Cours de probabilités MASS 61

Yoann Dabrowski

20 mai 2015

Chapitre 1

Bases des probabilités.

1 Espaces probabilisés

On appelle expérience aléatoire toute expérience, provoquée ou non, reproductible, mais dont la répétition, dans des conditions supposées être les mêmes pour un observateur ou un expérimentateur, conduit à un ensemble de résultats dont aucun ne peut être prévu à l'avance. Cette définition se rattache à la notion de phénomènes aléatoires qui se manifestent toujours avec un certain degré d'indétermination, ce qui empêche de prévoir exactement le résultat de leurs observations, par opposition aux phénomènes déterministes régis par des lois qui en déterminent le déroulement et le résultat. On peut cependant remarquer que tout phénomène comporte une part d'aléatoire qui traduit l'impossibilité de connaître ou de maîtriser tous les paramètres qui le définissent. En fait, notre environnement comporte de nombreux phénomènes aléatoires qui interviennent dans toutes les branches de l'activité humaine. D'où la nécessité et l'importance de disposer et de développer des moyens et des techniques qui aident à comprendre ces phénomènes, à déceler et dégager des propriétés de régularité et des tendances quantifiées afin de maîtriser, prévoir et agir. A travers cette quantification, on obtient un calcul déterministe décrivant les phénomènes aléatoires.

Deux approches s'imposent et se complètent : L'observation et la modélisation. Mais un traitement qui ne s'appuie pas sur un modèle solide atteint vite ses limites et peut conduire à des résultats incohérents appelés paradoxes. La modélisation fournit un cadre d'étude fondé sur une théorie cohérente qui possède et permet de développer des outils efficaces et rationnels. L'objet fondamental de cette modélisation est l'espace probabilisé qui est un triplet $(\Omega, \mathcal{T}, P)$ dont les termes sont rappelés dans cette section. Il s'agit d'espaces mesurés particuliers dont on rappelle dans cette section quelques propriétés de bases supposées connues. Le but de ce cours, par rapport aux cours de probabilités précédent en MASS, est d' apprendre à utiliser dans ce contexte les outils d'analyses utilisés couramment par les probabilistes (notamment la théorie de la mesure).

1.1 L'espace des réalisations

L'espace des réalisations Ω d'une expérience aléatoire est l'ensemble des résultats possibles de cette expérience. Cet espace est aussi appelé univers, référentiel, ensemble fondamental ou espace des épreuves.

Les éléments $\omega \in \Omega$ sont appelés réalisations de l'expérience en question.

L'ensemble Ω peut être fini, infini dénombrable ou infini non dénombrable.

Selon le point de vue et l'objectif de l'étude, on peut considérer des ensembles différents pour une même expérience. Par exemple, on peut s'intéresser aux six nombres amenés par le lancer d'un dé ou uniquement à leur parité. Dans le premier cas, on prendra $\Omega = \{1, 2, 3, 4, 5, 6\}$ et dans le second $\Omega = \{0, 1\}$.

1.2 Événements et tribu d'événements

Dans l'étude d'une expérience aléatoire on s'intéresse à des ensembles de réalisations appelés événements. Un événement lié à une expérience aléatoire et à son espace Ω s'identifie donc à une partie de Ω. Les événements élémentaires s'identifient aux singletons $\{\omega\}$.

Définition 1. Soit $A \subset \Omega$ un événement et $\omega \in \Omega$ une réalisation de l'expérience. On dit que l'événement A s'est réalisé lorsque $\omega \in A$ et ne s'est pas réalisé lorsque $\omega \notin A$.

Remarque 1. Tout événement étant une partie de Ω, une correspondance s'établit alors entre le langage naturel des événements et celui des sous-ensembles. Nous en donnons ici quelques exemples.
- L'espace Ω représente l'événement certain et l'événement impossible correspond à l'ensemble vide $\emptyset$.
- L'événement contraire à l'événement A est représenté par son complémentaire A^c.
- Deux événements A et B sont dits *incompatibles* lorsqu'ils sont disjoints, i.e. lorsque $A \cap B = \emptyset$.

Définition 2. On appelle *système complet d'événements* (SCE) toute suite finie ou infinie $(A_n)_{n\geq 1}$ d'événements deux à deux incompatibles telle que $\cup_{n\geq 1} A_n = \Omega$. (On parle aussi parfois de partition).

Nous verrons plus loin que cette notion est très utile car elle ramène à des situations où on dispose de plus d'informations pour traiter les événements étudiés.

Définition 3. La famille $\mathcal{T}$ des événements constitue une partie de $P(\Omega)$ ayant la structure d'une tribu qui, par définition, possède les propriétés suivantes :

1. $\emptyset \in \mathcal{T}$, $\Omega \in \mathcal{T}$.
2. Si $A \in \mathcal{T}$ alors $A^c \in \mathcal{T}$.
3. Pour toute suite finie ou infinie (dénombrable) $(A_n)_{n\geq 1}$ de parties de $\mathcal{T}$, alors $\cup_{n\geq 1} A_n \in \mathcal{T}$.

Par construction (ou hypothèse), la tribu $\mathcal{T}$ contient tous les événements auxquels on s'intéresse lors de l'étude d'une expérience aléatoire. La structure de tribu est importante pour l'efficacité du modèle et pour donner un sens aux différentes opérations que l'on est amené à faire sur les événements.

En pratique, on peut supposer que toutes ces propriétés sont satisfaites sans référence explicite à la tribu, d'autant plus que, *lorsque Ω est fini ou dénombrable, on peut prendre pour tribu $\mathcal{T}$ la famille $P(\Omega)$ de toutes les parties de Ω.*

Pour $\Omega = \mathbb{R}^n$, on utilisera en général la tribu des boréliens $\mathcal{B}(\mathbb{R}^n)$, plus petite tribu engendrée par les ensembles ouverts.

1.3 Probabilité sur une tribu d'événements

Ayant défini les événements et leur famille, on est amené à quantifier la vraisemblance de chacun d'eux et à évaluer ses chances de réalisation au cours d'une expérience aléatoire. Ceci se fait en associant à chaque événement un nombre compris entre 0 et 1 appelé sa probabilité, en respectant quelques règles inspirées par l'intuition. Ceci nous conduit à la définition suivante

Définition 4 (Définition d'une probabilité)**.** On appelle *probabilité* sur une tribu $\mathcal{T}$ toute application $P : \mathcal{T} \to [0, 1]$ ayant les propriétés suivantes :

1. $P(\emptyset) = 0, P(\Omega) = 1$.
2. Pour toute suite finie ou infinie (dénombrable) $(A_n)_{n\geq 1}$ d'événements deux à deux incompatibles,

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

Autrement dit, P **est une mesure positive avec** $P(\Omega) = 1$.

La propriété 2., appliquée à une suite finie, est l'additivité d'une probabilité et, appliquée à une suite infinie, est la σ-additivité d'une probabilité. Une probabilité est aussi appelée *loi* ou *distribution.*

1.4 Rappels des propriétés de base

Pour ces proriétés, on se reportera au cours de théorie de la mesure :

Proposition 1. *1.* $P(A^c) = 1 - P(A)$.

2. Si $A \subset B$ *alors* $P(B) = P(A) + P(B - A)$. *En particulier, si* $A \subset B$ *alors* $P(A) \leq P(B)$.

3. Si $A \cap B = \emptyset$, *alors* $P(A \cup B) = P(A) + P(B)$. *Dans le cas général,* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. *Plus généralement encore,*

4.

$$P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1\leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}).$$

(Formule du crible ou de Poincaré).

5. Si $(A_n)_{n\geq 1}$ *est une suite croissante,*

$$P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n) = \sup_{n\geq 1} P(A_n).$$

6. . Si $(A_n)_{n\geq 1}$ *est une suite décroissante,*

$$P(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n) = \inf_{n\geq 1} P(A_n).$$

Définition 5. Un événement A est dit *presque certain* (ou *presque sûr*) si $P(A) = 1$ et *presque impossible* si $P(A) = 0$.

