet exemple que M, $\int \frac{dM}{M_-}$, et X sont des $I\!\!P$-martingales, et que $X - \int \frac{d[M,X]}{M}$ et $X - \int \frac{d\langle M,X\rangle}{M_-}$ sont des Q-martingales.

16 Application aux marchés : arbitrage

La problématique de l'arbitrage est essentielle aux marchés financiers, puisqu'il s'agit de la possibilité d'y gagner de l'argent sans courir de risque.

Modélisons ceci en termes de stratégie financière (ou de gestion de portefeuille, donnant à chaque instant n la quantité des valeurs (ou actifs disponibles sur le marché) dont est constitué le portefeuille considéré), qui doit être un processus prévisible (cf définition 10.2), comme déjà signalé en section 10.

On notera ici $\varphi\cdot\sigma := \sum_{j=0}^{d}\varphi^j\,\sigma^j$ le produit scalaire usuel de 2 vecteurs $\varphi,\sigma\in\mathbb{R}^{1+d}$.

Notons $\phi=(\phi^0,\ldots,\phi^d)\,:\,\mathbb{N}\times\Omega\to\mathbb{R}^{1+d}$ un processus prévisible représentant une stratégie, et $V(\phi):=\phi\cdot S\,:\,\mathbb{N}\times\Omega\to\mathbb{R}$ la <u>valeur</u> du portefeuille correspondant. Ici le processus $S=(S^0,\ldots,S^d)\,:\,\mathbb{N}\times\Omega\to\mathbb{R}^{1+d}$ représente les cours successifs d'un vecteur d'actifs financiers (actions, obligations, options) présents sur le marché. Nous avons donc :

$$V_n(\phi)(\omega)=\sum_{j=0}^{d}\phi_n^j(\omega)\,S_n^j(\omega)\,,\ \text{ pour tous }\ n\geq 0\,,\ \omega\in\Omega\,.$$

L'actif S^0 joue un rôle particulier : il représente un placement sans risque (par exemple à la caisse d'épargne), et correspond donc à un taux d'intérêt déterministe (ou à l'indice général des prix). C'est dire mathématiquement qu'on a : $S_0^0=1,\ldots,S_n^0=\prod_{j=1}^{n}(1+r_j)$, pour une suite déterministe de taux ou indices successifs (r_n).
La <u>valeur actualisée</u> du portefeuille est ainsi donnée par le processus : $\tilde{V}(\phi):=V(\phi)/S^0$, et le vecteur des cours actualisés est le processus $\tilde{S}:=S/S^0=(1,S^1/S^0,\ldots,S^d/S^0)$.

Définition 16.1 *La stratégie ϕ est dite <u>autofinancée</u> lorsque $(\phi_{n+1}-\phi_n)\cdot S_n=0$ pour tout $n\geq 0$. La stratégie ϕ est dite <u>admissible</u> si elle est autofinancée et si $V_n(\phi)\geq 0$ pour tout $n\geq 0$. La stratégie ϕ est dite <u>d'arbitrage</u> lorsque qu'elle est admissible, de valeur initiale nulle, mais non de valeur nulle à tout instant $n\leq N$.*

L'autofinancement signifie que la valeur du portefeuille évolue sans apport ni retrait de fonds, mais seulement par modifications de sa structure (ϕ_n devient ϕ_{n+1}), en réinvestissant à chaque instant $(n+1)$ (dans les actifs n^o $0,\ldots,d$) exactement la valeur totale $V_n(\phi)$, cela connaissant seulement $\mathcal{F}_n$. On aura alors

$$V_{n+1}(\phi)-V_n(\phi)=\ \phi_{n+1}\cdot(S_{n+1}-S_n)\,,$$

ce qui signifie que pour une stratégie donnée, les variations de valeur ne sont dues qu'aux variations des cours.

L'admissibilité précise qu'en outre l'emprunt (global) n'est pas autorisé. Il est toutefois possible d'avoir des emprunts locaux (id est certains $\phi^j<0$), à condition que l'ensemble du portefeuille reste non débiteur.

Les marchés (bourses) s'appliquent généralement à exclure toute possibilité d'arbitrage, sans quoi ils ne seraient guère viables ! La caractérisation mathématique de cette exigence est élégante, et constitue le premier théorème fondamental des mathématiques financières.

Théorème 16.2 *Il n'existe pas de stratégie d'arbitrage si et seulement si il existe une probabilité équivalente à $\mathbb{P}$ sous laquelle les cours actualisés des actifs sont des martingales.*

Précisons qu'une probabilité équivalente à $\mathbb{P}$ est une probabilité Q sur $(\Omega, \mathcal{T})$ (l'espace probabilisable sur lequel est déjà définie $\mathbb{P}$) qui, comme dans la section 15, admet par rapport à $\mathbb{P}$ une densité $f > 0$: $\mathbb{E}^Q(h) = \mathbb{E}^{\mathbb{P}}(hf)$ pour toute fonction $\mathcal{T}$-mesurable positive h.

Preuve (i) Comme observé un peu plus haut, pour toute stratégie autofinancée ϕ nous avons $V_n(\phi) = V_0(\phi) + \sum_{k=1}^{n} \phi_k \cdot (S_k - S_{k-1})$ et donc $\tilde{V}_n(\phi) = V_0(\phi) + \sum_{k=1}^{n} \phi_k \cdot (\tilde{S}_k - \tilde{S}_{k-1})$. Supposant l'existence de Q comme dans l'énoncé, on voit (proposition 10.3) qu'il s'agit d'une Q-martingale, de sorte que $\mathbb{E}^Q[\tilde{V}_n(\phi)] = \mathbb{E}^Q[V_0(\phi)]$, qui est nulle si la valeur initiale est nulle. Si $\tilde{V}_n(\phi) \geq 0$, cela impose que $\tilde{V}_n(\phi) = 0$ Q-p.s., et donc aussi $\mathbb{P}$-p.s..

(ii) La preuve de la réciproque est moins facile. Nous nous limiterons au cas de Ω fini. L'hypothèse d'absence d'arbitrage s'écrit : $V_0(\phi) = 0 \Rightarrow (\forall\, n \leq N)\, \tilde{V}_n(\phi) \notin \mathcal{C}$, $\mathcal{C}$ désignant ici le cône convexe des variables aléatoires ≥ 0 non presque sûrement nulles.

L'ensemble $\mathcal{V}$ des v.a. de la forme $\tilde{V}_N(\phi)$, ϕ prévisible, est un ssev de l'espace $\mathbb{R}^\Omega$ de toutes les v.a.r. définies sur Ω. L'hypothèse assure que $\mathcal{V}$ est disjoint de $\mathcal{C}$, et donc a fortiori du compact convexe $K := \{X \in \mathcal{C} \mid \sum_{\omega\in\Omega} X(\omega) = 1\} \subset \mathcal{C}$. Le théorème de séparation des convexes (bien connu en analyse, cf par exemple l'appendice du livre de Lamberton et Lapeyre) fournit une v.a. λ (prendre $\lambda \in \mathcal{V}^\perp \ldots$ et faire un dessin) telle que :

a) $\sum_{\omega} \lambda(\omega) X(\omega) > 0$ pour tout $X \in K$; b) $\sum_{\omega} \lambda(\omega)\tilde{V}_N(\phi)(\omega) = 0$ pour tout ϕ prévisible.

Comme (a) implique $\lambda(\omega) > 0$ pour tout $\omega \in \Omega$, la probabilité Q définie par $Q(\omega) := \lambda(\omega) \Big/ \sum_{\omega'} \lambda(\omega')$ est équivalente à $\mathbb{P}$. Alors (b) signifie que $\mathbb{E}^Q\Big(\sum_{j=1}^{N} \phi_j(\tilde{S}_j - \tilde{S}_{j-1})\Big) = 0$, pour tout ϕ prévisible. Cela équivaut clairement à $\mathbb{E}^Q\Big(\sum_{j=1}^{N} \phi_j^i(\tilde{S}_j^i - \tilde{S}_{j-1}^i)\Big) = 0$, pour tout $i \in \{1, \ldots, d\}$ et tout processus réel ϕ^i prévisible. La proposition 10.4 assure alors que les cours actualisés $(\tilde{S}_n^i)$ sont des Q-martingales. $\diamond$

17 Modèle de Cox-Ross-Rubinstein

Il s'agit d'une version discrétisée du modèle de Black-Scholes (qui sera l'objet de la section IV.27), dans lequel on considère un unique actif à risque $S_n^1 = S_n$, et un taux d'intérêt fixe r. Donc $d = 1$, $S_n^0 = (1 + r)^n$, et on ne s'intéresse qu'aux temps $n \in \{0, \ldots, N\}$, pour un terme N fixé a priori.

On ne considère en outre qu'une évolution du cours S_n bien particulière, supposant l'existence de deux réels a et b fixés tels que $-1 < a < b$ et $T_n := \frac{S_n}{S_{n-1}} \in \{(1+a), (1+b)\}$ pour tout n.

On suppose le cours initial S_0 connu, de sorte que $\mathcal{F}_0 = \{\emptyset, \Omega\}$, et on prend $\mathcal{T} = \mathcal{P}(\Omega)$, et $\Omega := \{a, b\}^N$, chaque $\omega \in \Omega$ représentant la suite des valeurs $(T_1 - 1, \ldots, T_N - 1)$, puis

$\mathcal{F}_n = \sigma\{S_1,\dots,S_n\} = \sigma\{T_1,\dots,T_n\}$. On suppose enfin $0 < \mathbb{P}(T_n = 1+a) < 1$ $(\forall\, n)$.

Observons d'abord que le cours actualisé $(\tilde{S}_n)$ est une $\mathbb{P}$-martingale si et seulement si

$$\mathbb{E}(\tilde{S}_{n+1}\,|\,\mathcal{F}_n) = \tilde{S}_n \Leftrightarrow \mathbb{E}\Big(\tfrac{\tilde{S}_{n+1}}{\tilde{S}_n}\,\Big|\,\mathcal{F}_n\Big) = 1 \Leftrightarrow \mathbb{E}(T_{n+1}\,|\,\mathcal{F}_n) = 1+r \quad \text{pour tout } n < N\,.$$

Pour qu'il n'existe pas de stratégie d'arbitrage, selon le théorème 16.2, il doit donc exister une probabilité équivalente à $\mathbb{P}$ pour laquelle la moyenne de T_1 vaut $(1+r)$. Comme T_1 vaut soit $(1+a)$ soit $(1+b)$ et n'est pas p.s. constant, on doit avoir $1+a < 1+r < 1+b$, id est : $a < r < b$. $\big($On peut à la date 0 acheter $S_0 = 1$ d'actif risqué, et le revendre à la date N : le profit ainsi réalisé vaut $S_N - (1+r)^N$, qui est $\geq (1+a)^N - (1+r)^N$, cette inégalité étant stricte avec une probabilité non nulle. Ceci constitue donc un arbitrage si $r \leq a$. Raisonnement analogue si $r \geq b$.$\big)$

On suppose dorénavant que $a < r < b$, et on pose $p := (b-r)/(b-a)$.

Si les T_n sont i.i.d. de loi de Bernoulli $\mathcal{B}(1+a, 1+b\,; p, 1-p)$, alors $\mathbb{E}(T_{n+1}\,|\,\mathcal{F}_n) = \mathbb{E}(T_{n+1}) = 1+r$, de sorte que $(\tilde{S}_n)$ est une $\mathbb{P}$-martingale.
Réciproquement, si $(\tilde{S}_n)$ est une $\mathbb{P}$-martingale, alors

$$\begin{aligned} 1+r = \mathbb{E}(T_{n+1}|\,\mathcal{F}_n) &= (1+a)\mathbb{E}(1_{\{T_{n+1}=1+a\}}|\,\mathcal{F}_n) + (1+b)\mathbb{E}(1_{\{T_{n+1}=1+b\}}|\,\mathcal{F}_n) \\ &= 1+b-(b-a)\mathbb{E}(1_{\{T_{n+1}=1+a\}}|\,\mathcal{F}_n)\,, \end{aligned}$$

de sorte que $\mathbb{E}(1_{\{T_{n+1}=1+a\}}|\,\mathcal{F}_n) = p$, pour tout n. Cela détermine (par une récurrence immédiate) la loi du vecteur $(T_1,\dots,T_N)$, qui doit donc être comme ci-dessus. Ceci prouve qu'il y a une unique probabilité $\mathbb{P}$ (qu'on fixera pour la suite) sur $(\Omega, \mathcal{F}_N)$ faisant de $(\tilde{S}_n)$ une $\mathbb{P}$-martingale. On dit alors que le marché est complet.

Définition 17.1 *Une option est un titre donnant à son détenteur le droit (et non le devoir) d'acheter (pour une option d'achat) ou de vendre (pour une option de vente) une certaine quantité d'un actif financier (action, obligation, devise...), avant une date (échéance) fixée à l'avance N et à un prix (prix d'exercice) fixé à l'avance K.*
Le terme anglo-saxon pour "option d'achat" est call ; et put pour "option de vente".

Si l'option ne peut être exercée qu'à la date d'échéance N, l'option est dite européenne, et si elle peut être exercée également à toute date antérieure, elle est dite américaine.

Dans le cas de l'option d'achat européenne relative à l'actif de cours (S_n), il y a clairement 2 cas :
- soit $S_N > K$, et alors l'exercice de l'option permet un profit égal à $(S_N - K)$,
- soit $S_N \leq K$, et alors il est préférable de ne pas exercer l'option.

Au total, la valeur à l'instant terminal N de cette option est $(S_N - K)^+ = \max\{S_N - K, 0\}$.
Dans le cas de l'option de vente européenne relative à l'actif de cours (S_n), le raisonnement analogue produit la valeur $(S_N - K)^- = (K - S_N)^+$.