Définition 6. On appelle système presque complet d'événements (SPCE) toute suite finie ou infinie (dénombrable) $(A_n)_{n\geq 1}$ d'événements deux à deux incompatibles telle que

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = 1.$$

Proposition 2. *Soit* $(A_n)_{n\geq 1}$ *un SPCE. Alors, pour tout événement* B, *on a*

$$P(B) = \sum_{n\geq 1} P(A_n \cap B).$$

On utilisera souvent le lemme suivant vu en théorie de la mesure.

Lemme 3 (de classe monotone)**.** *Si deux probabilités* P_1, P_2 *sur* $(\Omega, \mathcal{T})$ *sont égales sur une classe* $\mathcal{C}$ *stable par intersection finie (c'est-à-dire* $\forall C \in \mathcal{C}, \mu(C) = \nu(C)$*) et telle que* $\sigma(\mathcal{C}) = \mathcal{T}$ *alors* $\mu = \nu$.

1.5 Exemple : Probabilité sur un ensemble fini ou dénombrable

Soit Ω un ensemble fini ou dénombrable. La donnée d'une probabilité P sur $P(\Omega)$ équivaut à la donnée d'une fonction $p : \Omega \to [0,1]$ telle que $\sum_{\omega\in\Omega} p(\omega) = 1$. La correspondance se fait de la manière suivante : Lorsque P est donnée, on définit p par les relations $p(\omega) = P(\{\omega\})$ et lorsque p est donnée, on définit P par l relation

$$P(A) = \sum_{\omega\in A} p(\omega).$$

En terme de théorie de la mesure on notera

$$\mathbf{P} = \sum_{\omega\in\boldsymbol{\Omega}} \mathbf{p}(\omega)\delta_\omega$$

où δ_ω est la mesure de Dirac en ω définie par $\delta_\omega(A) = 1$ si $\omega \in A$ et 0 sinon.

En effet, on retrouve bien la même valeur sur A qu'avant :

$$P(A) = \sum_{\omega\in\Omega} p(\omega)\delta_\omega(A) = \sum_{\omega\in A} p(\omega).$$

Quelques lois usuelles sur un ensemble fini ou dénombrable

La loi uniforme $\mathcal{U}(n)$, de paramètre $n \in \mathbb{N}^*$ est bien connue. On rappelle les lois usuelles rencontrées dans les cours précédents de probabilités.

Définition 7. La *loi binomiale* $B(n,p)$ de paramètres $(n,p) \in \mathbb{N}^* \times [0,1]$. C'est la probabilité sur $[\![0,n]\!]$ définie par les nombres $p_k = C_n^k p^k (1-p)^{n-k}$. En terme de théorie de la mesure

$$P = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \delta_k.$$

Définition 8. La *loi hypergéométrique* $H(N,n,p)$, $n \in [\![1,N]\!]$, $p = m/N$. C'est la probabilité sur $[\![0,n]\!]$ définie par les nombres $p_k = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}$. Notons que certains p_k peuvent être nuls. C'est la loi du nombre de succès (c'est-à-dire, le nombre de boules d'une catégorie) obtenus en n tirages sans remise, lorsque le nombre initial des boules de cette catégorie est m et le nombre total de boules dans l'urne est initialement N.

Définition 9. La *loi de Poisson* $P(\lambda)$, de paramètre $\lambda > 0$. C'est la loi sur $\mathbb{N}$ définie par les nombres $p_n = e^{-\lambda}\frac{\lambda^n}{n!}$. C'est la loi du nombre de succès associé à un phénomène discret qui se déroule dans le temps ou dans l'espace. C'est la limite (en loi au sens que l'on verra ultérieurement) de $B(n, \frac{\lambda}{n})$ quand $n \to \infty$ En théorie de la mesure elle s'écrit :

$$P = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \delta_n.$$

Définition 10. La loi géométrique G(p), de paramètre $p \in 0,1[$. C'est la loi sur N^* définie par les nombres $p_n = pq^{n-1}$ où $q = 1-p$. C'est la loi d'attente du premier succès dans une succession infinie dénombrable d'essais identiques et indépendants, lorsque la probabilité du succès à chaque essai est p. En théorie de la mesure elle s'écrit :

$$P = \sum_{n=1}^{\infty} p(1-p)^{n-1} \delta_n.$$

Définition 11. La loi géométrique $G_0(p)$, de paramètre $p \in]0,1[$. C'est la loi sur N définie par les nombres $p_n = pq^n$ où $q = 1-p$. C'est la loi du nombre d'échecs précédant le premier succès dans une succession infinie dénombrable d'essais identiques et indépendants, lorsque la probabilité du succès à chaque essai est p. En théorie de la mesure elle s'écrit :

$$P = \sum_{n=0}^{\infty} p(1-p)^n \delta_n.$$

1.6 Exemple : Probabilité à densité sur $\mathbb{R}$ (ou $\mathbb{R}^n$)

Une densité de probabilité f (sur $\mathbb{R}$) est une fonction positive $f : \mathbb{R} \to \mathbb{R}_+$, intégrable (par rapport à la mesure de Lebsegue) telle que $\int_{-\infty}^{\infty} f(x)dx = 1$. (Par exemple f continue par morceau et tout segment et $\int_{-\infty}^{\infty} f(x)dx := \lim_{K\to\infty} \int_{-K}^{K} f(x)dx = 1$)

Définition 12. Une proba P sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite continue (ou à densité, ou absolument continue par rapport à la mesure de Lebesgue) si pour toute partie borélienne

$$\forall B \in \mathcal{B}, P(X \in B) = \int_B f(x)dx.$$

Cette densité est alors unique et on dit que f est la densité de P.
De façon équivalent il suffit de vérifier :
pour tout intervalle $I = [a,b] \subset \mathbb{R}, P(X \in [a,b]) = \int_a^b f(x)dx$.

Remarque 2. Soit P une variable continue de densité f Pour tout $a \in \mathbb{R}$, $P(\{a\}) = \int_{\{a\}} f(x)dx = 0$.

Lois à densité classiques

Exemple 1. La **loi uniforme sur le segment** $[a,b]$, notée $\mathcal{U}([a,b])$ est la loi de densité $f(x) = \frac{1}{b-a} 1_{[a,b]}(x)$.

Exemple 2. La **loi exponentielle de paramètre** λ, notée $\mathcal{E}(\lambda)$. a pour densité $f(x) = \lambda e^{-\lambda x} 1_{[0,\infty[}(x)$. Elle modélise la durée de vie d'un matériel ou un processus sans mémoire car $P(X-t > s | X > t) = P(X > s)$. En effet, sa fonction de répartition est $F_X(x) = 1_{[0,\infty[}(x)(1 - e^{-\lambda x})$.

Exemple 3. La **loi normale** $\mathcal{N}(m,\sigma)$ (ou loi gaussienne) a pour densité :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Elle intervient dans le théorème centrale limite.

Exemple 4. La **loi semicirculaire** (ou de Wigner) a pour densité :

$$f(x) = 1_{[-R,R]}(x) \frac{2}{\pi R^2} \sqrt{R^2 - x^2}.$$

Elle intervient dans la loi des valeurs propres des grandes matrices aléatoires.

2 Variables aléatoires

En probabilité, on cherche à obtenir des résultats qui dépendent le moins possible de la modélisation de l'espace probabilisé de référence $(\Omega, \mathcal{T}, P)$. L'objet principal est donc les variables aléatoires que l'on définit sur cet espace.

Définition 13. Soit $(E, \mathcal{E})$ un espace mesurable. Une variable aléatoire (v.a.) définie sur Ω est une application mesurable $X : \Omega \to E$, c'est-à-dire telle que, pour tout $B \in \mathcal{E}$

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

Le plus souvent l'espace d'arrivée sera $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, cas des variables aléatoires réelles (v.a.r.) ou $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, cas des variables aléatoires vectorielles (v.a.v.) ou vecteurs aléatoires.

Comme cette tribu est engendrée par les produits d'intervalles, une application $X : \Omega \to \mathbb{R}^n$ est un v.a.v si et seulement si $X^{-1}(\prod_{i=1}^n [a_i, b_i]) \in \mathcal{T}$, pour tout $a_i < b_i$.

Définition 14. Soit $X : \Omega \to (E, \mathscr{E})$ une v.a. On appelle **loi de** X (ou distribution, ou mesure image de X) la probabilité sur $(E, \mathscr{E})$, notée P_X, définie par

$$P_X(B) = P(\{X \in B\}) = P(X^{-1}(B)), B \in \mathscr{E}$$

On vérifie que P_X ainsi définie est effectivement une probabilité (car l'image inverse respecte unions et complémentaires). Cela ramène l'étude des probabilités associées aux vecteurs aléatoires à l'étude de probabilités sur $\mathbb{R}^n$ puisqu'on veut des résultats ne dépendant pas de Ω mais seulement de la loi P_X.

Définition 15 (Egalité en loi)**.** Deux v.a. X et Y ayant même loi $P_X = P_Y$ sont dites équidistribuées ou égales en loi ou *identiquement distribuées*, ce que l'on peut noter $X =_{\mathscr{L}} Y$. Insistons sur le fait que l'égalité en loi est distincte de l'égalité ponctuelle.

Pour les vecteurs aléatoires, on a la conséquence importante du lemme de classe monotone (cf. Th de la mesure, utilisé dans l'unicité de la mesure de Lebesgue)

Proposition 4. *Soient $X, Y : \Omega \to \mathbb{R}^n$. Les propriétés suivantes sont équivalentes*

1. *X, Y sont égales en loi : $P_X = P_Y$.*
2. *Pour tout $h : \mathbb{R}^n \to \mathbb{R}$, continue bornée, $\int h(X)dP = \int h(Y)dP$*
3. *Pour tout ouvert O de $\mathbb{R}^n$, $P_X(O) = P_Y(O)$.*
4. *pour tout $(x_1, ..., x_n) \in \mathbb{R}^n$:*

$$P_X(]-\infty, x_1] \times ... \times]-\infty, x_n]) = P_Y(]-\infty, x_1] \times ... \times]-\infty, x_n]).$$

La fonction $F_X(x_1, ..., x_n) = P_X(]-\infty, x_1] \times ... \times]-\infty, x_n])$ appelée **fonction de répartition** caractérise donc la loi.

Démonstration. Les produits d'intervalles $]-\infty, x_1] \times ... \times]-\infty, x_n]$ et les ouverts sont des familles stables par intersection finie et engendrent la tribu des boréliens de $\mathbb{R}^n$ (car par intersection et complémentaire on obtient les boules carrées de la norme infini et que tout ouvert de $\mathbb{R}^n$ est union dénombrable de telles boules, de centre un point de$\mathbb{Q}^n$ par densité de$\mathbb{Q}^n$.) On applique donc le lemme de classe monotone 3 pour obtenir les 2 dernières équivalences. 1 implique 2 vient du th de transfert plus bas comme l'équivalence de 2 avec : Pour tout $h : \mathbb{R}^d \to \mathbb{R}$, $\int_{\mathbb{R}^d} h(x)dP_X(x) = \int_{\mathbb{R}^d} h(x)dP_Y(x)$.

Pour montrer 3 à partir de 2 et conclure, il suffit de remarquer que $h_n(x) = \max(1, nd(., O^c))$ sont des fonctions continues bornées par 1 (car la distance à un fermé $x \mapsto d(x, O^c) = \inf\{d(x, y), y \in O^c\}$ est continue, cf. MASS 31 TD1). Si $x \in O^c$, $h_n(x) = 0$ et sinon, h_n est une suite croissante qui tend vers $h_n(x) \to 1_O(x)$ (car si $x \in O$, $nd(., O^c) \to \infty$ donc ≥ 1 pour n assez grand donc $h_n(x) = 1$ pour n assez grand). Donc par convergence monotone, $\int_{\mathbb{R}^d} h_n(x) dP_X(x) \to P_X(O)$ d'où l'égalité du 3. par celle du 2.

□

Exemple 5. **V.a. discrètes, v.a. continues. Si $X = (X_1, ..., X_n)$ v.a. continues, X_i est continue mais la réciproque est fausse. Loi de $X = 1_{\{U \leq p\}}$ si $U \sim \mathscr{U}([0, 1])$.**

2.1 Espérance et Théorème de transfert

Définition 16. Soit $X : \Omega \to \mathbb{R}$ une v.a., l'espérance de X est notée

$$\mathbf{E}[X] = \int_\Omega X(\omega) dP(\omega).$$

Elle est définie dans deux cas :

1. si $X \geq 0$, et alors $\mathbf{E}(X) \in [0, \infty]$.
2. si $X \in L^1(\Omega, dP)$ i.e. si $\mathbf{E}(|X|) = \int |X| dP < \infty$.

Si $X : \Omega \to \mathbb{R}^n$, on note $\mathbf{E}(X) = (\mathbf{E}(X_1), ..., \mathbf{E}(X_n))$.

Exemple 6. Si $X \sim \mathcal{U}([a, b])$

$$\mathbf{E}(X) \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

Exemple 7. Si $X \sim \mathcal{E}(\lambda)$

$$\mathbf{E}(X) = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Exemple 8. Si $X \sim \mathcal{N}(m, \sigma^2)$ $\mathbf{E}(X) = m$.

La moyenne est une forme linéaire croissante sur l'espace vectoriel $L^1(\Omega, \mathcal{T}, P)$ des v.a. intégrables. En d'autres termes, on a toutes les propriétés usuelles de l'intégrale, dont les résultats de convergence :

Proposition 5. *Si X et Y sont deux v.a. intégrables, alors*

1. *$L^1(\Omega, \mathcal{T}, P)$ est un espace vectoriel et $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$.*
2. *$X \leq Y \Rightarrow \mathbf{E}[X] \leq \mathbf{E}[Y]$. En particulier, $X \geq 0 \Rightarrow \mathbf{E}[X] \geq 0$.*
3. *$|Z| \leq |X|$ implique Z intégrable.*
4. *$|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$.*
5. *$\mathbf{E}[c] = c$, $a \leq X \leq b \Rightarrow a \leq \mathbf{E}[X] \leq b$.*
6. *(Convergence monotone TCM) Si $X_n \geq 0$, X_n suite de v.a.r croissante et tend simplement vers X alors $\mathbf{E}(X_n) \to \mathbf{E}(X)$.*
7. *(Lemme de Fatou) Si $X_n \geq 0$ $\mathbf{E}(\liminf X_n) \leq \liminf \mathbf{E}(X)$.*
8. *(Convergence dominée TCD) Si $|X_n| \leq Y$, $\mathbf{E}[Y] < \infty$ et $P(X_n \to X) = 1$ (on dit $X_n \to X p.s.$) alors $\mathbf{E}(X_n) \to \mathbf{E}(X)$.*

Le premier théorème fondamental montrant que seule la loi P_X d'une v.a. est importante et pas la réalisation concrète sur Ω est la suivante :

Théorème 6 (Théorème de transfert). *Soit $X : (\Omega, \mathcal{T}, P) \to (E, \mathscr{E})$ une v.a. et $h : (E, \mathscr{E}) \to (I\!R, \mathcal{B}(I\!R))$ une fonction mesurable. Alors, si h est à valeur positive :*

$$\mathbf{E}(h(X)) = \int_{I\!R} h(x) dP_X(x).$$

De plus, si h n'est pas à valeur positive $h(X) \in L^1(\Omega, P)$ si et seulement si $h \in L^1(E, P_X)$ et on a encore $\mathbf{E}(h(X)) = \int h(x) dP_X(x)$.