Traitons la question du juste prix (non actualisé) C_n que doit valoir à l'instant $n < N$ une option d'achat européenne, et de même, du juste prix (non actualisé) P_n que doit valoir à l'instant $n < N$ une option de vente européenne, en fonction de (N, K) (fixés).
Notons qu'on doit avoir $C_N = (S_N - K)^+$, et que $(\tilde{C}_n)$ doit être une martingale : le juste prix à l'instant n doit être la projection sur $\mathcal{F}_n$ du juste prix à l'instant $(n+1)$.
D'où l'expression : $C_n = (1+r)^{n-N}\, \mathbb{E}[(S_N - K)^+ \,|\, \mathcal{F}_n]$; et idem pour P_n (avec $(K - S_N)^+$).
Nous avons tout d'abord la relation de parité suivante :

$$C_n - P_n = (1+r)^{n-N}\, \mathbb{E}[(S_N - K)^+ - (K - S_N)^+ \,|\, \mathcal{F}_n] = (1+r)^{n-N}\, \mathbb{E}[S_N - K \,|\, \mathcal{F}_n]$$
$$= S_n - K(1+r)^{n-N}.$$

Vérifions ensuite que C_n est une fonction de n et de S_n. En effet, tirant parti de l'indépendance des T_k vis-à-vis de $\mathcal{F}_n$ dès que $k > n$, nous avons :

$$C_n\,(1+r)^{N-n} = \mathbb{E}\left[(S_N - K)^+ \,|\, \mathcal{F}_n\right] = \mathbb{E}\left[\left(S_n \prod_{k=n+1}^{N} T_k - K\right)^+ \,\middle|\, \mathcal{F}_n\right]$$
$$= \sum_{j=0}^{N-n} C_{N-n}^{j}\, p^j (1-p)^{N-n-j}\, \mathbb{E}\left[\left(S_n\,(1+a)^j (1+b)^{N-n-j} - K\right)^+ \,\middle|\, \mathcal{F}_n\right]$$
$$= \sum_{j=0}^{N-n} C_{N-n}^{j}\, p^j (1-p)^{N-n-j} \left(S_n\,(1+a)^j (1+b)^{N-n-j} - K\right)^+ =: c(n, S_n)\,(1+r)^{N-n}.$$

Cherchons ensuite sous la forme $H_n = \gamma(n, S_{n-1})$ une stratégie (rappelons que H doit être prévisible) produisant C_n, c'est-à-dire telle que $C_n = H_n^0 (1+r)^n + H_n\, S_n$.
Cela se réécrit :

$$c\Big(n, S_{n-1}\, T_n\Big) = c(n, S_n) = H_n^0 (1+r)^n + \gamma(n, S_{n-1}) S_{n-1}\, T_n.$$

Comme T_n est indépendant des autres variables ($\mathcal{F}_{n-1}$-mesurables) de cette équation, on en déduit les deux équations :

$$c\Big(n, S_{n-1}\,(1+a)\Big) = H_n^0 (1+r)^n + \gamma(n, S_{n-1})\, S_{n-1}\,(1+a),$$

$$c\Big(n, S_{n-1}\,(1+b)\Big) = H_n^0 (1+r)^n + \gamma(n, S_{n-1})\, S_{n-1}\,(1+b).$$

Par soustraction, on obtient :

$$H_n = \gamma(n, S_{n-1}) = \frac{c\Big(n, S_{n-1}\,(1+b)\Big) - c\Big(n, S_{n-1}\,(1+a)\Big)}{(b-a)\, S_{n-1}}.$$

Conclusion Nous avons obtenu une formule explicite à la fois pour le prix (C_n) d'une option d'achat européenne, et pour une stratégie (H_n) produisant (simulant) ce prix :

$$C_n = c(n, S_n) = H_n^0 (1+r)^n + H_n \, S_n \; ; \quad H_n = \gamma(n, S_{n-1}) \, ,$$

les deux fonctions c, γ étant explicitées plus haut. En particulier à l'instant 0 :

$$C_0 = (1+r)^{-N} \, \mathbb{E} \left[\left(S_0 \prod_{k=1}^{N} T_k \, - K \right)^{+} \right] ,$$

pour des v.a.r.i.i.d. T_k , de Bernoulli de paramètre p. De même, pour une option de vente européenne, soit en utilisant la relation de parité

$$C_n - P_n = S_n - K(1+r)^{n-N},$$

soit directement en remplaçant la fonction $S \mapsto (S-K)^+$ par la fonction $S \mapsto (K-S)^+$, on obtient le prix P_n , et en particulier

$$P_0 = (1+r)^{-N} \, \mathbb{E} \left[\left(K - S_0 \prod_{k=1}^{N} T_k \right)^{+} \right] .$$

Pour les grandes valeurs de N, on peut approcher ces expressions par des intégrales gaussiennes. Cela donne les formules de Black-Scholes, objet de la section IV.27, où elles seront obtenues plus directement.

18 Arrêt optimal

Dans la section précédente, nous avons calculé le prix non actualisé (C_n) d'une option d'achat européenne (et le prix (P_n) d'une option de vente), dans le modèle de Cox-Ross-Rubinstein.

Considérons ici plutôt le prix (non actualisé) (C'_n) (respectivement (P'_n)) d'une option d'achat américaine, pour le même modèle. Notons que puisque l'option américaine offre plus de choix que l'option européenne, on aura automatiquement $C'_n \geq C_n$ et $P'_n \geq P_n$, pour tout $n \in \{0, \dots, N\}$. Plutôt qu'essayer d'évaluer C'_n , nous allons ici nous demander quand il convient d'exercer son droit d'option américaine.

Utilisons toujours la situation de marché complet, et son unique probabilité $\mathbb{P}$ faisant du cours actualisé $(\tilde{S}_n)$ une $\mathbb{P}$-martingale (même si les T_n ne sont plus nécessairement ici des variables de Bernoulli comme dans la section 17).

La valeur (i.e. le profit qu'elle permet de réaliser) non actualisée de l'option considérée est encore $Z_N = (S_N - K)^+$ (respectivement $Z_N = (K - S_N)^+$) à l'instant terminal N, mais d'après la définition d'une option d'achat américaine, on doit également considérer sa valeur $Z_n = (S_n - K)^+$ (respectivement $Z_n = (K - S_n)^+$) à tout instant $n \in \{0, \ldots, N\}$. La valeur actualisée correspondante est bien sûr $\tilde{Z}_n = (1+r)^{-n} Z_n$. Notons que les suites (Z_n) et $(\tilde{Z}_n)$ sont positives et adaptées à la filtration $(\mathcal{F}_n)$.

Nota Bene La notion de "prix" (on pourrait aussi dire : de "juste prix") utilisée ici est à comprendre comme la valeur maximale, potentielle, qu'on doit pouvoir tirer de la suite (Z_n) des valeurs, i.e. de la définition de l'option et de la fluctuation du cours S_n. Elle correspond à la situation de "couverture" minimale du vendeur de l'option.

Le prix terminal doit évidemment être la valeur terminale (non actualisée) :

$C'_N = C_N = Z_N$. Par récurrence, quel doit être le prix (non actualisé) C'_{n-1}, connaissant le prix (non actualisé) C'_n ?

Puisque l'option peut être exercée à l'instant $(n-1)$, le prix non actualisé est le maximum de la valeur Z_{n-1} à cet instant et du montant qu'il aurait sans cette possibilité à l'instant $(n-1)$, à savoir (comme dans le cas d'une option européenne considérée entre les instants $(n-1)$ et n) : $(1+r)^{-1}\mathbb{E}(C'_n \,|\, \mathcal{F}_{n-1})$. On obtient donc :

$$C'_{n-1} = \max\big\{Z_{n-1}\,, (1+r)^{-1}\mathbb{E}(C'_n \,|\, \mathcal{F}_{n-1})\big\}.$$

Utilisant les prix actualisés $\tilde{C}'_n = (1+r)^{-n}\, C'_n$, ceci se réécrit de façon équivalente :

$$\tilde{C}'_{n-1} = \max\big\{\tilde{Z}_{n-1}\,, \mathbb{E}(\tilde{C}'_n \,|\, \mathcal{F}_{n-1})\big\}.$$

Proposition 18.1 *(i) Le processus $\{\tilde{C}'_n \,|\, 0 \le n \le N\}$ est la plus petite surmartingale majorant le processus $\{\tilde{Z}_n \,|\, 0 \le n \le N\}$.*

(ii) $\tau := \min\Big\{n \in \{0, \ldots, N\} \,\Big|\, \tilde{C}'_n = \tilde{Z}_n\Big\}$ est un temps d'arrêt, et la surmartingale arrêtée $\{\tilde{C}'_{\min\{n,\tau\}} | \, 0 \le n \le N\}$ est une martingale.

(iii) τ est un temps d'arrêt optimal : $\tilde{C}'_0 = \mathbb{E}(\tilde{Z}_\tau \,|\, \mathcal{F}_0) = \sup_{\tau'} \mathbb{E}(\tilde{Z}_{\tau'} \,|\, \mathcal{F}_0)$, où τ' varie parmi l'ensemble de tous les temps d'arrêt à valeurs dans $\{0, \ldots, N\}$.

(iv) τ est le plus petit des temps d'arrêt optimaux (i.e., des temps d'arrêt $\tilde{\tau}$ qui vérifient $\mathbb{E}(\tilde{Z}_{\tilde{\tau}} \,|\, \mathcal{F}_0) = \sup_{\tau'} \mathbb{E}(\tilde{Z}_{\tau'} \,|\, \mathcal{F}_0)$).

Preuve (*i*) On a évidemment pour tout n : $\tilde{C}'_n \ge \tilde{Z}_n$, et $\mathbb{E}(\tilde{C}'_n \,|\, \mathcal{F}_{n-1}) \le \tilde{C}'_{n-1}$. Soit (M_n) une surmartingale majorant $(\tilde{Z}_n)$. Nous avons d'une part $M_N \ge \tilde{Z}_N = \tilde{C}'_N$, et d'autre part, si $M_n \ge \tilde{C}'_n$, alors

$$\tilde{C}'_{n-1} = \max\{\tilde{Z}_{n-1}\,, \mathbb{E}(\tilde{C}'_n \,|\, \mathcal{F}_{n-1})\} \le \max\{M_{n-1}\,, \mathbb{E}(M_n \,|\, \mathcal{F}_{n-1})\} \le M_{n-1}\,.$$

(*ii*) Il est clair que $\tau \le N$, et que τ est un temps d'arrêt, comme temps d'atteinte de 0 par le processus adapté $\tilde{C}'_n - \tilde{Z}_n$. Ensuite, notant $C''_n := \tilde{C}'_{\min\{n,\tau\}}$, nous avons :

$$C''_n - C''_{n-1} = 1_{\{\tau \ge n\}}(\tilde{C}'_n - \tilde{C}'_{n-1}) = (1 - 1_{\{\tau \le n-1\}})\Big(\tilde{C}'_n - \mathbb{E}(\tilde{C}'_n \,|\, \mathcal{F}_{n-1})\Big),$$

d'où le résultat en prenant l'espérance conditionnelle par rapport à $\mathcal{F}_{n-1}$.

(iii) Par (ii) et le théorème d'arrêt 11.3, et puisque τ' peut valoir en particulier τ, nous avons aussitôt :

$$\tilde{C}'_0 = \mathbb{E}(\tilde{C}'_\tau \,|\, \mathcal{F}_0) = \mathbb{E}(\tilde{Z}_\tau \,|\, \mathcal{F}_0) \leq \sup_{\tau'} \mathbb{E}(\tilde{Z}_{\tau'} \,|\, \mathcal{F}_0).$$

Et pour tout τ', $\tilde{C}'_0 \geq \mathbb{E}(\tilde{C}'_{\tau'} \,|\, \mathcal{F}_0) \geq \mathbb{E}(\tilde{Z}_{\tau'} \,|\, \mathcal{F}_0)$, par (i) et le théorème d'arrêt 11.3.

(iv) Soit $\tilde{\tau}$ un autre temps d'arrêt optimal. (i), (iii), et le théorème d'arrêt entraînent : $\mathbb{E}(\tilde{C}'_{\tilde{\tau}} \,|\, \mathcal{F}_0) \leq \tilde{C}'_0 = \mathbb{E}(\tilde{Z}_{\tilde{\tau}} \,|\, \mathcal{F}_0) \leq \mathbb{E}(\tilde{C}'_{\tilde{\tau}} \,|\, \mathcal{F}_0)$, et donc $\mathbb{E}(\tilde{Z}_{\tilde{\tau}} \,|\, \mathcal{F}_0) = \mathbb{E}(\tilde{C}'_{\tilde{\tau}} \,|\, \mathcal{F}_0)$. Puisque $\tilde{Z}_{\tilde{\tau}} \leq \tilde{C}'_{\tilde{\tau}}$, cela impose : $\tilde{Z}_{\tilde{\tau}} = \tilde{C}'_{\tilde{\tau}}$. Par définition de τ, on doit donc avoir $\tau \leq \tilde{\tau}$. $\diamond$

Conclusion À quel instant est-il préférable d'exercer son droit d'option américaine d'achat ? Il faut éviter tout instant n tel que $\tilde{C}'_n > \tilde{Z}_n$, pour ne pas perdre l'option dont la valeur potentielle réelle (le "prix", comme noté plus haut) est $\tilde{C}'_n$, contre une valeur effective $\tilde{Z}_n$ (une fois l'option exercée). Donc il faut préférer un temps n tel que $\tilde{C}'_n = \tilde{Z}_n$. En l'absence de délit d'initié, on ne peut considérer que des temps d'arrêt. Cela conduit en particulier au temps d'arrêt τ de la proposition 18.1. Du même coup, cela justifie l'appellation de "temps d'arrêt optimal". Une aute façon de dire est qu'on doit choisir d'acheter en un temps optimisant la valeur, ce qui correspond bien à la définition (dans (iv) de la proposition 18.1) d'un temps d'arrêt optimal. En outre, pour optimiser au plus tôt il convient de privilégier un temps d'exercice minimal. Au total, le temps optimal τ de la proposition 18.1 ressort comme le meilleur des choix possibles.