Autrement dit on ramène une intégrale sur Ω à une intégrale sur $\mathbb{R}$: $\int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} h(x) dP_X(x)$.
Preuve : On procède comme pour la construction de l'intégrale. Si $h = 1_B$ avec $B \in \mathscr{E}$, $h \circ X = 1_{X^{-1}(B)}$ et donc $\mathbf{E}(h(X)) = P(X^{-1}(B)) = P_X(B) = \int h(x) dP_X(x)$. Par linéarité on obtient le cas de h étagé. Si h positive, h est la limite croissante d'une suite de fonctions étagées, par exemple :

$$h_n(x) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{[\frac{k}{2^n}, \frac{k+1}{2^n}[}(h(x)) = \sum_{k=0}^{4^n} \frac{k}{2^n} 1_{h^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}[)}(x)$$

qui est étagée mesurable (car h est mesurable). En effet, voyons que $h_n(x) \leq h_{n+1}(x)$, si $h(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}[$, soit $h(x) \in [\frac{2k}{2.2^n}, \frac{2k+1}{2.2^n}[$ et alors $h_n(x) = h_{n+1}(x)$, soit $h(x) \in [\frac{2k+1}{2.2^n}, \frac{2k+2}{2.2^n}[$ et alors $h_n(x) = k/2^n \leq h_{n+1}(x) = (2k+1)/2^{n+1}$. Dans les cas en dehors de ces intervalles, $h_n(x) = 0$ qui est forcément inférieur aux valeurs positives prises par h_{n+1}. Comme $h_n(x) \to h(x)$ par construction, on applique le théorème de convergence monotone aux deux mesures :

$$\mathbf{E}(h(X)) = \lim_{n\to\infty} \mathbf{E}(h_n(X)) = \lim_{n\to\infty} \int h_n(x) dP_X(x) = \int h(x) dP_X(x).$$

Le dernier résultat du cas intégrable est évident par le cas positif pour l'équivalence et par linéarité pour l'égalité. ∎

***Exemple* 9. Loi de $Y = \frac{1}{1+U}$ si $U \sim \mathscr{U}([0,1])$.**

Remarquons que lorsque X et Y sont égales en loi, on voit que, pour toute fonction h, $E[h(X)] = E[h(Y)]$, pourvu que l'un des deux membres existe. En particulier, X et Y ont mêmes paramètres (moyenne, variance, fonction de répartition, fonctions caractéristiques etc).

***Exemple* 10. Cas discret, cas continu.**

Proposition 7 (Lemme de Doob-Dynkin). *Soit X une variable aléatoire, $X : (\Omega, \mathcal{T}, P) \to (E, \mathscr{E})$, $\sigma(X) = \{A = X^{-1}(B), B \in \mathscr{E}\}$ est une tribu (appelé **tribu engendrée par** X) et si $Y : \Omega \to I\!R^n$ v.a. est $\sigma(X)$-mesurable si et seulement si il existe $h : (E, \mathscr{E}) \to (I\!R^n, \mathcal{B}(I\!R^n))$ tel que $Y = h(X)$.*

De même, si $X = (X_i)_{i\in I}$ une famille de variables aléatoires, $X_i : (\Omega, \mathcal{T}, P) \to (E_i, \mathscr{E}_i)$, on note $\sigma(X) = \sigma(X_i^{-1}(B_i), B_i \in \mathscr{E}_i\}$ somme tribu engendrée par les tribus $\sigma(X_i)$.
Preuve : La condition suffisante est évidente. Réciproquement, on raisonne comme pour le transfert par le cas étagé $Y = \sum_i \lambda_i 1_{A_i}$ et $A_i = X^{-1}(B_i)$ et alors $h = \sum \lambda_i 1_{B_i}$ convient. Sinon, si Y positive, on la prend pour limite simple de Y_n étagée de la forme $h_n(X)$ par le cas étagé, et $h(x) = \liminf_{n\to\infty} h_n(x)$ convient car mesurable et car $Y(\omega) = \lim_n h_n(X(\omega)) = h(X(\omega))$ en $X(\omega)$ h_n converge d'après le choix de Y_n. ∎

Moments, variance et rappels sur les espaces L^p

Vous avez vu durant le cours d'intégration, les espaces

$$L^p(\Omega, \mathcal{T}, P) = \{X : \Omega \to \mathbb{R} \ v.a. \ |\mathbf{E}(|X|^p) < \infty\},$$

pour $p \in [1, \infty[$ (ou plutôt l'espace des classes d'équivalence de fonctions égales sur un ensemble de proba 1), espace vectoriel normé complet pour la norme

$$||X||_p = \mathbf{E}(|X|^p)^{1/p}.$$

On définit $||X||_\infty = \lim_{p\to\infty} ||X||_p$ et $L^\infty(\Omega, \mathcal{T}, P) = \{X : \Omega \to \mathbb{R} \ v.a. \ |||X||_\infty < \infty\}$. On rappelle l'inégalité de Hölder :si $p, q \in [1, \infty]$ tels que $1/p + 1/q = 1/r \leq 1$, $X \in L^p, Y \in L^q$ alors $XY \in L^r$ et

$$||XY||_r \leq ||X||_p||Y||_q.$$

On rappelle aussi qu'en prenant $Y = 1$, on obtient :

$$L^p(\Omega, \mathcal{T}, P) \subset L^q(\Omega, \mathcal{T}, P) \ si \ p \geq q.$$

Définition 17. Soit $k \in \mathbb{N}$ Soit X une v.a. dans $L^k(\Omega, P)$. Le *moment d'ordre k* de X est défini par la quantité

$$m_k = \mathbf{E}[X^k] = \int_{-\infty}^{\infty} x^k dP_X(x).$$

Il est important de retenir la conséquence suivante du théorème de Fubini :

Proposition 8. *Soit $X \geq 0$ une variable aléatoire positive, alors pour $p \in]0, \infty[$*

$$E(X^p) = \int_0^\infty dtpt^{p-1}P(X > t).$$

Preuve : On remarque que $X^p = \int 0^X pt^{p-1}dt = \int 0^\infty pt^{p-1}1_{\{X>t\}}dt$. Or par Fubini-Tonelli,

$$E(X^p) = \int dP \int_0^\infty dtpt^{p-1}1_{\{X>t\}} = \int_0^\infty dtpt^{p-1} \int dP1_{\{X>t\}} = \int_0^\infty dtpt^{p-1}P(X > t).$$

■

Définition 18. La *variance* de $X \in L^2(\Omega, P)$ est définie par la quantité $v = \sigma^2 = V[X] = \mathbf{E}[(X - m)^2] = \mathbf{E}[X^2] - m^2$, avec $m = E(X)$. *L'écart-type* de X est la quantité $\sigma = (V[X])^{1/2}$.

La variable centrée réduite associée à X est $U = \frac{X-m}{\sigma}$. On a $\mathbf{E}[U] = 0$ et $V[U] = 1$.

Soient $X_1, ..., X_n \in L^2(\Omega)$, la *matrice de covariance* de $(X_1, ..., X_n)$ est la matrice $n \times n$ (symétrique) $(Cov(X_i, X_j))_{ij}$, avec :

$$Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = E\left[(X_i - E(X_i))(X_j - E(X_j))\right].$$

Exemple 11. La loi $\mathcal{N}(m, \sigma^2)$ a pour moyenne m et variance σ^2, sa variable centrée réduite a pour loi $\mathcal{N}(0, 1)$ (loi normale centrée réduite).