III. Mouvement Brownien

19 Mouvement Brownien réel

L'existence du mouvement brownien peut s'obtenir par limite, à partir d'une suite de marches aléatoires simples dont le pas élémentaire tend vers 0 (voir le théorème de Donsker 19.7 ci-dessous). Commençons par une définition formelle.

Définition 19.1 *On appelle mouvement brownien réel standard tout processus* $(B_t \,|\, t \in \mathbb{R}_+)$ *qui est à la fois :*

- à accroissements indépendants : pour toute subdivision $0 = t_0 < t_1 < \ldots < t_N < \infty$, *les* N *variables aléatoires* $(B_{t_j} - B_{t_{j-1}})$ *sont indépendantes ;*

- à accroissements stationnaires : pour tous $0 < s < t$, $(B_t - B_s)$ *a la même loi que* B_{t-s} *;*

- tel que pour tout $t > 0$, *la loi de* B_t *est* $\mathcal{N}(0,t)$ *(gaussienne centrée de variance* t*) :*

$$\mathbb{P}[B_t > y] = \int_y^\infty e^{-x^2/(2t)} \frac{dx}{\sqrt{2\pi\, t}} \,; \quad \textit{ou encore :} \quad \mathbb{E}[f(B_t)] = \int_{-\infty}^\infty f(x)\, e^{-x^2/(2t)} \frac{dx}{\sqrt{2\pi\, t}} \,;$$

- continu : $\lim_{s \to t} B_s = B_t$ *, pour tout* $t \in \mathbb{R}_+$ *;*

- issu de 0 : $B_0 = 0$ *.*

Sa loi est une loi sur l'espace $\mathcal{C}_0^0(\mathbb{R}_+, \mathbb{R})$ *des fonctions de* $\mathbb{R}_+$ *dans* $\mathbb{R}$ *continues et nulles en 0, appelée mesure de Wiener standard.*

Cela entraîne en particulier l'énoncé suivant.

Proposition 19.2 *Le mouvement brownien réel standard* (B_t) *est gaussien et centré : pour toute subdivision* $0 \leq t_1 < \ldots < t_N < \infty$ *, toute combinaison linéaire* $\sum_{j=1}^{N} \lambda_j B_{t_j}$ *à coefficients réels* $\lambda_1, \ldots, \lambda_N$ *constants est une variable aléatoire gaussienne centrée.*

Voici une caractérisation des mouvements browniens réels standards, due au fait que la loi d'un processus gaussien est déterminée par sa moyenne et sa fonction de covariance :

Proposition 19.3 *Tout processus réel* (B_t) *sur* $\mathbb{R}_+$ *qui soit gaussien centré de covariance* $(s,t) \longmapsto \mathbb{E}(B_s B_t) = \min\{s,t\}$ *est un mouvement brownien réel standard.*

Les processus $t \mapsto B_{a+t} - B_a$, $t \mapsto c^{-1} B_{c^2 t}$, et $t \mapsto t\, B_{1/t}$ vérifient les conditions de la proposition 19.3 ci-dessus. On en déduit aussitôt les propriétés fondamentales suivantes :

Corollaire 19.4 *Tout mouvement brownien réel standard* (B_t) *vérifie :*

1) la propriété de Markov (faible) : pour tout $a \in \mathbb{R}_+$ *,* $(B_{a+t} - B_a)$ *est encore un mouvement brownien réel standard, indépendant de* $\mathcal{F}_a := \sigma(\{B_s \,|\, 0 \leq s \leq a\})$ *;*

2) l'autosimilarité ou invariance d'échelle : pour tout $c > 0$ *,* $(c^{-1} B_{c^2 t})$ *est encore un mouvement brownien réel standard ;*

3) $(-B_t)$ *et* $(t\, B_{1/t})$ *sont aussi des mouvements browniens réels standards.*

La propriété d'autosimilarité ci-dessus signifie que la courbe brownienne (définie par toute trajectoire $(B_t \,|\, t \geq 0)$) est un ensemble aléatoire fractal.

Exercice n^o 19.1 Vérifier les propositions 19.2 et 19.3, et le corollaire 19.4.

La propriété de Markov ci-dessus s'étend aux temps d'arrêt (définis en temps continu comme en temps discret ; voir la section 21 ci-dessous), comme suit.

Proposition 19.5 *(Propriété de Markov forte) Pour tout temps d'arrêt (de la filtration de B) fini T , le processus* $(B_{T+t} - B_T)$ *est encore un brownien réel standard, indépendant de la tribu* $\mathcal{F}_T$ *des événements antérieurs à T.*

Preuve Le résultat se déduit sans difficulté de la propriété de Markov faible lorsque $T = \sum_{j \in \mathbb{N}} \alpha_j 1_{A_j}$, avec $A_j \in \mathcal{F}_{\alpha_j}$ pour chaque j . Dans le cas général on considère une suite T_n de temps d'arrêt du type précédent qui décroît vers T (voir juste avant la proposition 21.5) : on obtient ainsi une suite de browniens standards indépendants de $\mathcal{F}_T$ qui converge (uniformément sur les compacts) vers $(B_{T+t} - B_T)$. $\diamond$

19.6 Construction-approximation du brownien réel

Précisons ici comment construire ou bien approcher le mouvement brownien réel standard. Cela peut servir notamment à des simulations numériques du mouvement brownien.

19.6.1 Le théorème de Donsker

Soit $S_k = X_1 + \ldots + X_k$ une marche aléatoire sur $\mathbb{Z}$ (i.e.: les v.a. à valeurs entières X_j sont i.i.d.), de pas $X_j \in L^4$ et centré. Notons σ^2 la variance de (X_j). Pour $n \in \mathbb{N}^*$, $t \in \mathbb{R}_+$, posons

$$S_t^n := \frac{1}{\sigma\sqrt{n}}\Big(S_{[nt]} + (nt - [nt])\,X_{[nt]+1}\Big)\,.$$

S^n est un processus continu, affine sur chaque intervalle de temps $[k/n, (k+1)/n]$. Notons que dans cette normalisation la variance de S_t^n est (environ) t, et que par conséquent c'est la bonne normalisation pour obtenir une convergence lorsque $n \to \infty$. De fait, nous avons le résultat essentiel suivant (de convergence en loi).

Théorème 19.7 *(de Donsker) Il existe un mouvement brownien réel standard (B_t) et une suite de processus $\tilde{S}_t^n$ tels que : (1) pour chaque n $\tilde{S}^n$ et S^n aient la même loi ; (2) pour tout $T \in \mathbb{R}_+$ on ait $\sup\limits_{0\le t\le T} |\tilde{S}_t^n - B_t| \xrightarrow{p.s.} 0$ lorsque $n \to \infty$.*

19.7.1 Une expression multiéchelle

Théorème 19.8 *Soit $\{\varphi_k \,|\, k \in \mathbb{N}^*\}$ une base orthonormée de $L_0^2([0,1], \mathbb{R})$ (fonctions de carré intégrable et d'intégrale nulle sur $[0,1]$), telle que la série $\sum\limits_{k=1}^{\infty} \phi_k^2$ converge uniformément sur $[0,1]$ $\big($où ϕ_k est la primitive de φ_k nulle en 0, et donc en $1\big)$. Soit $\{\xi_k \,|\, k \in \mathbb{N}\}$ une suite de v.a.i.i.d. de loi normale $\mathcal{N}(0,1)$. Alors*

$$B_t := \xi_0\, t + \sum_{k\in\mathbb{N}^*} \xi_k\, \phi_k(t)\,, \qquad \text{pour tout } t \in [0,1]\,,$$

définit un mouvement brownien réel standard sur $[0,1]$.

<u>Justification rapide</u> La convergence uniforme de la série $\sum\limits_k \phi_k^2$ assure que la série $\sum\limits_k \xi_k\phi_k$ converge uniformément dans L^2, vers un processus gaussien continu, dont la covariance est $\sum\limits_k \phi_k(s)\phi_k(t) = s \wedge t - st$, id est celle du processus gaussien continu $(B_t - t\,B_1)$.

On a en effet plus généralement la formule de Parseval : pour toute fonction f dérivable sur $[0,1]$ nulle en 0 et en 1,

$$f = \sum_{k\in\mathbb{N}^*} \Big(\int_0^1 f'\varphi_k\Big)\,\phi_k\ .$$

Appliquant ceci à $t \mapsto s\wedge t - st$, nous obtenons bien $\quad s\wedge t - st = \sum_k \phi_k(s)\phi_k(t)$.

Exemples : 1) la suite trigonométrique : cela donne l'écriture du brownien réel en Fourier :

$$B_t = \xi_0\,t + \frac{1}{\pi\sqrt{2}}\sum_{k\in\mathbb{N}^*}\frac{1}{k}\Big(\xi_{2k}\,(1-\cos[2\pi kt]) + \xi_{2k-1}\sin[2\pi kt]\Big) \quad \text{p.s. ;}$$

$$\text{ou encore :}\qquad B_t = \xi_0\,t + \frac{\sqrt{2}}{\pi}\sum_{k\in\mathbb{N}^*}\frac{\xi_k}{k}\sin[\pi kt] \quad \text{p.s. ;}$$

2) une suite d'ondelettes, la base de Haar :

$\varphi_{k,j} := 2^{k/2}\Big(1_{\left[(j-1)2^{-k},\left(j-\frac{1}{2}\right)2^{-k}\right]} - 1_{\left]\left(j-\frac{1}{2}\right)2^{-k},\,j2^{-k}\right]}\Big)$, pour $k\in\mathbb{N}$, $1\le j\le 2^k$; alors

$\phi_{k,j}(t) = 2^{-k/2}\,\Phi(2^k t - j + 1)$, avec $(\forall u\in\mathbb{R})$ $\Phi(u) := 1_{\left[0,\frac{1}{2}\right]}(u)\times u + 1_{\left]\frac{1}{2},1\right]}(u)\times(1-u)$.

D'où $\quad B_t = \xi_0\,t + \sum_{k\in\mathbb{N}}\sum_{j=1}^{2^k}\xi_{k,j}\,\phi_{k,j}(t)\quad$ *p.s.* avec $\{\xi_0\}\cup\{\xi_{k,j}|k\in\mathbb{N}, 1\le j\le 2^k\}$ suite i.i.d. $\mathcal{N}(0,1)$.

Nota Bene Le processus gaussien continu $(B_t - t\,B_1)$, nul aux temps 0 et 1, est appelé pont brownien.

19.9 Autres mouvements browniens réels

Les mouvements browniens réels sont simplement déduits d'un mouvement brownien réel standard par transformation affine : $B_t \mapsto \sigma\,B_t + m$. Ainsi la moyenne devient $B_0 = m\in\mathbb{R}$, et la variance de B_t devient $\sigma^2\,t>0$ (pour $t>0$).

Une notion légèrement plus générale est celle de $(\mathcal{F}_t)$-mouvement brownien (réel), étant donnée une filtration a priori $(\mathcal{F}_t)$ sur $(\Omega,\mathcal{F},\mathbb{P})$.

Définition 19.10 *Pour toute filtration continue à droite* $(\mathcal{F}_t)$, *on appelle* $(\mathcal{F}_t)$*-mouvement brownien réel tout processus* (X_t) *réel* $(\mathcal{F}_t)$*-adapté p.s. continu, nul en* 0, *tel que pour tous* $0\le s<t$ $(X_t - X_s)$ *soit indépendante de* $\mathcal{F}_s$ *et ait pour loi* $\mathcal{N}(0,t-s)$.

Notons qu'un $(\mathcal{F}_t)$-mouvement brownien (réel) (X_t) est automatiquement un mouvement brownien dans sa filtration propre $\mathcal{F}'_t := \sigma\{X_s\,|\,0\le s\le t\}$, et que le mouvement brownien vu précédemment (construit par 19.6.1 ou 19.7.1) est un exemple de $(\mathcal{F}'_t)$-mouvement brownien.

20 Propriétés trajectorielles du mouvement brownien

La définition 19.1 entraîne beaucoup de propriétés. En voici quelques unes relatives à l'allure des trajectoires browniennes, parmi les plus frappantes.

Proposition 20.1 *(i) (B_t) passe p.s. une infinité de fois par toute valeur réelle.*
(ii) Les trajectoires de (B_t) ne sont monotones sur aucun sous-intervalle de temps.

Proposition 20.2 *Fixons $t \in \mathbb{R}_+$, et notons $\mathcal{P} := \{0 = t_0 < t_1 < \ldots < t_N = t\}$ une subdivision de $[0,t]$, de pas $|\mathcal{P}|$. Soit $V_{\mathcal{P}} := \sum_{j=0}^{N-1} (B_{t_{j+1}} - B_{t_j})^2$. Lorsque $|\mathcal{P}| \to 0$, $V_{\mathcal{P}}$ converge vers t dans L^2, et c'est aussi vrai presque sûrement si on se restreint à une suite croissante de subdivisions.*

Corollaire 20.3 *Les trajectoires browniennes sont presque sûrement à variation infinie (voir la définition 22.1, plus loin) sur chaque intervalle de temps non trivial.*

<u>Preuve</u> Fixons une suite croissante de subdivisions (t_j^n) de $[0,t]$ dont le pas tend vers 0. Utilisant la proposition 20.2, nous avons (p.s.) :

$$t = \lim_n V_{\mathcal{P}_n} = \lim_n \sum_{j=0}^{N_n-1} (B_{t_{j+1}^n} - B_{t_j^n})^2 \leq Var(B,[0,t]) \times \limsup_n \sup_j |B_{t_{j+1}^n} - B_{t_j^n}| \,,$$

et donc $t = 0$ si $Var(B,[0,t])$ est finie. La propriété de Markov montre que de même p.s. $Var(B,[a,b]) = \infty$ pour $a < b \in \mathbb{Q}_+$. $\diamond$

Le résultat suivant montre qu'on ne peut pas vraiment tracer une trajectoire brownienne.