Remarque 3.
1. La moyenne et l'écart-type d'une v.a. X constituent un résumé intéressant de la répartition de X. La moyenne donne une idée de l'ordre de grandeur de la variable et l'écart-type une idée de la dispersion des valeurs de X autour de leur moyenne. Ces deux paramètres interviennent dans plusieurs inégalités.

2. Le recours à la variable centrée réduite permet et facilite certains calculs. Il permet aussi de comparer entre elles de façon cohérente plusieurs variables.

La variance est définie sur l'espace vectoriel $L^2(\Omega)$ des v.a. d'ordre 2.

Proposition 9. *si X et Y sont deux v.a. dans L^2, on a les propriétés suivantes :*

1. $V[c] = 0$ *si* $c \in I\!R$, $V[X] \geq 0$ *et* $V[X] = 0 \Leftrightarrow X = m$.
2. $V[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$.
3. $V(aX + b) = a^2 V(X), a, b \in I\!R$.

Proposition 10 (Inégalité de Markov)**.** *Soit Z une v.a. positive et $a > 0$. Alors,*

$$P(Z \geq a) \leq \frac{\mathbf{E}[Z]}{a}.$$

Proposition 11 (Inégalité de Bienaymé-Tchébychev)**.** *Soit $X \in L^2$ une v.a. de moyenne m et d'écart-type σ. Alors, pour tout $\epsilon > 0$,*

$$P(|X - m| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Preuve : $P(Z \geq a) = \mathbf{E}(1_{Z \geq a}) \leq \mathbf{E}(\frac{Z}{a})$ car $a 1_{Z \geq a} \leq Z$ en décomposant les cas vu $Z \geq 0$. La deuxième inégalité vient du cas $Z = |X - m|^2$ vu que $P(|X - m| \geq \epsilon) = P(Z \geq \epsilon^2)$. ■

2.2 Fonction Caractéristique (Transformée de Fourier)

Définition 19. La *fonction caractéristique* (f.c. ou transformée de Fourier) du v.a. $(X_1, ..., X_n) : \Omega \to \mathbb{R}^n$ est définie par

$$\Phi_{(X_1,...,X_n)}(t_1, ..., t_n) = \mathbf{E}[e^{i\langle t,X\rangle}],$$

pour tout $t = (t_1, ..., t_n) \in \mathbb{R}^n$ et en notant le produit scalaire $\langle t, X\rangle := \sum_{i=1}^n t_i X_i$.

La fonction φ_X caractérise la loi de X par le théorème d'injectivité de la transformée de Fourier/ théorème d'inversion de la transformée de Fourier ci-dessous. On utilisera aussi plus tard au chapitre 2 la fonction caractéristique pour caractériser une notion de convergence, au chapitre 3 pour l'introduction des vecteurs gaussiens qui seront la base du chapitre 5 sur le mouvement brownien. C'est une notion FONDAMENTALE...

Lemme 12. *Soit $X \sim \mathcal{N}(m, \sigma^2)$ de loi normale alors $\Phi_X(t) = exp(-\frac{t^2\sigma^2}{2} + imt)$.*

Preuve : On a vu une preuve à l'exercice 8 du TD 3 de MASS 31 utilisant que la partie imaginaire est nulle par parité et le calcul de la partie réelle en établissant une équation différentielle par intégration dépendant d'un paramètre.

On donne ici une autre preuve par prolongement analytique. Par transfert, on doit montrer $\int \frac{1}{\sigma\sqrt{2\pi}} e^{ixt - \frac{(x-m)^2}{2\sigma^2}} = exp(-\frac{t^2\sigma^2}{2} + imt)$ en faisant le changement de variables $u = (x - m)/\sigma$ on se ramène au cas $\sigma = 1, m = 0$.

En prenant $m = z$ dans le calcul de la densité, on a pour $z \in \mathbb{R}$

$$\int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+z^2-2xz}{2}} = \int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-z)^2}{2}} = 1.$$

Pour $z \in \mathbb{C}$, en appliquant le résultat précédent

$$\sum_{n=0}^{\infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} = \lim_{N\to\infty} \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{N} \frac{|zx|^n}{n!} e^{-\frac{x^2}{2}} \leq \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+|zx|} \leq exp(\frac{|z|^2}{2}) < \infty$$

La première bornitude permet d'appliquer le TCD pour les séries (ou Fubini pour la mesure discrète) et intervertir somme et série :

$$\sum_{n=0}^{\infty} z^n \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} \frac{x^n}{n!} e^{-\frac{x^2}{2}} = \int_{\mathbb{R}} dx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+zx}$$

la fonction de droite est donc la somme d'une série entière $exp(\frac{z^2}{2})$ pour $z \in \mathbb{R}$, donc par identification des coefficients, elle vaut cette valeur pour tout $z \in \mathbb{C}$, en particulier pour $z = it$ et on trouve le résultat. ∎

On démontrera le théorème suivant dans la prochaine section puisque la preuve utilise des propriétés générales de l'indépendance importante à noter pour elles-mêmes :

Théorème 13 (Théorème d'injectivité de la transformation de Fourier)**.** *Deux v.a.* $(X_1, ..., X_n), (Y_1, ..., Y_n)$ *tels que*

$$\Phi_{(X_1,...,X_n)}(t) = \Phi_{(Y_1,...,Y_n)}(t) \forall t \in I\!R^n$$

sont égales en loi, c'est à dire :

$$P_{(X_1,...,X_n)} = P_{(Y_1,...,Y_n)}.$$

De plus, si $\Phi_X \in L^1(I\!R^n, Leb)$ *alors* $P_{(X_1,...,X_n)}$ *a une densité par rapport à la mesure de Lebesgue donnée par (la transformée de Fourier inverse) qui est une fonction continue :*

$$f_{(X_1,...,X_n)}(x) = \frac{1}{(2\pi)^n} \int_{I\!R^n} \Phi_{(X_1,...,X_n)}(t) exp(-i\langle x, t\rangle) dt.$$

3 Indépendance, premiers résultats asymptotiques

Définition 20 (Indépendance)**.** . Soit $n \in N^*$.

1. Les événements de la suite $(A_k)_{1\leq k\leq n}$ sont *indépendants* (ou mutuellement indépendants) si pour toute partie $J \subset [\![1, n]\!]$,
$$P(\cap_{k\in J} A_k) = \prod_{k\in J} P(A_k).$$
2. Les tribus $(\mathcal{A}_k)_{1\leq k\leq n}$ sont *indépendantes* (ou mutuellement indépendantes) si pour tout événement $A_k \in \mathcal{A}_k$ la famille $(A_k)_{1\leq k\leq n}$ est indépendante.
3. Des v.a. $(X_k)_{1\leq k\leq n}$ (à valeur $(E_k, \mathscr{E}_k)$) sont dites indépendantes si les tribus $(\sigma(X_k))_{1\leq k\leq n}$ sont indépendantes, c'est à dire, si et seulement si :
$$\forall A_k \in \mathscr{E}_k,\ P(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

Remarque 4. Lorsque (X_i) sont des v.a. indépendantes et f_i des fonctions mesurables, alors les v.a. $(f_i(X_i))$ sont encore indépendantes. (car $\sigma(f_i(X_i)) \subset \sigma(X_i)$)

Théorème 14 (Loi des v.a. indépendantes)**.** *n variables aléatoires $X_1, ..., X_n$ sont indépendantes si et seulement si la loi du n-uplet $(X_1, ..., X_n)$ est la mesure produit des lois :*

$$P_{(X_1,...,X_n)} = P_{X_1} \otimes ... \otimes P_{X_n}.$$

Remarque 5. En conséquence, en utilisant aussi le théorème de transfert, si $X_i : \Omega \to \mathbb{R}$ et $f : \mathbb{R}^n \to \mathbb{R}_+$ mesurable :

$$\mathbf{E}[f(X_1, ..., X_n)] = \int_{\mathbb{R}^n} f(x_1, ..., x_n) dP_{X_1}(x_1)...dP_{X_n}(x_n)$$