Proposition 20.4 *Les trajectoires browniennes ne sont (p.s.) dérivables en aucun point.*

Il y a tout de même une certaine forme de régularité de ces trajectoires.

Proposition 20.5 *Les trajectoires browniennes sont p.s. localement höldériennes d'ordre α pour tout $\alpha < \frac{1}{2}$ · Id est : p.s. pour tout $T > 0$ il existe $A > 0$ tel que $0 \leq s,t \leq T \Longrightarrow |B_s - B_t| \leq A\,|s-t|^\alpha$.*

La régularité relative de ces trajectoires est précisée comme suit.

Proposition 20.6 *(module de continuité de Lévy) Nous avons presque sûrement :*

$$\limsup_{\varepsilon \searrow 0} \sup_{0 \leq s < t \leq 1\,;\, t-s \leq \varepsilon} \frac{|B_s - B_t|}{\sqrt{2\varepsilon \,\log(1/\varepsilon)}} = 1 \,.$$

Les trajectoires du brownien réel ne sont donc p.s. pas localement höldériennes d'ordre $\frac{1}{2}$ ·

Le module de continuité de Lévy des trajectoires browniennes ci-dessus est uniforme. Le résultat suivant montre en particulier que le module de continuité local (en un point) de ces trajectoires n'est pas le même.

Proposition 20.7 *(Loi du logarithme itéré) Nous avons presque sûrement :*

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\log(\log t)}} = 1 = -\liminf_{t\to\infty} \frac{B_t}{\sqrt{2t\log(\log t)}}\ ,$$

$$\text{et} \qquad \limsup_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log(1/t)}} = 1 = -\liminf_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log(1/t)}}\ .$$

21 Martingales à temps continu

Considérons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, muni d'une filtration continue à droite $(\mathcal{F}_t)$, i.e. une famille croissante de sous-tribus de $\mathcal{F}$, telle que $\mathcal{F}_{t+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon} = \mathcal{F}_t$.

Définition 21.1 *Une $(\mathcal{F}_t)$-martingale est un processus réel $\{X_t \,|\, t \in \mathbb{R}_+\}$ tel que :*
i) X_t est $\mathcal{F}_t$-mesurable (cela définit un processus adapté) et intégrable : $X_t \in L^1(\mathcal{F}_t\,, \mathbb{P})$;
ii) $X_t = \mathbb{E}(X_{t+s} \,|\, \mathcal{F}_t)$ presque sûrement, pour tous $s, t \geq 0$.
Si dans ii), $=$ est remplacé par $\leq$ ou par $\geq$, le processus (X_t) est nommé sousmartingale ou surmartingale, respectivement.

Pour imaginer une martingale, une sousmartingale ou une surmartingale, penser à un jeu d'argent équitable, favorable ou défavorable, en temps continu (de façon approchée, lorsque le jeu va très vite ; penser à un marché financier).

Noter que pour une sousmartingale, $t \mapsto \mathbb{E}(X_t)$ croît (au sens large), tandis que pour une surmartingale, $t \mapsto \mathbb{E}(X_t)$ décroît (au sens large).

Les propriétés des martingales discrètes s'étendent aux martingales en temps continu.

Exemples: - (voir l'exercice n^o 10.0) si (X_t) est une sousmartingale, alors $(X_t - a)^+$ est une sousmartingale ; et si (X_t) est une surmartingale, alors $(t \mapsto \min\{X_t\,, a\})$ est une surmartingale (pour tout a réel).

- Si Z est intégrable, alors $\mathbb{E}(Z \,|\, \mathcal{F}_t)$ définit une $(\mathcal{F}_t)$-martingale.

- Le mouvement brownien (B_t) fournit des martingales (de carré intégrable). Nous avons en effet l'importante propriété suivante.

Proposition 21.2 *Si (X_t) est un $(\mathcal{F}_t)$-mouvement brownien, alors (X_t), $(X_t^2 - t)$, et (pour tout λ réel) la "martingale exponentielle" $(\exp[\lambda X_t - \lambda^2 t/2])$ sont des $(\mathcal{F}_t)$-martingales.*

Réciproquement, si (X_t) un processus continu nul en 0 tel que pour tout λ réel $(\exp[\lambda X_t - \lambda^2 t/2])$ est une $(\mathcal{F}_t)$-martingale, alors (X_t) est un $(\mathcal{F}_t)$-mouvement brownien.

Preuve Pour le sens direct, ce sont des petits calculs simples fondés sur l'indépendance de $(X_t - X_s)$ et de $\mathcal{F}_s$ et sur la loi de Gauss. Réciproquement, la condition de l'énoncé entraîne que $\mathbb{E}\Big(\exp[\lambda(X_t - X_s)]\Big|\,\mathcal{F}_s\Big) = \exp[\lambda^2(t-s)/2]$, et donc par injectivité de la transformée de Laplace (relativement à $\mathbb{P}(\cdot|\mathcal{F}_s)$) que $(X_t - X_s)$ est indépendant de $\mathcal{F}_s$, puis que $\mathcal{N}(0, t-s)$ est bien la loi de $(X_t - X_s)$. $\diamond$

Exercice n^o 21.1 Vérifier le sens direct de la proposition 21.2.

Une surmartingale ne peut pas trop osciller. La remarque suivante peut être déduite de l'étude des surmartingales discrètes.

Remarque 21.3 Soit $\{X_t \,|\, t \in \mathbb{Q}_+\}$ une surmartingale. Alors presque sûrement elle admet en tout $t \in \mathbb{Q}_+$ une limite à gauche (si $t \neq 0$) et une limite à droite.
Soit $\{X_t \,|\, t \in \mathbb{R}_+\}$ une surmartingale continue à droite. Alors presque sûrement elle est bornée sur tout intervalle borné et admet en tout $t > 0$ une limite à gauche.

Ceci n'exclut nullement l'existence de sauts. Mais nous n'en considérerons pas.

Nota Bene *Nous ne considérerons que des martingales continues.*

Le théorème suivant se déduit du théorème 12.1, par approximations dyadiques.

Théorème 21.4 *(Inégalités de Doob) Soit (X_t) une sousmartingale continue. On a :*

i) Pour tous $\lambda, t > 0$: $\lambda\, \mathbb{P}\Big[\sup_{s \le t} X_s^+ \ge \lambda\Big] \le \mathbb{E}(X_t^+)$;

ii) si (X_t) est positive ou bien est une martingale, alors :

$$\Big\| \sup_{s \le t} |X_s| \Big\|_2 \le 2\, \|X_t\|_2 \quad \textit{pour tout temps constant } t\,, \quad \textit{et} \quad \Big\| \sup_t |X_t| \Big\|_2 \le 2 \sup_t \|X_t\|_2 \,.$$

Exemple : Utilisant la martingale exponentielle de la proposition 21.2, nous avons par le théorème 21.4(i) : pour tous $a, b, t > 0$, $\mathbb{P}\Big(\sup_{0 \le s \le t} (B_s - as/2) \ge b\Big) \le e^{-ab}$.

Les temps d'arrêt se définissent en temps continu exactement comme en temps discret : un $(\mathcal{F}_t)$-temps d'arrêt est une variable aléatoire S à valeurs dans $\overline{\mathbb{R}}_+$ telle que $\{S \le t\} \in \mathcal{F}_t$ pour tout $t > 0$. Cela implique $\{S < t\} \in \mathcal{F}_t$ pour tout t, et c'est en fait équivalent, puisque la filtration est continue à droite.

Des exemples classiques de temps d'arrêt sont les temps d'atteinte d'ensembles ouverts par des processus adaptés continus.

Il est facile de vérifier que si S et T sont des temps d'arrêt, alors leur supremum et leur infimum sont aussi des temps d'arrêt. Et toute limite monotone d'une suite de temps d'arrêt est un temps d'arrêt.

Un temps d'arrêt peut toujours être approché uniformément par une suite décroissante de temps d'arrêt prenant leurs valeurs dans un ensemble discret : pour approcher S, il suffit de considérer $S_n := \sum\limits_{k\in\mathbb{N}} \frac{k+1}{2^n}\, 1_{\{\frac{k}{2^n}\leq S<\frac{k+1}{2^n}\}}$.

Exercice n^o 21.2   Vérifier toutes les assertions suivant le théorème 21.4.

La proposition 11.2 s'étend au cas continu.

Proposition 21.5 *Si (X_t) est une surmartingale continue à droite et si T est un temps d'arrêt (relativement à la filtration $(\mathcal{F}_t)$), alors la "surmartingale arrêtée" $(X_{t\wedge T})$ est encore une surmartingale.*

Le théorème d'arrêt 11.3 de Doob s'étend aussi au cas continu.

Théorème 21.6 *Soit $\{X_t\,|\,t\in\mathbb{R}_+\}$ une surmartingale continue à droite. Soient S et T deux temps d'arrêt p.s. bornés, tels que p.s. $S\leq T$. Alors on a p.s. $\mathbb{E}(X_T\,|\,\mathcal{F}_S)\leq X_S$.*

Réciproquement, si un processus adapté intégrable (X_t) vérifie p.s. $\mathbb{E}(X_T)\leq\mathbb{E}(X_S)$ pour tout couple $S\leq T$ de temps d'arrêt p.s. bornés, alors c'est une surmartingale.

Preuve (abrégée)   La seconde assertion a déjà été vue au théorème 11.3, de même que la première lorsque S et T sont à valeurs discrètes. Le résultat est donc assuré pour chaque couple (S_m,T_n) tel que $m\geq n$, où S_n,T_n désignent les approximations mentionnées ci-dessus. On a en effet nécessairement p.s. $S_m\leq S_n\leq T_n\leq A$. La conclusion s'obtient (moyennant quelques points à régler) par continuité à droite, en faisant tendre d'abord m vers l'infini, puis n. $\diamond$

Le théorème de convergence 12.2 s'étend de même du cas discret au cas continu.

Théorème 21.7 *Soit $\{X_t\,|\,t\in\mathbb{R}_+\}$ une surmartingale continue à droite, telle que $\sup\limits_t \mathbb{E}(X_t^-)<\infty$. Alors (X_t) est bornée dans L^1, et converge presque sûrement vers une variable intégrable X_∞.*

Idem pour le corollaire 12.3 (cas discret).

Corollaire 21.8 *Toute surmartingale $\{X_t\,|\,t\in\mathbb{R}_+\}$ continue à droite et p.s. positive (id est p.s. $X_t\geq 0\ \forall t$) converge presque sûrement vers une variable intégrable X_∞, et on a pour tout $n\in\mathbb{N}$: $\mathbb{E}(X_\infty\,|\,\mathcal{F}_t)\leq X_t$.*

Corollaire 21.9 *Soient $p > 1$ et (X_t) une martingale bornée dans L^p. Alors (X_t) converge presque sûrement et dans L^p, et on a pour tout $t \in \mathbb{R}_+$:* $\quad \mathbb{E}(X_\infty \,|\, \mathcal{F}_t) = X_t$.

On a souvent besoin d'élargir un peu la notion de martingale.

Définition 21.10 *Un processus continu M est une martingale locale par rapport à $(\mathcal{F}_t, \mathbb{P})$ s'il est adapté et s'il existe une suite de temps d'arrêt (T_n) croissant vers l'infini qui "réduit" M, au sens où pour chaque n le processus $(M_{t\wedge T_n})$ est une martingale.*

Exercice n^o 21.3 Vérifier que dans la définition 21.10 on peut prendre la martingale $(M_{t\wedge T_n})$ bornée.

IV. Intégrale stochastique

Commençons par une généralisation classique de l'intégrale de Riemann.

22 Intégrale de Stieltjes

Nous ne considérons que des fonctions réelles continues à droite ("càd") sur $\mathbb{R}_+$.

Définition 22.1 *Une fonction $A = (t \mapsto A_t)$ continue à droite de $\mathbb{R}_+$ dans $\mathbb{R}$ est à variation finie lorsque pour tout $t \in \mathbb{R}_+^*$ on a $Var(A, [0,t]) := \sup_{\mathcal{P}_t} \sum_{j=1}^{N} |A_{t_j} - A_{t_{j-1}}| < \infty$, où $\mathcal{P}_t$ désigne l'ensemble des subdivisions finies $(0 = t_0 < \ldots < t_N = t)$ de $[0,t]$.*
Sa variation totale est alors la fonction croissante de $\mathbb{R}_+$ dans $\mathbb{R}_+$ définie par $t \mapsto Var(A, [0,t])$.

Remarque 22.2 (i) Une fonction A est à variation finie ssi elle est différence de 2 fonctions croissantes $\big($pour le sens direct, écrire $A = Var(A) - [Var(A) - A]\big)$.
(ii) Une fonction A à variation finie se décompose en partie continue et partie à sauts : $A_t = A_t^c + \sum_{s\le t} \Delta A_s$, où $\Delta A_s := A_s - A_{s-}$, la somme étant dénombrable.

Exemples - Toute fonction de classe C^1 sur $\mathbb{R}_+$ est à variation finie (via l'inégalité des accroissements finis), avec $Var(A,[0,t]) \le t \times \sup_{0\le s\le t} |A'_s|$.
- Les trajectoires browniennes ne sont pas à variation finie (corollaire 20.3).

Les fonctions à variation finie sont des intégrateurs naturels, en procédant par sommes de Riemann (comme on le fait surtout pour la fonction identique : $A_t \equiv t$).

Définition 22.3 *La* <u>*somme de Riemann*</u> *associée à une fonction à variation finie A, une subdivision finie $(0 = t_0 < \ldots < t_N)$ de $[0,t]$, et une fonction ("intégrande") f borélienne localement bornée sur $I\!R_+$ (i.e. bornée sur $[0,t]$ pour tout t) est :* $\sum_{j=1}^{N} f(t_j)\,(A_{t_j} - A_{t_{j-1}})$.

Rappel : le <u>*pas*</u> *de la subdivision $\mathcal{P}_t$ ci-dessus est* $|\mathcal{P}_t| := \max_{1\le j\le N}(t_j - t_{j-1})$.

L'<u>*intégrale de Riemann-Stieltjes*</u> *de f (borélienne localement bornée) par rapport à la fonction à variation finie A est la fonctionnelle qui à $t \in I\!R_+$ associe la limite, lorsque le pas tend vers 0 et lorsqu'elle existe, des sommes de Riemann ci-dessus :*

$$\int_0^t f\,dA \equiv \int_0^t f(s)\,dA_s := \lim_{|\mathcal{P}_t|\to 0} \sum_{j=1}^{N} f(t_j)\,(A_{t_j} - A_{t_{j-1}})\,.$$

Bien entendu, comme dans le cas de l'intégrale de Riemann (qui correspond à $A_t \equiv t$) d'une fonction continue, l'intégrale de Riemann-Stieltjes (ou de Stieltjes tout court) est une fonctionnelle linéaire de l'intégrande f, et vérifie la relation de Chasles. Elle vérifie de plus d'autres propriétés classiques de l'intégrale de Riemann.

Proposition 22.4 *(Intégration par parties ; analogue à la proposition 14.2) Pour A et B à variation bornée, (désignant par A_{s-} la limite à gauche en s de A) nous avons :*

$$A_tB_t - A_0B_0 = \int_0^t A_s\,dB_s + \int_0^t B_{s-}\,dA_s = \int_0^t A_{s-}\,dB_s + \int_0^t B_{s-}\,dA_s + \sum_{s\le t} \Delta A_s\,\Delta B_s\,.$$

L'énoncé suivant (analogue à la proposition 14.2) annonce la formule d'Itô (théorème 24.5 ci-dessous). Noter qu'il devient très simple (analogue au cas de $A_t \equiv t$) si A est continu.

Proposition 22.5 *Pour F de classe C^1 et A à variation finie, $F\circ A$ est à variation finie, et* $$F(A_t) - F(A_0) = \int_0^t F'(A_{s-})\,dA_s + \sum_{s\le t}\Big(F(A_s) - F(A_{s-}) - F'(A_{s-})\Delta A_s\Big)\,.$$

<u>Exemple</u> (analogue à la proposition 14.3) : Pour A càd à variation finie, la seule solution localement bornée de l'équation $dX_t = X_{t-}\,dA_s$ est $X_t = X_0\times\prod_{s\le t}(1+\Delta A_s)\times\exp(A_t^c - A_0^c)$.

23 Intégrale d'Itô

Il s'agit de définir la notion d'intégrale stochastique en temps continu. C'est la célèbre intégrale d'Itô. Le problème est de donner du sens à des expressions de la forme $\int_0^t H_s\,dB_s$, où (B_s) est un mouvement brownien réel, qui n'est differentiable nulle part et a une variation infinie sur tout intervalle ouvert (corollaire 20.3), de sorte que la section précédente 22 ne peut pas s'appliquer.

La solution consiste à restreindre convenablement la classe des processus (H_s) à intégrer.

Considérons un espace de probabilité $(\Omega, \mathcal{F}, I\!\!P)$, muni d'une filtration $(\mathcal{F}_t)$ continue à droite. Nous dirons qu'un processus stochastique réel (H_s) est progressif ssi pour tout $t > 0$, l'application $(\omega, s) \mapsto H_s(\omega)$ de $\Big(\Omega \times [0,t], \mathcal{F}_t \otimes \mathcal{B}([0,t])\Big)$ dans $\Big(\mathbb{R}, \mathcal{B}(\mathbb{R})\Big)$ est mesurable.

L'énoncé suivant assure qu'il ne manque pas de processus progressifs.

Proposition 23.1 *Tout processus adapté continu à gauche ou à droite est progressif.*

Preuve Il suffit de constater que pour $0 \le s \le t$, dans le cas continu à droite on a :

$$X_s(\omega) = \lim_{n\to\infty} \sum_{k=1}^{2^n} 1_{\{t\frac{k-1}{2^n} \le s < t\frac{k}{2^n}\}} \, X_{t\frac{k}{2^n}}(\omega) + 1_{\{s=t\}} \, X_t(\omega) \,.$$

Et dans le cas continu à gauche :

$$X_s(\omega) = 1_{\{s=0\}} \, X_0(\omega) + \lim_{n\to\infty} \sum_{k=0}^{2^n-1} 1_{\{t\frac{k}{2^n} < s \le t\frac{k+1}{2^n}\}} \, X_{t\frac{k}{2^n}}(\omega). \quad \diamond$$

Exemple Si S est un temps d'arrêt, $(\omega, s) \mapsto 1_{\{s<S(\omega)\}}$ (respectivement $(\omega, s) \mapsto 1_{\{s\le S(\omega)\}}$) est adapté et continu à droite (respectivement à gauche), et donc progressif.

Les processus progressifs les plus simples sont les processus en escalier, i.e. les processus de la forme $H_t(\omega) := \sum_{j=0}^{N-1} U_j(\omega) \, 1_{[t_j, t_{j+1}[}(t)$, avec $0 = t_0 < t_1 < \ldots < t_N$ et $U_j \in b\mathcal{F}_{t_j}$.

Notons Λ^∞ l'espace (de Hilbert) des processus progressifs (H_s) tels que $I\!\!E\left[\int_0^\infty H_s^2 \, ds\right]$ soit finie. Observons que les intégrales de ces processus sont presque sûrement continues, puisque $\left|\int_{t-\varepsilon}^t H_s \, ds\right|^2 \le \varepsilon \int_0^t H_s^2 \, ds$.

Les processus en escalier sont denses dans Λ^∞, au sens où pour tout $H \in \Lambda^\infty$, il existe une suite de processus en escalier $H^{(n)}$ telle que $\lim_{n\to\infty} \int_0^\infty I\!\!E\Big[|H_s - H_s^{(n)}|^2\Big] ds = 0$.

De même, soit Λ^t l'espace des processus progressifs (H_s) tels que $I\!\!E\left[\int_0^t H_s^2 \, ds\right] < \infty$.

Considérons maintenant un $(\mathcal{F}_t)$-mouvement brownien (B_t). Par définition, l'intégrale d'Itô d'un processus en escalier (H écrit comme ci-dessus) est :

$$\int_0^t H \, dB \equiv \int_0^t H_s \, dB_s := \sum_{j=0}^{N-1} U_j \, (B_{t_{j+1}\wedge t} - B_{t_j\wedge t}).$$

De façon équivalente, notant $N(t)$ le plus grand entier j tel que $t_j \le t$ et posant $U_N = 0$:

$$\int_0^t H\, dB \;=\; \sum_{j=0}^{N(t)-1} U_j\,(B_{t_{j+1}} - B_{t_j}) + U_{N(t)}\,(B_t - B_{t_{N(t)}}).$$

Exercice n^o 23.1 (i) Vérifier que $\int H\,dB$ est une martingale continue $\Big($considérer d'abord le cas $N = 2$ et $s < t_1 < t < t_2$, pour comprendre pourquoi $\mathbb{E}\Big[\int_0^t HdB \,\Big|\, \mathcal{F}_s\Big] = \int_0^s HdB\Big)$.

(ii) Vérifier le lemme suivant 23.2, qui découle aisément d'un simple calcul, dans lequel tous les termes rectangulaires ont une espérance nulle, par définition d'un $(\mathcal{F}_t)$-mouvement brownien.

Lemme 23.2 *Pour tout processus en escalier H et tout t , nous avons l'identité isométrique d'Itô :*

$$\mathbb{E}\left[\left|\int_0^t H_s\, dB_s\right|^2\right] = \mathbb{E}\left[\int_0^t H_s^2\, ds\right].$$

On peut par conséquent étendre par densité l'intégrale d'Itô aux processus $H \in \Lambda^\infty$, définissant ainsi une application linéaire Is de Λ^∞ dans l'espace $\mathcal{M}_c^\infty$ des martingales continues bornées dans L^2. Nous avons alors en outre :

$$\mathbb{E}\left[\int_0^\infty H_s^2\, ds\right] = \mathbb{E}[Is(H)_\infty^2] =: \|Is(H)\|^2 .$$

$Is(H)_t$ est usuellement noté $\int_0^t H_s\, dB_s$ ou $\int_0^t H\, dB$, pour tout $t \le \infty$. Cette notation est cohérente avec la propriété suivante.

Lemme 23.3 *Pour tout temps d'arrêt τ et tout $H \in \Lambda^\infty$, $t \mapsto H_t\, 1_{[0,\tau[}(t) \in \Lambda^\infty$,*
$$\int_0^\infty 1_{\{s<\tau\}}\, H_s\, dB_s = Is(H)_\tau =: \int_0^\tau H\, dB\,, \quad et \quad \mathbb{E}\left[\left|\int_0^\tau H\, dB\right|^2\right] = \mathbb{E}\left[\int_0^\tau H_s^2\, ds\right].$$

Preuve Pour les processus en escalier et les temps d'arrêt à valeurs discrètes, la propriété est obtenue par calcul direct, utilisant que

$$\int_0^\infty 1_{\{s<\tau\}}\, H_s\, dB_s = \int_0^\infty H_s\, dB_s - \int_0^\infty 1_{\{s\ge\tau\}}\, H_s\, dB_s\, .$$

Ensuite, si τ_n décroît vers τ , $H\, 1_{[0,\tau_n]}$ converge vers $H\, 1_{[0,\tau]}$ dans Λ^∞, de sorte que, par continuité de la martingale du second membre, le résultat s'ensuit pour les processus en escalier et les temps d'arrêt quelconques. Finalement, observons que si H_n converge vers H dans Λ^∞, alors $H_n\, 1_{[0,\tau]}$ converge vers $H\, 1_{[0,\tau]}$ dans Λ^∞, d'où (par l'inégalité de Doob) la convergence uniforme presque sûre des deux intégrales stochastiques. $\diamond$

Remarque 23.4 *De la même façon, pour tout couple de temps d'arrêt* $\sigma \le \tau$ *et tout* (H_t) *dans* Λ^∞, $(t \mapsto H_t\, 1_{[\sigma,\tau]}(t))$ *appartient à* Λ^∞, *et*

$$\int_0^\infty 1_{\{\sigma \le s \le \tau\}} H_s\, dB_s = Is(H)_\tau - Is(H)_\sigma =: \int_\sigma^\tau H_s\, dB_s \equiv \int_\sigma^\tau H\, dB\,.$$

De plus l'identité isométrique d'Itô s'écrit: $\quad I\!\!E\left[\left|\int_\sigma^\tau H\, dB\right|^2\right] = I\!\!E\left[\int_\sigma^\tau H_s^2\, ds\right].$

On étend légèrement la définition de l'intégrale d'Itô en considérant l'espace $\Lambda := \bigcap_{t>0} \Lambda^t$ des processus progressifs tels que $I\!\!E\left[\int_0^t H_s^2\, ds\right] < \infty$ pour tout t. Il est assez clair que l'intégrale d'Itô $\int_0^t H\, dB$ s'étend en une application linéaire de Λ dans l'espace $\mathcal{M}_c$ des martingales continues de carré intégrable, et que l'identité isométrique d'Itô reste vraie :

$$I\!\!E\left[\left|\int_0^t H\, dB\right|^2\right] = I\!\!E\left[\int_0^t H_s^2\, ds\right]. \qquad \text{Par polarisation, cela équivaut à :}$$

$$I\!\!E\left[\left(\int_0^t H\, dB\right) \times \left(\int_0^t K\, dB\right)\right] = I\!\!E\left[\int_0^t H_s K_s\, ds\right], \quad \text{pour tous } H\,,\, K \in \Lambda\,. \tag{4}$$

Comme ci-dessus, pour tout $H \in \Lambda$ et tous temps d'arrêt $\sigma \le \tau$, $H\, 1_{[\sigma,\tau]} \in \Lambda$ et $\int_0^t 1_{\{\sigma \le s \le \tau\}} H_s\, dB_s = \int_{t\wedge\sigma}^{t\wedge\tau} H\, dB$. De plus, nous avons :

$$I\!\!E\left[\left|\int_{t\wedge\sigma}^{t\wedge\tau} H\, dB\right|^2\right] = I\!\!E\left[\int_{t\wedge\sigma}^{t\wedge\tau} H_s^2\, ds\right] \text{ pour tous } H \in \Lambda,\ t \ge 0\,,\ \text{temps d'arrêt } \sigma \le \tau.$$

En particulier, $(t \mapsto B_{t\wedge\tau})$ est une martingale continue de carré intégrable.

Exercice n^o 23.2 Montrer que si $I\!\!E(\tau)$ est fini, alors $I\!\!E(B_\tau) = 0$ et $I\!\!E(B_\tau^2) = I\!\!E(\tau)$. (Utiliser l'inégalité de Doob.) Que peut-on dire du temps d'atteinte de 1 par B ?

La Proposition 20.2 sur la variation quadratique du brownien se généralise aux intégrales stochastiques $\int H dB$, fournissant du même coup une généralisation de la convergence des sommes de Riemann.

Proposition 23.5 *Fixons* $t \in I\!\!R_+$, *et une subdivision* $\mathcal{P}_t := \{0 = t_0 < t_1 < \ldots < t_N = t\}$ *de* $[0,t]$, *de pas* $|\mathcal{P}_t|$. *Soit* $H \in \Lambda$, *uniformément borné.*

Alors, lorsque $|\mathcal{P}_t| \to 0$, $\sum_{j=0}^{N-1} \left(\int_{t_j}^{t_{j+1}} H\, dB\right)^2$ *converge dans* L^2 *vers* $\int_0^t H_s^2\, ds$.

De plus, pour tout processus continu adapté Y *tel que* $\sup_{[0,t]} |Y|$ *est de carré intégrable,*

$$\sum_{j=0}^{N-1} Y_{t_j} \int_{t_j}^{t_{j+1}} H\, dB \ \text{ converge dans } L^2 \text{ (lorsque } |\mathcal{P}_t| \to 0) \text{ vers } \int_0^t Y\, H\, dB\,.$$

24 Formule d'Itô

Nous commençons par un cas particulier, qui est une formule d'intégration par parties.