Si f n'est pas positive mais seulement intégrable, (i.e. $\mathbf{E}(|f(X_1, ..., X_n)|] < +\infty$) la formule s'applique aussi. En particulier pour $f_i : \mathbb{R} \to \mathbb{R}_+$ (ou $f_i : \mathbb{R} \to \mathbb{R}$ avec $E(|f_i(X_i)|] < \infty$) mesurables, en utilisant de plus le théorème de Fubini, on obtient :

$$\mathbf{E}[\prod_{i=1}^{n} f_i(X_i)] = \prod_{i=1}^{n} \mathbf{E}[f_i(X_i)].$$

Preuve : Pour $F_i \in \mathscr{E}_i$,

$$P_{(X_1,...,X_n)}(F_1 \times ... \times F_n) = P(X_1 \in F_1, ..., X_n \in F_n)$$

et

$$(P_{X_1} \otimes ... \otimes P_{X_n})(F_1 \times ... \times F_n) = P(X_1 \in F_1)...P(X_n \in F_n).$$

L'indépendance est donc par définition équivalente à l'égalité des 2 mesures sur les produits $F_1 \times ... \times F_n$ ce qui équivaut à l'égalité des mesures par le lemme de classe monotone. ∎

Exemple 12. $U \sim \mathscr{E}(2), V \sim \mathscr{U}([0,1])$ **indépendantes, Mq,** $X = \sqrt{U}\cos(2\pi V), Y = \sqrt{U}\sin(2\pi V)$ **sont indépendantes.**

Corollaire 15 (Regroupement par paquets)**.** *Si $X_1, ..., X_n$ sont des v.a. indépendantes, et $0 < n_1 < n_2... < n_k = n$ Alors $Y_1 = (X_1, ..., X_{n_1}), Y_2 = (X_{n_1+1}, ..., , X_{n_2}), ..., Y_k = (X_{n_{k-1}+1}, ..., X_{n_k})$ sont indépendantes.*

Preuve : Par le Thm et l'indépendance, on a $P_{(X_1,...,X_n)} = P_{X_1} \otimes ... \otimes P_{X_n}$, $P_{Y_i} = P_{X_{n_{i-1}+1}} \otimes ... \otimes P_{X_{n_i}}$ donc par associativité de la mesure produit

$$P_{(Y_1,...,Y_k)} = P_{(X_1,...,X_n)} = P_{X_1} \otimes ... \otimes P_{X_n} = P_{Y_1} \otimes ... \otimes P_{Y_n}$$

d'où l'indépendance par le théorème de nouveau. ∎

Proposition 16 (Caractérisations de l'indépendance)**.** *Soient $(X_1, ..., X_n)$ des v.a. réelles. On a équivalence des assertions :*

1. *$X_1, .., X_n$ sont indépendantes.*
2. *$\forall a_1 \in \mathbb{R}, ..., \forall a_n \in \mathbb{R}$*

$$P(X_1 \le a_1, ..., X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i)$$

3. *$\forall (t_1, ..., t_n) \in \mathbb{R}^n$*

$$\Phi_{(X_1,...,X_n)}(t_1, ..., t_n) = \prod_{i=1}^{n} \Phi_{X_i}(t_i).$$

PREUVE : (i) implique (ii) ou (iii) est évident par le théorème de la loi produit et sa conséquence. (ii) implique (i) vient de la caractérisation de la loi par les fonctions de répartition (Prop 4). (iii) implique (i) vient de même en utilisant l'injectivité de la transformée de Fourier. ■

Définition 21. Une suite $(X_n)_{n\geq 1}$ de variables aléatoires est dite **indépendante** si toutes les sous-suites finies $(X_n)_{1\leq n\leq k}$ sont indépendantes $\forall k \in \mathbb{N}$.

Proposition 17. *Si $(X_n)_{n\geq 1}$ sont des v.a. indépendantes, alors $\sigma(X_1, ..., X_n)$ et $\sigma(X_{n+1}, X_{n+2}, ...)$ sont des tribus indépendantes.*

PREUVE : Par définition et rassemblement par paquet il est clair que $\sigma(X_1, ..., X_n)$ et $\sigma(X_{n+1}, ..., X_{n+K})$ sont indépendantes pour tout K. Or $\mu(B) = P((X_1, ..., X_n) \in A \text{ et } (X_{n+1}, X_{n+2}, ...) \in B)$ et $\nu(B) = P((X_1, ..., X_n) \in A)P((X_{n+1}, X_{n+2}, ...) \in B)$ sont des mesures (finies de même masse) sur $\sigma(X_{n+1}, X_{n+2}, ...)$, qui s'accordent sur la partie génératrice stable par intersection finie $\cup_{K=1}^{\infty}\sigma(X_{n+1}, ..., X_{n+K})$ d'après l'indépendance ci-dessus, donc par le lemme 3 de classe monotone $\mu = \nu$ d'où le résultat. ■

Il n'est pas évident que l'on puisse construire une suite de v.a. indépendantes, mais c'est en effet possible. (cf remarque 7)

3.1 Sommes de variables aléatoires indépendantes, Lois des grands nombres

Vous avez probablement vu en TD de théorie de la mesure la définition de la convolution que l'on rappelle ici et relie aux sommes de variables aléatoires indépendantes.

Définition 22 (Convolution)**.** Soit μ une mesure de Proba sur $S \subset \mathbb{R}^d$ et $f : \mathbb{R} \to \mathbb{R}$ une fonction mesurable telle que pour tout $x \in S$, $y \mapsto f(x - y)$ est dans $L^1(\mathbb{R}^d, \mu)$, la convolution de f et μ est la fonction $f * \mu$ définie par :

$$(f * \mu)(x) = \int_{\mathbb{R}^d} f(x - y)d\mu(y).$$

Si μ est absolument continue par rapport à la mesure de Lebesgue de densité g, on note aussi $f * g$.

Proposition 18. *Soient $X, Y : \Omega \to \mathbb{R}^d$ des v.a. indépendantes :*

1. *$\forall t \in \mathbb{R}^d, \Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$*
2. *Si X_i, Y_i sont dans $L^2(\Omega)$, $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j)$.*
3. *Si $P_X(dx) = f(x)dx, P_Y(dx) = g(y)dy$ alors P_{X+Y} est absolument continue par rapport à Lebesgue (sur $\mathbb{R}^d$) de densité $f * g$ définie Lebesgue p.p. :*

$$P_{X+Y}(dz) = (f * g)(z)dz.$$

4. *Si seulement X est de loi absolument continue mais de densité continue bornée f, alors quel que soit Y, P_{X+Y} est absolument continue par rapport à Lebesgue (sur $\mathbb{R}^d$) de densité $f * P_Y$ (définie partout). De plus, pour tout h continue bornée :*

$$E((h * f)(Y)) = E(h(X + Y)).$$

PREUVE : 1. On a $\Phi_{X+Y}(t) = \mathbf{E}[e^{it(X+Y)}] = \mathbf{E}[e^{itX}e^{itY}] = \mathbf{E}[e^{itX}]\mathbf{E}[e^{itY}] = \Phi_X(t)\Phi_Y(t)$ l'avant dernière égalité par indépendance car $f(x) = e^{itx}$ est bornée donc intégrable (par rapport à une probabilité).

2. En général par bilinéarité $Cov(X_i + Y_i, X_j + Y_j) = Cov(X_i, X_j) + Cov(Y_i, Y_j) + Cov(Y_i, X_j) + Cov(Y_i, X_j)$, mais ici par indépendance les deux derniers termes sont nuls.

3.Il faut d'abord vérifier que $f*g$ est bien définie. Par Fubini-Tonelli vu le caractère positif :

$$\int_{\mathbb{R}^n} dx \int_{\mathbb{R}^n} dy f(x-y)g(y) = \int_{\mathbb{R}^n} dy (\int_{\mathbb{R}^n} dx f(x-y))g(y) = \int_{\mathbb{R}^n} dy g(y) = 1$$

donc $\int_{\mathbb{R}^n} dy f(x-y)g(y)$ existe et est fini p.p.

En prenant h mesurable positive et en appliquant le transfert, on obtient par changement de variables $z = x+y$ dans l'intégrale sur y obtenue par Fubini :

$$E(h(X+Y)) = \int_{\mathbb{R}^{2d}} h(x+y)f(x)dxP_Y(dy) = \int_{\mathbb{R}^{2d}} h(z)f(z-y)dzP_Y(dy) = \int_{\mathbb{R}^d} h(z)(f*P_Y)(z)dz$$

ce qui donne le calcul de densité (égalité de la loi avec seulement le cas $h = 1_B$). Dans le cas de 4. on raisonne pareil sauf que f continue bornée donne $x \mapsto f(x-y)$ intégrable par rapport à la proba P_Y directement. L'application de Fubini vient de $\int_{\mathbb{R}^{2d}} |h(z)f(z-y)|dzP_Y(dy) \leq ||h||_\infty$. L'égalité intermédiaire donne aussi $E(h(X+Y)) = \int_{\mathbb{R}^d}(h*f)(y)P_Y(dy) = E((h*f)(Y))$ par transfert. ■

Exemple 13. **Loi binomiale comme somme de Bernoulli indépendantes**

Théorème 19 (Loi faible des grands nombres)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables indépendantes de même loi (i.i.d), telle que $X_1 \in L^2(\Omega, P)$ et $m = E(X_n)$. Alors pour tout $\epsilon > 0$, on a*

$$\left\| \frac{X_1 + ... + X_n}{n} - m \right\|_2 \to_{n\to\infty} 0,$$

$$P(|\frac{X_1 + ... + X_n}{n} - m| > \epsilon) \to_{n\to\infty} 0.$$

Preuve : Comme $\mathbf{E}(\frac{X_1+...+X_n}{n}) = m$,

$$||\frac{X_1 + ... + X_n}{n} - m||_2^2 = Var(\frac{X_1 + ... + X_n}{n}) = \frac{\sum_{i=1}^n Var(X_i)}{n^2} = \frac{Var(X_1)}{n} \to 0.$$

En appliquant l'inégalité de Bienaimé-Tchébychev, on déduit la seconde convergence vers 0. ■

Théorème 20 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables indépendantes de même loi, telle que $X_1 \in L^1(\Omega, P)$ c'est-à-dire $\mathbf{E}(|X_1|) < \infty$ et soit $m = \mathbf{E}(X_1)$.*

$$P(\{\omega : \frac{X_1(\omega) + ... + X_n(\omega)}{n} \to m\}) = 1.$$

On donnera une preuve en application du chapitre 4 des théorèmes de convergence des martingales. On donne ici une preuve simple avec une hypothèse supplémentaire.

Preuve dans le cas $\mathbf{E}(X_1^4) < \infty$. : On peut supposer $m = 0$, en remplaçant X_i, par $X_i - m$. Alors (cf détail en cours à connaître)

$$\mathbf{E}\left[\left(\frac{X_1 + ... + X_n}{n}\right)^4\right] = \frac{E(X_1^4)}{n^3} + \frac{E(X_1^2X_2^2)3n(n-1)}{n^4} \leq \frac{C}{n^2}$$

donc

$$\mathbf{E}\left[\sum_{n=1}^{\infty}\left(\frac{X_1 + ... + X_n}{n}\right)^4\right] \leq \sum_{n=1}^{\infty}\frac{C}{n^2} < \infty$$

donc l'événement $A = \{\sum_{n=1}^\infty \left(\frac{X_1+...+X_n}{n}\right)^4$ converge $\}$ a probabilité $P(A) = 1$ (sans quoi l'espérance serait infinie), donc $\{\omega : \frac{X_1(\omega)+...+X_n(\omega)}{n} \to 0\} \supset A$ a proba 1. ■

Exemple 14. $\lim_{n\to\infty}\int_{-2}^{2}dx_1...\int_{-2}^{2}dx_n\frac{\sqrt{4-x_1^2}...\sqrt{4-x_n^2}}{(2\pi)^n}f(\frac{x_1+...+x_n}{n}) = ...$ pour $f:[-2,2]\to\mathbb{R}$ continue.
Exemple 15. Aiguille de Buffon.

On va voir dans la section 3.3 que le fait d'avoir probabilité 1 (ou 0 pour leur complémentaire) pour des événements dépendants de limites comme celle de $\frac{X_1+...+X_n}{n}$ est une conséquence très générale de l'indépendance.

3.2 Preuve [Facultative] du Thm d'injectivité de la transformée de Fourier

On va utiliser les lois gaussiennes pour se ramener au cas avec densité tout en exploitant leurs propriétés de stabilité par cette transformée.

Lemme 21. *Soit g_σ la densité sur $\mathbb{R}^n$ d'un n-uplet de variable gaussienne i.i.d. $\mathcal{N}(0,\sigma^2)$. Pour tout $h:\mathbb{R}^n\to\mathbb{R}$ continue bornée, $(h*g_\sigma)(x)\to_{\sigma\to 0}h(x)$. On a même convergence uniforme sur tout compact.*

En terme de convergence en loi, cela signifiera au chapitre 2 que si $(X_1(\sigma),...,X_n(\sigma))$ sont les variables de densités g_σ, alors $x+(X_1(\sigma),...,X_n(\sigma))\to_{\sigma\to 0}x$ en loi en utilisant la proposition 18.(4) au cas $Y=x$.
Preuve : Par transfert et changement de variables

$$(h*g_\sigma)(x)-h(x)=\int_{\mathbb{R}^d}(h(x-\sigma z)-h(x))g_1(z)dz.$$

En prenant, en prenant le supremum sur un compact K :

$$\sup_{x\in K}|(h*g_\sigma)(x)-h(x)|\leq\int_{\mathbb{R}^d}\sup_{x\in K}|(h(x-\sigma z)-h(x))|g_1(z)dz$$

la limite vient de la convergence dominée par une constante $2||h||_\infty$ puisque une constante est intégrable par rapport à une probabilité comme $g_1(z)dz$, et la limite ponctuelle en z vient de la continuité de h qui est donc uniformément continue sur $K+B(0,|z|)$ et donc pour $|\sigma|<1, x-\sigma z, x$ sont dans ce compact de distance $\sigma|z|$ tendant vers 0. Si h est uniformément continue sur $\mathbb{R}^d$ on a même convergence uniforme sur $\mathbb{R}^d$. ■
Preuve du Thm 13 : Pour montrer l'injectivité, par le lemme 4, il suffit de montrer que l'égalité des transformée de Fourier implique égalité de $\mathbf{E}(h(X))$ pour tout h continue bornée.

Or par le lemme précédent, $(h*g_\sigma)(x)\to h(x)$ tout en étant borné par $||h||_\infty$ donc par TCD :

$$\mathbf{E}(h(X))=\lim_{\sigma\to 0}\mathbf{E}((h*g_\sigma)(X))=\lim_{\sigma\to 0}\mathbf{E}(h(X+Y_\sigma))$$

la dernière égalité avec Y_σ de densité g_σ et indépendant de X par la proposition 18 (4) puisque la densité g_σ est continue bornée. Or la transformée de Fourier de $X+Y_\sigma$ est $\Phi_{X+Y_\sigma}(t)=\Phi_X(t)\Phi_{Y_\sigma}(t)$ par la proposition 18 (2) et donc

$$\Phi_{X+Y_\sigma}(t)=\Phi_X(t)exp(-\frac{||t||_2^2\sigma^2}{2})$$

par le calcul du lemme 12. Comme ceci est intégrable, on s'attend à avoir la formule d'inversion de Fourier de la deuxième partie qui va donner $\mathbf{E}(h(X+Y_\sigma))$ en fonction de $\Phi_{X+Y_\sigma}(t)$, nous allons donc la montrer à la main dans ce cas pour conclure la preuve.