Proposition 24.1 *Considérons 2 intégrales d'Itô* $X = \int H dB$ *et* $Y = \int K dB$ *, avec* $H, K \in \Lambda$ *. Alors pour tout* $t \geq 0$ *nous avons :*

$$X_t\, Y_t = \int_0^t H\, Y\, dB + \int_0^t X\, K\, dB + \int_0^t H_s\, K_s\, ds\,.$$

Ceci s'étend aux processus à variation bornée, généralisant la proposition 22.4 (dans le cas continu ; noter de nouveau que $H, K \in \Lambda \Rightarrow X, Y, Z$ continus).

Proposition 24.2 *Considérons une intégrale d'Itô* $X = \int H dB$ *, et des processus à variation bornée* $Y = \int K_s\, ds$ *,* $Z = \int H_s\, ds$ *, avec* $H, K \in \Lambda$ *. Alors p.s. pour tout* $t \geq 0$ *nous avons :*

$$X_t\, Y_t = \int_0^t H\, Y\, dB + \int_0^t X_s\, K_s\, ds\ ; \quad Z_t\, Y_t = \int_0^t H_s\, Y_s\, ds + \int_0^t Z_s\, K_s\, ds\,.$$

Ces deux énoncés conduisent à introduire une classe de processus permettant de les unifier, comme d'unifier des énoncés futurs. La filtration de la définition suivante est celle du mouvement brownien B.

Définition 24.3 *Étant donné un mouvement brownien réel* B *, notons* $\mathcal{SM}(B)$ *l'espace des* <u>*semi-martingales browniennes*</u>*, c'est-à-dire des processus réels* X *qui s'écrivent :* $X = X_0 + \int H\, dB + \int K_s\, ds$ *, pour* $H, K \in \Lambda$ *et* X_0 $\mathcal{F}_0$*-measurable, et définissons leur covariation quadratique, ou* <u>*crochet*</u>*, par :*

$$\left\langle X_0 + \int H\, dB + \int K_s\, ds\,,\, \tilde{X}_0 + \int \tilde{H}\, dB + \int \tilde{K}_s\, ds \right\rangle := \int H_s\, \tilde{H}_s\, ds\,.$$

Introduisons aussi la notation : $\quad d\Big(X_0 + \int H\, dB + \int K_s\, ds\Big) := H_s\, dB_s + K_s\, ds\,.$

Il n'y a aucune ambigüité dans cette definition 24.3 : la decomposition $X = X_0+M+A$ d'une semi-martingale brownienne $X \in \mathcal{SM}(B)$ comme somme d'une martingale M et d'un processus A à variation bornée est unique, comme le montre l'énoncé suivant, qui comporte aussi la formule complète d'intégration par parties.

Proposition 24.4 *(Intégration par parties) Pour toutes semi-martingales browniennes $X, Y \in \mathcal{SM}(B)$, nous avons p.s. :*

$$X_t\, Y_t = X_0\, Y_0 + \int_0^t Y_s\, dX_s + \int_0^t X_s\, dY_s + \langle X, Y\rangle_t \quad \textit{pour tout } t \in \mathbb{R}_+ \,. \tag{5}$$

La décomposition (de la définition 24.3) d'une semi-martingale brownienne $X \in \mathcal{SM}(B)$ est unique ; en particulier, le crochet $\langle X, Y\rangle$ de la définition 24.3 est bien défini.

Preuve La formule (5) découle aussitôt des propositions 24.1 et 24.2 par bilinéarité. Déduisons-en l'unicité de la décomposition $X = X_0 + \int H\, dB + \int K_s\, ds \in \mathcal{SM}(B)$. Si $X \equiv 0$, alors clairement $X_0 = 0$, et selon la formule (5) nous avons :

$$0 = X_t^2 = 2\int_0^t X_s\, dX_s + \int_0^t H_s^2\, ds = \int_0^t H_s^2\, ds \quad \text{p.s. pour tout } t > 0\,,$$

de sorte que H est négligeable pour $\mathbb{P} \otimes ds$. Ceci implique que $\int H \cdot dB \equiv 0$ p.s., et donc aussi que $\int K_s\, ds \equiv 0$ p.s., de sorte que K est négligeable pour $\mathbb{P} \otimes ds$. $\diamond$

Voici maintenant la formule principale du calcul stochastique d'Itô (avec l'identité isométrique d'Itô (4)). Comparer avec la proposition 22.5.

Théorème 24.5 *(Formule d'Itô) Soit B un mouvement brownien réels, $X \in \mathcal{SM}(B)$ une semi-martingale brownienne, et F une fonction de classe C^2 sur $\mathbb{R}$. Alors $F \circ X \in \mathcal{SM}(B)$, et nous avons p.s. pour tout $t \in \mathbb{R}_+$:*

$$F(X_t) = F(X_0) + \int_0^t F'(X_s)\, dX_s + \tfrac{1}{2}\int_0^t F''(X_s)\, d\langle X, X\rangle_s \,. \tag{6}$$

Preuve La formule d'intégration par parties de la proposition 24.4 ci-dessus implique que si la formule (6) est valable pour F, alors elle l'est aussi pour $x \mapsto x \times F(x)$. De sorte qu'elle est valable pour tout polynôme F. Par localisation, il est suffisant de considérer le cas de X à valeurs dans un intervalle $[0, t]$. Mais alors F est limite uniforme de polynômes (théorème de Stone-Weierstrass), ce qui permet de conclure par passage à la limite dans chacun des termes de la formule (6) . $\diamond$

Remarque 24.6 *1)* Dans le cas déterministe, ou si A est un processus continu à variation bornée, alors la formule d'Itô se réduit à la règle usuelle : *pour toute f de classe C^1 de $I\!R$ dans $I\!R$ et pour tout $t \in I\!R$ nous avons :* $f(A_t) = f(A_0) + \int_0^t f'(A_s)\, dA_s$.

2) Une écriture équivalente de la formule d'Itô est :

$$d[F(X_t)] = F'(X_t)\, dX_t + \tfrac{1}{2}\, F''(X_t)\, d\langle X, X\rangle_t \,.$$

3) La formule (5) d'intégration par parties de la proposition 24.4 est un cas particulier (choisissant $F(X) = X^2$) de la formule d'Itô (6). Particularisant encore, en prenant $Y \equiv f$ fonction à support compact de classe C^1 sur $I\!R_+^*$, nous obtenons :
$\int_0^\infty f(s)\, dX_s = -\int_0^\infty X_s\, f'(s)\, ds$. Ceci montre que dX_s est p.s. la différentielle au sens des distributions de la fonction continue (mais en général non différentiable) $s \mapsto X_s$.

La formule d'Itô (6) admet la généralisation multidimensionnelle suivante (avec plus ou moins la même preuve).

Théorème 24.7 *(Formule d'Itô multidimensionnelle) Soient $B^1, \ldots, B^d$ d mouvements browniens réels indépendants, $X^1, \ldots, X^n$ des semi-martingales browniennes, $X := (X^1, \ldots, X^n)$, et F une fonction de classe C^2 sur $I\!R^n$. Alors $F \circ X \in \mathcal{SM}(B)$, et nous avons p.s. pour tout $t \in I\!R_+$:* $d\langle B^j, B^k\rangle_t = 1_{\{j=k\}}\, dt$, *et*

$$F(X_t) = F(X_0) + \sum_{j=1}^n \int_0^t \frac{\partial F}{\partial x^j}(X_s)\, dX_s^j + \tfrac{1}{2} \sum_{j,k=1}^n \int_0^t \frac{\partial^2 F}{\partial x^j \partial x^k}(X_s)\, d\langle X^j, X^k\rangle_s \,. \qquad (7)$$

Voici un premier exemple d'application.

Corollaire 24.8 (i) *Pour toute martingale brownienne (réelle ou complexe) M et tout $\lambda \in \mathbb{C}$, $\exp(\lambda M - \lambda^2 \langle M\rangle/2)$ est une martingale (locale) continue. En particulier, si B est un mouvement brownien réel et si $f \in L^2_{loc}(I\!R_+)$ est une fonction déterministe, alors $t \mapsto \exp\left(\int_0^t f(s)\, dB_s - \tfrac{1}{2}\int_0^t f(s)^2\, ds\right)$ est une martingale.*

(ii) *Pour toute martingale brownienne réelle M nulle en 0 telle que $\langle M\rangle \leq c\, Id$, et pour tout $t, \alpha > 0$, nous avons :* $I\!P\left[\sup_{\{0 \leq s \leq t\}} |M_s| \geq \alpha\right] \leq 2\, \exp[-\alpha^2/(2ct)]$.

Preuve Pour (i), appliquer la formule d'Itô (7) avec $n=2$, $X^1=M$ et $X^2=\langle M\rangle$. Pour le cas particulier, la martingale locale est bien une vraie martingale, puisque l'exponentielle d'une variable gaussienne est intégrable. Pour déduire (ii), utiliser la martingale exponentielle de (i), avec $\lambda=\alpha/(ct)$. Utilisant l'inégalité de Doob, on voit par convergence dominée que c'est une martingale d'espérance 1. Ensuite l'hypothèse et l'inégalité maximale de Doob (théorème 21.4(i)) impliquent (optimisant a posteriori par rapport à λ) :

$$\mathbb{P}\left[\sup_{0\le s\le t} M_s\ge\alpha\right]\le\mathbb{P}\left[\sup_{0\le s\le t}\exp(\lambda M_s-\lambda^2\langle M\rangle_s/2)\ge\exp(\lambda\alpha-\lambda^2ct/2)\right]\le e^{\frac{-\alpha^2}{2ct}}.$$

Idem avec $-M$. $\diamond$

25 Théorème de Girsanov

Il s'agit essentiellement d'un changement de probabilité, comme dans la section 15. Considérons deux probabilités équivalentes $\mathbb{P}$ et $\mathcal{Q}$ sur $(\Omega,(\mathcal{F}_t),\mathcal{F}_\infty)$. Notons Z_t la dérivée de Radon-Nikodym de $\mathcal{Q}$ par rapport à $\mathbb{P}$ sur $(\Omega,\mathcal{F}_t)$, ce qui signifie que pour toute variable aléatoire $\mathcal{F}_t$-mesurable bornée F_t, nous avons : $\displaystyle\int F_t\,d\mathcal{Q}=\int F_t\,Z_t\,d\mathbb{P}$.

C'est aussi l'espérance conditionnelle de Z_{t+s} par rapport à $\mathcal{F}_t$ (revoir le théorème 9.1), et donc cela définit (sous $\mathbb{P}$) une martingale strictement positive (donc p.s. convergente), que nous supposerons continue. Cette hypothèse entraîne que $1/Z$ est continue aussi, et donc que $R_\cdot:=\displaystyle\int_0^\cdot Z^{-1}dZ$ définit une martingale locale continue (sous $\mathbb{P}$), telle que $Z_\cdot=Z_0+\displaystyle\int_0^\cdot Z\,dR$. Cette équation simple se résout aussitôt : nous avons $Z_t=Z_0\exp\big[R_t-\frac12\langle R\rangle_t\big]$ pour tout $t\ge0$, comme le montre aisément la formule d'Itô (R est une sorte de logarithme stochastique de Z). Voici l'analogue continu du théorème 15.2.

Proposition 25.1 *(Girsanov) Soit M une martingale locale continue sous $\mathbb{P}$. Alors $N:=M-\langle M,R\rangle=M-\displaystyle\int_0^\cdot Z^{-1}d\langle M,Z\rangle$ définit une martingale locale continue sous $\mathcal{Q}$, et nous avons $\langle N\rangle=\langle M\rangle$. La variation quadratique a le même sens sous $\mathbb{P}$ et sous $\mathcal{Q}$.*

Preuve Nous avons

$$d(NZ)=[MdZ+ZdM+d\langle M,Z\rangle]-[\langle M,R\rangle dZ+Z.Z^{-1}d\langle M,Z\rangle]=NdZ+ZdM,$$

ce qui montre que NZ est une martingale locale par rapport à $\mathbb{P}$, et donc que N est une martingale locale par rapport à $\mathcal{Q}$. Le second résultat est clair à partir de la définition de N, si l'on sait que la variation quadratique a le même sens sous $\mathbb{P}$ et sous $\mathcal{Q}$. Or les propositions 20.2 et 23.5 montrent que $\langle N\rangle$ peut s'obtenir comme limite presque sûre (donc sous $\mathbb{P}$ comme sous $\mathcal{Q}$) d'une sous-suite de sommes de carrés d'accroissements de N. $\diamond$

Théorème 25.2 *(Girsanov) Fixons* $\big(\Omega, (\mathcal{F}_t), I\!\!P\big)$, $T>0$, *et un* $(\mathcal{F}_t)$*-mouvement brownien d-dimensionnel* $(B_t \,|\, 0 \le t \le T)$*. Soit* $\gamma = (\gamma_t \,|\, 0 \le t \le T)$ *un processus prévisible borné à valeurs dans* $I\!\!R^d$*. Notons* $Y_\cdot := B_\cdot - \displaystyle\int_0^\cdot \gamma_s\, ds$*. Alors pour toute* F *borélienne positive ou bornée sur* $C([0,T], I\!\!R^d)$*, nous avons*

$$I\!\!E\Big[F\big(B[0,T]\big)\Big] = I\!\!E\left[F\big(Y[0,T]\big) \times \exp\Big(\int_0^T \gamma_s \cdot dB_s - \tfrac{1}{2}\int_0^T |\gamma_s|^2 ds\Big)\right],$$

ou encore

$$I\!\!E\Big[F\big(Y[0,T]\big)\Big] = I\!\!E\left[F\big(B[0,T]\big) \times \exp\Big(-\int_0^T \gamma_s \cdot dB_s - \tfrac{1}{2}\int_0^T |\gamma_s|^2 ds\Big)\right].$$

Preuve Le corollaire 24.8(i) assure que $Z_t := \exp\Big(\int_0^t \gamma_s\, dB_s - \frac{1}{2}\int_0^t |\gamma_s|^2 ds\Big)$ définit une martingale locale continue, et le corollaire 24.8(ii), compte tenu que $|\gamma|^2 \le c$, entraîne que $I\!\!P\left[\sup\limits_{0\le s\le t}\Big|\displaystyle\int_0^s \gamma\, dB\Big| \ge \alpha\right] \le 2\exp[-\alpha^2/(2ct)]$, ce qui suffit à assurer que Z est une martingale sur $[0,T]$, qu'on peut prolonger par Z_T sur $[T,\infty[$.
Soit $\mathcal{Q}$ la probabilité ayant Z_T pour densité par rapport à $I\!\!P$. Elle est équivalente à $I\!\!P$, et nous pouvons donc appliquer la proposition 25.1 ci-dessus, avec $M = B$ et $N = Y$: Y est une martingale sous $\mathcal{Q}$, de même variation quadratique que le brownien B. Un théorème de Lévy assure alors que Y est un brownien sous $\mathcal{Q}$.
La première formule s'en déduit aussitôt. La seconde se déduit de la première simplement en changeant γ en $-\gamma$ $\big($changeant donc $B_\cdot$ en $\big(B_\cdot + \int_0^\cdot \gamma\big)\big)$ et en changeant de fonction F :

$$I\!\!E\Big[F\big(Y[0,T]\big)\Big] = I\!\!E\Big[F\Big(\big(B_\cdot - \textstyle\int_0^\cdot \gamma\big)[0,T]\Big)\Big] = I\!\!E\left[F\big(B[0,T]\big)\, e^{-\int_0^T \gamma_s\cdot dB_s - \frac{1}{2}\int_0^T |\gamma_s|^2 ds}\right]. \ \diamond$$

Nota Bene : l'hypothèse de bornitude sur γ ci-dessus est relaxable ; elle n'a servi qu'à garantir que $\{Z_t \,|\, 0 \le t \le T\}$ est une martingale, i.e. (via l'inégalité maximale de Doob) que $I\!\!E(Z_T) = 1$. Le résultat demeure donc dès qu'on sait qu'il est vrai pour $F \equiv 1$.

Le cas particulier suivant du théorème de Girsanov 25.2 est déjà important.

Corollaire 25.3 *(Formule de Cameron-Martin) Pour* B *brownien et tous* $T \in I\!\!R_+$, $\lambda \in I\!\!R^d$*, et* F *borélienne positive ou bornée sur* $C([0,T], I\!\!R^d)$*, nous avons :*

$$I\!\!E\Big[F\Big(\big\{B_t + \lambda t \,|\, 0 \le t \le T\big\}\Big)\Big] = I\!\!E\left[F\Big(\big\{B_t \,|\, 0 \le t \le T\big\}\Big) \times \exp\Big(\lambda\cdot B_T - |\lambda|^2\, T/2\Big)\right].$$

Nota Bene : Pour mémoriser le signe devant $\int_0^T \gamma_s\, dB_s$ ou $\lambda\, B_T$: si $F\big(B[0,T]\big) = 1_{\{B_T>0\}}$ et $T \to \infty$: $1 = \lim\limits_{T\to\infty} I\!\!E\Big[1_{\{B_T+T>0\}}\Big] \ne \lim\limits_{T\to\infty} I\!\!E\Big[1_{\{B_T>0\}}\, e^{-B_T - T/2}\Big] = 0$!
On peut aussi se rappeler qu'enlever $\int\gamma$ à B doit être compensé par l'ajout de $e^{\int \gamma\cdot dB}$.

Exemple : Transformée de Laplace des temps d'atteinte du brownien avec dérive $\lambda \in \mathbb{R}$.
Fixons $x, \alpha \in \mathbb{R}_+^*$, un brownien réel standard B, et considérons le temps d'atteinte de x par B : $T_x = T_x(B)$, et idem pour $B + \lambda Id$: $T_x^\lambda := T_x(B + \lambda Id)$.
La formule de Cameron-Martin ci-dessus permet d'écrire :

$$\mathbb{E}\big(e^{-\alpha T_x^\lambda}\big) = \lim_{T\to\infty} \mathbb{E}\big(e^{-\alpha T_x^\lambda} 1_{\{T_x^\lambda \le T\}}\big) = \lim_{T\to\infty} \mathbb{E}\Big(e^{-\alpha T_x} 1_{\{T_x \le T\}} \times e^{\lambda B_T - \lambda^2 T/2}\Big)$$

$$= \lim_{T\to\infty} \mathbb{E}\Big(e^{-\alpha T_x} 1_{\{T_x \le T\}} \times e^{\lambda B_{T_x} - \lambda^2 T_x/2}\Big) = e^{\lambda x} \times \lim_{T\to\infty} \mathbb{E}\Big(e^{-(\alpha+\lambda^2/2)\, T_x} 1_{\{T_x \le T\}}\Big)$$

$$= e^{\lambda x} \times \mathbb{E}\big(e^{-(\alpha+\lambda^2/2)\, T_x}\big) = e^{\lambda x - \sqrt{2\alpha+\lambda^2}\,|x|} = e^{(\lambda - \sqrt{2\alpha+\lambda^2}\,)\, x},$$

puisque $1 = \mathbb{E}\big(e^{sgn(x)\sqrt{2a}\, B_{T_x\wedge T} - a\,(T_x \wedge T)}\big) \longrightarrow e^{\sqrt{2a}\,|x|} \times \mathbb{E}\big(e^{-a\, T_x}\big)$, pour $a \ge 0$ et $T \to \infty$.
Faisant décroître α vers 0, nous obtenons $\mathbb{P}\big(T_x^\lambda < \infty\big) = e^{(\lambda - |\lambda|)\, x} = e^{-2\lambda^- x}$.
Et si $\lambda > 0$, T_x^λ est donc p.s. fini, et même intégrable puisque $\mathbb{E}(T_x^\lambda) = -\frac{d_o}{d\alpha}\, \mathbb{E}\big(e^{-\alpha T_x^\lambda}\big) = \frac{d_o}{d\alpha}\,(x\sqrt{2\alpha+\lambda^2}\,) = \frac{x}{\lambda}$. De plus, par intégration, on peut vérifier que sa transformée de Laplace ci-dessus assigne au temps T_x^λ la densité sur $\mathbb{R}_+$:

$$s \longmapsto \frac{x}{\sqrt{2\pi\, s^3}}\, e^{-\frac{(x-\lambda\, s)^2}{2s}}.$$

26 Équations différentielles stochastiques (E.D.S.)

On nomme ainsi toute équation du type $X_s^x = x + \int_0^s \sigma(t, X_t^x)\, dB_t + \int_0^s b(t, X_t^x)\, dt$,
où (B_t) est un mouvement brownien réel, σ est une fonction de $\mathbb{R}_+ \times \mathbb{R}$ dans $\mathbb{R}$, b est une fonction de $\mathbb{R}_+ \times \mathbb{R}$ dans $\mathbb{R}$, $x \in \mathbb{R}$, s décrit $\mathbb{R}_+$, et l'inconnue est le processus X_s^x.
Cela se note aussi : $dX_t = \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt$.

Définition 26.1 *Il y a unicité en loi dans l'E.D.S. lorsque 2 solutions quelconques (qui sont égales à l'instant 0) ont la même loi. Il y a unicité trajectorielle lorsque 2 solutions quelconques (qui sont égales à l'instant 0) sont p.s. égales à tout instant $t > 0$. Une solution est dite forte lorsqu'elle est adaptée à la filtration de B. Elle est dite faible sinon.*

Lemme 26.2 *En cas d'unicité trajectorielle, toute solution est nécessairement forte.*

Voici un théorème d'existence et d'unicité.

Théorème 26.3 *Fixons des fonctions boréliennes σ, b sur $\mathbb{R}_+ \times \mathbb{R}^d$, sous-linéaires et lipschitziennes au sens suivant : pour tout $s > 0$ il existe une constante C_s telle que pour tous $t \in [0, s]$, $x, x' \in \mathbb{R}^d$:*

$$|\sigma(t,x) - \sigma(t,x')| + |b(t,x) - b(t,x')| \le C_s\, |x - x'|\,.$$

Fixons un mouvement brownien (B_t) *et* $x \in \mathbb{R}$. *Alors il existe un unique processus continu* (X_s^x), *solution forte de l'E.D.S. :* $X_s^x = x + \int_0^s \sigma(t, X_t^x)\, dB_t + \int_0^s b(t, X_t^x)\, dt\,.$

Le cas le plus important est le suivant. Selon le théorème précédent 26.3, (X_s^x) est l'unique solution forte si σ et b sont sous-linéaires et lipschitziennes sur $\mathbb{R}$. La seconde assertion (ii) est directement déduite de la formule d'Itô.

Théorème 26.4 *Soit* $X = (X_s^x)$ *un processus réel solution forte de l'E.D.S.*

$$X_s^x = x + \int_0^s \sigma(X_t^x)\, dB_t + \int_0^s b(X_t^x)\, dt\,,$$

dans laquelle les fonctions $\sigma = \sigma(x)$ *et* $b = b(x)$ *ne dépendent que de* x. *Alors*

(i) (X_s^x) *est* <u>*fortement markovien*</u> *: pour tout temps d'arrêt fini* T, *le processus (décalé dans le temps)* $(X_{T+s}^x - X_T^x)_{s\geq 0}$ *a la même loi que* (X_s^0), *et, connaissant* X_T^x, *est indépendant de la tribu* $\mathcal{F}_T = \sigma\{X_{\min\{s,T\}}^x \,|\, s \geq 0\}$ *du passé de* X^x *avant* T.

(ii) *Pour toute fonction* f *de classe* C^2 *sur* $\mathbb{R}$ *et tous* $x \in \mathbb{R}$, $s \geq 0$, *nous avons p.s. :*

$$f(X_s^x) = f(x) + \int_0^s f'(X_t^x)\, \sigma(X_t^x)\, dB_t + \int_0^s \mathcal{L}f(X_t^x)\, dt\,,$$

où $\quad \mathcal{L} := \frac{1}{2}\,\sigma(x)^2 \frac{d^2}{dx^2} + b(x)\,\frac{d}{dx}\,, \qquad$ *i.e.* $\quad \mathcal{L}f(x) = \frac{1}{2}\,\sigma(x)^2\, f''(x) + b(x)\, f'(x)\,.$

Définition 26.5 *Le processus de Markov* X *du théorème 26.4 est ce qu'on nomme une* <u>*diffusion*</u> *réelle.*

L'opérateur différentiel du second ordre $\mathcal{L}$ *est nommé* <u>*générateur infinitésimal*</u> *de* X.

Le <u>*semi-groupe*</u> *associé est la famille des opérateurs* $\{P_t \,|\, t \in \mathbb{R}_+\}$ *définis par :*

$P_t f(x) := \mathbb{E}_x[f(X_t)] \equiv \mathbb{E}[f(X_t^x)]$, *pour toute fonction* f *mesurable bornée ou positive sur* $\mathbb{R}$, *tout* $x \in \mathbb{R}$, *et tout* $t \in \mathbb{R}_+$.

Théorème 26.6 (i) *Les operateurs* P_t *sont de norme 1 (dans* L^1, L^2, L^∞*), et vérifient :*

$$P_0 = Id\,,\;\; P_t 1 = 1\,,\;\; P_{s+t} = P_s P_t \quad (\textit{pour tous } s,t \geq 0).$$

De plus, $(t,x) \mapsto P_t f(x)$ *est continue bornée si* f *l'est et si* (X_t^x) *est continue comme fonction de* (t,x).

(ii) *Supposons* σ *et* b *bornées et lipschitziennes et la fonction* f *de classe* C^2, *bornée à dérivées bornées. Alors le processus* $\left(f(X_s^x) - f(x) - \int_0^s \mathcal{L}f(X_t^x)\, dt\right)$ *est une martingale,*

$P_s f = f + \int_0^s P_t\, \mathcal{L}f\, dt\,,\;\; \mathcal{L}f = \frac{d_o}{dt} P_t f\,,\;$ *et* $\;\mathcal{L}\, P_t f = \frac{d}{dt} P_t f = P_t\, \mathcal{L}f\;$ *pour tous* $s,t \in \mathbb{R}_+$.

C'est la célèbre équation de <u>*Fokker-Planck*</u>, *dite aussi* <u>*équation de la chaleur*</u>.

(iii) *Dans le cas où* $X = B$ *est brownien réel, nous avons simplement* $\mathcal{L} = \frac{1}{2}\,\frac{d^2}{dx^2}$.

26.7 Exemple des E.D.S. linéaires

Proposition 26.8 *Considérons des fonctions déterministes a, b, u, v de $\mathbb{R}_+$ dans $\mathbb{R}$, de classe C^1, un mouvement brownien réel B, et l'équation différentielle stochastique sur $\mathbb{R}$:*

$$dX_t \,=\, (a_t\,X_t + b_t)\,dB_t + (u_t\,X_t + v_t)\,dt\,.$$

X_0 étant fixé, cette équation admet pour seule solution :

$$X_t = \Big(X_0 + \int_0^t b_s\,\Lambda_s\,dB_s + \int_0^t (v_s - a_s b_s)\,\Lambda_s\,ds\Big)\Big/\Lambda_t\,,$$

où $\quad \Lambda_t := \exp\Big(-\int_0^t a_s\,dB_s - \int_0^t (u_s - a_s^2/2)\,ds\Big)\,.$

<u>Preuve</u> L'unicité découle du théorème 26.3. Il suffit donc de vérifier que l'expression, disons Y_t , donnée dans l'énoncé pour X_t est bien solution. La formule d'Itô donne d'une part :

$$d\Lambda_s^{-1} = \Lambda_s^{-1} \times \Big(a_s\,dB_s + (u_s - a_s^2/2)\,ds\Big) + \tfrac{1}{2}\,\Lambda_s^{-1}\,a_s^2\,ds = (a_s\,dB_s + u_s\,ds) \times \Lambda_s^{-1}\,.$$

D'autre part, posant $Z_t := \Lambda_t\,Y_t$ et appliquant la formule d'intégration par parties à $Y_t = Z_t \times \Lambda_t^{-1}$, nous obtenons :

$$dY_s = (dZ_s)/\Lambda_s + Z_s\,d\Lambda_s^{-1} + d\langle Z, \Lambda^{-1}\rangle_s$$

$$= b_s\,dB_s + (v_s - a_s b_s)\,ds + (a_s\,dB_s + u_s\,ds)\,Y_s + a_s b_s\,ds = (a_s Y_s + b_s)\,dB_s + (u_s Y_s + v_s)\,ds\,. \ \diamond$$

<u>Exercice</u> Calculer $\mathbb{E}(X_t)$ et $\mathbb{E}(X_t^2)$ (commencer par obtenir une équation différentielle ordinaire satisfaite par chacune de ces expressions).

26.9 Exemple : le processus d'Ornstein-Uhlenbeck

Il s'obtient en prenant dans la proposition 26.8 ci-dessus $a = v = 0$, $b = 1$, $u = -c$, où $c > 0$ est une constante positive. Il s'agit donc de la solution de l'équation de <u>Langevin</u> :

$$dX_t = dB_t - c\,X_t\,dt\,, \qquad \text{avec } \ c > 0\,.$$

La solution peut s'exprimer sous la forme (corollaire de la proposition 26.8) :

$$X_t = e^{-ct} \times \Big(X_0 + \int_0^t e^{cs}\,dB_s\Big)\,, \quad \text{ou bien encore : } \ X_t = e^{-ct} \times \Big(X_0 + B\big[(e^{2ct} - 1)/2c\big]\Big).$$

Son générateur est $\mathcal{L} = \frac{1}{2}\,\dfrac{d^2}{dx^2} - c\,x\,\dfrac{d}{dx}$ (selon le théorème 26.4 et la définition 26.5).

Pour X_0 déterministe, c'est aussi le processus gaussien caractérisé par la moyenne $e^{-ct}X_0$ et la fonction de covariance $e^{-c|s-t|}\times\big(1-e^{-2c\min\{s,t\}}\big)/(2c)\cdot$

Pour X_0 gaussienne de loi $\mathcal{N}\big(0,\frac{1}{2c}\big)$ indépendante de B, X est "stationnaire" : pour tout $t\ge 0$ la loi de X_t est $\mathcal{N}\big(0,\frac{1}{2c}\big)$, et la fonction de covariance devient $e^{-c|s-t|}/(2c)$.

Le semi-groupe (P_t) de (X_t) est donné par :

$$P_t f(x) := \mathbb{E}[f(X_t)\,|\,X_0=x] = \int_{\mathbb{R}} f\Big(e^{-ct}x+\sqrt{(1-e^{-2ct})/(2c)}\;y\Big)\;e^{-y^2/2}\,dy/\sqrt{2\pi}\;.$$

Il jouit de la propriété remarquable de symétrie suivante : pour toutes f,g mesurables bornées sur $\mathbb{R}$ et tout $t\ge 0$

$$\int_{\mathbb{R}} P_t f(x)\,g(x)\,e^{-c\,x^2}\,dx \;=\; \int_{\mathbb{R}} f(x)\,P_t g(x)\,e^{-c\,x^2}\,dx\,.$$

De façon équivalente, pour toutes f,g de classe C^2 sur $\mathbb{R}$, bornées à dérivées bornées, nous avons :

$$\int_{\mathbb{R}} \mathcal{L}f(x)\,g(x)\,e^{-c\,x^2}\,dx \;=\; \int_{\mathbb{R}} f(x)\,\mathcal{L}g(x)\,e^{-c\,x^2}\,dx\,.$$

Cette dernière propriété se vérifie par intégration par parties, et entraîne la précédente, en remplaçant f par $P_s f$ et g par $P_{t-s}g$, via l'équation de la chaleur $\big($théorème 26.6(ii)$\big)$.

27 Modèle de Black-Scholes

Le cours d'une action (ou d'une obligation) au temps $t\in[0,T]$ est un processus aléatoire (X_t) qu'on suppose vérifier l'équation différentielle stochastique : $dX_t = X_t\times(r\,dt+\sigma dB_t)$, linéaire (du type 26.7), où le taux r et la volatilité σ sont ici supposés constants (ce qui sous-entend que le terme T ne doit pas être grand).

Intéressons-nous à une option européenne, c'est-à-dire à un contrat comportant le droit soit de vendre ("put") soit d'acheter ("call") au temps fixé T à un prix fixé K. On note $Z_t = F(t,X_t)$ le prix d'un tel contrat à l'instant $t\in[0,T]$. On a clairement $Z_T = g(X_T)$, avec $g(x) := (K-x)^+$ dans le cas du "put" et $g(x) := (x-K)^+$ dans le cas du "call".

Considérons un portefeuille, c'est-à-dire un panaché d'actions ou obligations et d'options, dont la valeur à l'instant t se présente ici sous la forme aX_t+Z_t, avec $a>0$ supposé constant. On dit qu'il y a couverture totale lorsque la valeur du portefeuille évolue plus favorablement que le placement au taux d'intérêt fixe β. Le cas limite de la plus petite couverture totale correspond à l'équation : $(*)\quad d(aX_t+Z_t) = \beta\,(aX_t+Z_t)\,dt$.

Nous allons trouver la solution Z de cette équation, c'est-à-dire trouver le juste prix Z de l'option considérée.

Z est une semimartingale continue, et la formule d'Itô (7) montre que $(*)$ équivaut à

$$(r\,dt+\sigma dB_t)aX_t + \Big(dt\frac{\partial F}{\partial t} + (rX_t dt+\sigma X_t dB_t)\frac{\partial F}{\partial x} + \frac{\sigma^2}{2}X_t^2 dt\frac{\partial^2 F}{\partial x^2}\Big)(t,X_t) = \beta(aX_t+Z_t)dt,$$

id est
$$\Big(a+\frac{\partial F}{\partial x}(t,X_t)\Big)\sigma X_t dB_t+\Big(\frac{\partial F}{\partial t}+rX_t\frac{\partial F}{\partial x}+\frac{\sigma^2}{2}X_t^2\frac{\partial^2 F}{\partial x^2}-\beta F+a(r-\beta)X_t\Big)(t,X_t)dt=0,$$
ou encore
$$a+\frac{\partial F}{\partial x}(t,X_t)=0=\frac{\partial F}{\partial t}(t,X_t)+rX_t\frac{\partial F}{\partial x}(t,X_t)+\frac{\sigma^2}{2}X_t^2\frac{\partial^2 F}{\partial x^2}(t,X_t)-\beta F(t,X_t)+a(r-\beta)X_t\,.$$
Nous cherchons donc $F=F(t,x)$ sur $[0,T]\times I\!R_+^*$ telle que
$$(**)\qquad 0=a+\frac{\partial F}{\partial x}=\frac{\partial F}{\partial t}+\beta\,x\,\frac{\partial F}{\partial x}+\tfrac{\sigma^2}{2}\,x^2\,\frac{\partial^2 F}{\partial x^2}-\beta F=F(T,x)-g(x)\,.$$

Soient $Y_t:=Y_0\,\exp\Big[\sigma B_t+(\beta-\frac{\sigma^2}{2})\,t\Big]$ et $G(t,x)=e^{\beta\,(t-T)}\times I\!E_x\Big[g(Y_{T-t})\Big]$.

Pour h mesurable bornée ou positive sur $I\!R_+^*$ et $t>0$, notons aussi $P_th(x):=I\!E_x[h(Y_t)]$.
La formule d'Itô montre pour tout $t>0$ que $dY_t=\sigma Y_t\,dB_t+\beta Y_t\,dt$, et que si h est de classe C^2 à dérivées bornées sur $I\!R_+^*$:
$$h(Y_t)=h(Y_0)+\int_0^t h'(Y_s)\sigma Y_s dB_s+\int_0^t\Big(\tfrac{1}{2}\,h''(Y_s)\sigma^2Y_s^2+h'(Y_s)\beta Y_s\Big)ds\,,$$
et donc $\quad P_th(x)=h(x)+\displaystyle\int_0^t I\!E_x\Big(\tfrac{1}{2}\,h''(Y_s)\sigma^2Y_s^2+h'(Y_s)\beta Y_s\Big)ds\,.$

Il est immédiat (par rééchelonnement de B) que le terme sous la dernière intégrale est continu en $s=0$. Nous avons donc $\frac{d_o}{dt}P_th(x)=\frac{1}{2}\,\sigma^2x^2h''(x)+\beta xh'(x)=:\mathcal{L}h(x)$ pour tout $x>0$ et pour h régulière (noter que c'est un cas particulier du théorème 26.4).

Par ailleurs la propriété de Markov (celle de B entraîne celle de Y) montre que pour h mesurable bornée ou positive sur $I\!R$ et $x,s,t>0$ on a (voir aussi le théorème 26.6(i)) :
$$P_sP_th(x)=I\!E_x\Big[I\!E_{Y_s}[h(Y_t)]\Big]=I\!E_x[h(Y_{s+t})]=P_{s+t}h(x)\,.$$
Or de même que B_t, Y_t admet pour $t>0$ une densité bornée de classe C^∞ sur $I\!R_+^*$, ce qui entraîne que P_th, qui s'écrit comme convolée de h et de cette densité, est aussi de classe C^∞, pour tout $t>0$, et a ses premières dérivées bornées, pour peu que h soit suffisamment intégrable. En particulier, pour tout $t>0$, écrivant (voir aussi le théorème 26.6(ii))
$\frac{P_{s+t}g-P_tg}{s}=\frac{P_s-P_0}{s}\,P_t\,g\underset{s\searrow 0}{\longrightarrow}\frac{d_o}{ds}P_s\,P_t\,g\,,\quad$ nous obtenons : $\frac{d}{dt}P_tg=\mathcal{L}P_tg\,.$

Or $G(t,x)=e^{\beta\,(t-T)}\times P_{T-t}g(x)$, et donc pour tout $t\in[0,T[$ nous avons $\frac{d}{dt}G=\beta\,G-\mathcal{L}G$. De plus nous avons aussi G continue sur $[0,T]\times I\!R_+^*$, et $G(T,x)=g(x)$. Oubliant la première équation de $(**)$, nous voyons que $F-G$ est solution, avec $g\equiv 0$, c'est-à-dire correspond au prix d'un contrat rendant ce prix évidemment nul. On doit donc avoir $F=G$, c'est-à-dire
$$F(t,x)=e^{\beta\,(t-T)}\times I\!E\Big[g\Big(x\,\exp\Big[\sigma B_{T-t}+(\beta-\tfrac{\sigma^2}{2})\,(T-t)\Big]\Big)\Big]$$
$$=e^{\beta\,(t-T)}\int_{-\infty}^{\infty}g\Big(x\,\exp\Big[\sigma\sqrt{T-t}\,y+(\beta-\tfrac{\sigma^2}{2})\,(T-t)\Big]\Big)\times e^{-y^2/2}\frac{dy}{\sqrt{2\pi}}\,.$$

Dans le cas d'une option d'achat ("call"), cela donne

$$F(t,x) = \frac{e^{\beta\,(t-T)}}{\sqrt{2\pi}} \int_{\frac{\log(K/x)-(\beta-\frac{\sigma^2}{2})\,(T-t)}{\sigma\sqrt{T-t}}}^{\infty} \Big[x \,\exp\Big[\sigma\sqrt{T-t}\; y + (\beta - \tfrac{\sigma^2}{2})\,(T-t)\Big] - K\Big] e^{-\frac{y^2}{2}} dy$$

$$= x\,\Phi\left[\frac{\log(K/x) - (\beta + \frac{\sigma^2}{2})\,(T-t)}{\sigma\sqrt{T-t}}\right] - Ke^{-\beta\,(T-t)}\Phi\left[\frac{\log(K/x) - (\beta - \frac{\sigma^2}{2})\,(T-t)}{\sigma\sqrt{T-t}}\right],$$

où $\Phi(\alpha) := (2\pi)^{-1/2}\int_\alpha^\infty e^{-y^2/2}dy\,.$ Soit aussi $\tilde{\Phi}(\alpha) := (2\pi)^{-1/2}\int_{-\infty}^\alpha e^{-y^2/2}dy\,.$
Pour passer du "call" au "put", il suffit de changer $(x-K)^+$ en $(K-x)^+$, ce qui revient à changer Φ en $\Phi - 1 = -\tilde{\Phi}\,.$ Donc dans le cas d'une option de vente ("put") nous avons

$$F(t,x) = Ke^{-\beta\,(T-t)}\tilde{\Phi}\left[\frac{\log(K/x) - (\beta - \frac{\sigma^2}{2})\,(T-t)}{\sigma\sqrt{T-t}}\right] - x\,\tilde{\Phi}\left[\frac{\log(K/x) - (\beta + \frac{\sigma^2}{2})\,(T-t)}{\sigma\sqrt{T-t}}\right].$$

Bibliographie

DOTHAN M. *Prices in Financial Markets.* Oxford University Press, 1990.

FOATA D., FUCHS A. *Calcul des probabilités.* Dunod, Paris 1998, 2003, 2012.

FOATA D., FUCHS A. *Processus stochastiques.* Dunod, Paris 2002.

LAMBERTON D., LAPEYRE B. Mathématiques et Applications, Ellipses, 1991.
Introduction au calcul stochastique appliqué à la finance.

MÉTIVIER M., NEVEU J. *Cours de probabilités.* École polytechnique, Palaiseau 1979.

MIKOSCH T. *Elementary stochastic calculus.* World Scientific, 1998.

ROSS S. *Introduction to probability models.* Academic press, 1980.