Or en interprétant la densité comme une variante de la transformée de Fourier dans le cas gaussien :

$$(g_\sigma * P_X)(x) = \int_{\mathbb{R}^d} \frac{1}{\sigma^d (2\pi)^{d/2}} exp(-\frac{||x-y||_2^2}{2\sigma^2}) P_X(dy) = \int_{\mathbb{R}^{2d}} P_X(dy) dv \frac{1}{\sigma^d (2\pi)^d} exp(-\frac{||v||^2}{2} + i\langle \frac{y-x}{\sigma}, v\rangle))$$

soit par le changement de variables $u = v/\sigma$ de jacobien σ^{-d} on obtient

$$\mathbf{E}(h(X+Y_\sigma)) = \int_{\mathbb{R}^d} dx h(x)(g_\sigma * P_X)(x) = \int_{\mathbb{R}^{3d}} dx P_X(dy) dv h(x) \frac{1}{(2\pi)^d} exp(-\frac{\sigma^2 ||v||^2}{2} + i\langle y-x, v\rangle))$$

soit en appliquant Fubini sur les intégrales en y, v

$$\mathbf{E}(h(X+Y_\sigma)) = \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} exp(-\frac{\sigma^2 ||v||^2}{2} - i\langle x, v\rangle))\Phi_X(v) = \int_{\mathbb{R}^{2d}} dx dv \frac{h(x)}{(2\pi)^d} exp(-i\langle x, v\rangle))\Phi_{X+Y_\sigma}(v)$$

qui est la formule souhaitée qui ne dépend bien que de la transformée de Fourier Φ_X et conclut l'injectivité.

Maintenant si Φ_X est intégrable $|h(x)\Phi_{X+Y_\sigma}(v)| \leq h(x)|\Phi_X(v)|$ est une domination (si h est à support compacte) et puisque $\Phi_{X+Y_\sigma}(v) \to_{\sigma\to 0} \Phi_X(v)$ par les formules précédentes, on obtient par le TCD la formule souhaitée pour la densité à la limite. La continuité de la densité vient du Théorème de continuité des intégrales à paramètres. On remarque qu'en utilisant $\mathbf{E}(h(X))) = \int_{\mathbb{R}^d} dx h(x) f_X(x)$ pour tout h positive continue à support compact, on déduit f_X positive (sinon par continuité elle est négative sur un ouvert dans lequel on peut prendre le support de h pour contredire positivité de l'intégrale) et par convergence monotone et faisant tendre $h \to 1$, on déduit f_X intégrable et densité de proba. D'où on peut utiliser $\mathbf{E}(h(X))) = \int_{\mathbb{R}^d} dx h(x) f_X(x)$ (maintenant valable pour h continue bornée car f_X peut servir de domination) pour identifier $P_X(dx) = f_X(x)dx$ en utilisant le lemme 4. ■

3.3 Loi du 0-1, Lemme de Borel Cantelli

Définition 23. Soit $(X_n)_{n\geq 1}$ une suite de v.a. on note $\mathcal{F}^n = \sigma(X_p, p \geq n)$. La *tribu asymptotique* est la tribu $\mathcal{F}^{+\infty} = \cap_{n\geq 1}\mathcal{F}^n$.

Exemple 16. L'événement de la loi forte des grands nombres $\{\lim_{n\to\infty} \frac{X_1+...+X_n}{n} = m\} \in \mathcal{F}^{+\infty}$.

On va voir que cette tribu satisfait une alternative :

Théorème 22 (Loi du 0-1 de Kolmogorov)**.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes et $\mathcal{F}^{+\infty}$ la tribu asymptotique associée, alors tout $A \in \mathcal{F}^{+\infty}$ vérifie $P(A) = 0$ ou $P(A) = 1$.*

Preuve : On va montrer que $\mathcal{F}^{+\infty}$ est indépendante d'elle même. Soit $A \in \mathcal{F}^{+\infty}$ tel que $P(A) \neq 0$. On a vu de la proposition 17, $A \in \mathcal{F}^{n+1}$ est indépendant de $\sigma(X_1, ...X_n)$, donc comme n est arbitraire, A est indépendant de $\mathcal{B} = \cup_{n=1}^{\infty}\sigma(X_1, ...X_n)$. Or $\mathcal{B}$ est stable par intersection finie, donc les mesures de probabilités P et $P(.|A) = \frac{P(B\cap A)}{P(A)}$, qui s'accordent sur $\mathcal{B}$, s'accordent par le lemme de classe monotone sur $\sigma(B)$. Or $\sigma(X_i) \subset \sigma(B)$ pour tout i, donc $\mathcal{F}^{+\infty} \subset \mathcal{F}^1 \subset \sigma(\mathcal{B})$.

Donc A est indépendant de lui-même, soit $P(A) = P(A\cap A) = P(A)^2$ donc comme on a supposé $P(A) \neq 0$ on déduit $P(A) = 1$. ■

Il est intéressant de savoir distinguer lequel des cas de la conclusion de la loi du 0-1 est vérifiée. Le prochain théorème est important dans ce but.

Définition 24. Soit une suite d'événements (A_n), on appelle

$$\limsup_n A_n = \cap_{n\geq 1} \cup_{k\geq n} A_k = \{\omega : \#\{n : \omega \in A_n\} = \infty\}.$$

$$\liminf_n A_n = \cup_{n\geq 1} \cap_{k\geq n} A_k = \{\omega : \omega \in A_n \text{ à partir d'un certain rang}\}.$$

Remarque 6. 1. Par les formules de de Morgan, $\limsup_n A_n^c = (\liminf_n A_n)^c$, on a aussi $\liminf_n A_n \subset \limsup_n A_n$.

2. Il ne faut pas confondre lim sup, lim inf de suites et d'événements. Pour une suite $x_n \in \mathbb{R}$, $\limsup_n x_n = \inf_{n\geq 1} \sup_{k\geq n} x_k$ (c'est aussi la plus grande valeur d'adhérence), $\liminf_n x_n = \sup_{n\geq 1} \inf_{k\geq n} x_k$. (c'est aussi la plus petite valeur d'adhérence)
Par contre on utilise souvent les relations suivantes si X_n sont des v.a. et M_n des nombres :

$$\limsup_n \{X_n \geq M_n\} \subset \{\limsup_n X_n \geq \liminf_n M_n\},$$

$$\liminf_n \{X_n \leq M_n\} \subset \{\limsup_n X_n \leq \limsup_n M_n\}.$$

On a aussi les relations

$$1_{\liminf_n A_n} = \liminf_n 1_{A_n}, 1_{\limsup_n A_n} = \limsup_n 1_{A_n}.$$

En effet, si $\omega \in \liminf_n A_n$ ssi $\omega \in A_n$ pour n assez grand ssi $1_{A_n}(\omega) = 1$ pour n assez grand ssi $\liminf_n 1_{A_n}(\omega) = 1$. En effet pour une suite de $0, 1$ $\liminf_n 1_{A_n}(\omega) = 1$ ssi $\lim_n 1_{A_n}(\omega) = 1$ ssi la suite est stationnaire égale à 1. De plus $1_{\limsup_n A_n} = 1 - 1_{\liminf_n A_n^c} = 1 - \liminf_n 1_{A_n^c} = \limsup_n(1 - 1_{A_n^c}) = \limsup_n 1_{A_n}$.

Exemple 17. Si $A_n \in \sigma(X_n)$ alors $\limsup_n A_n, \liminf_n A_n \in \mathcal{F}^{+\infty}$ la tribu asymptotique des X_n.

Théorème 23 (Lemme de Borel-Cantelli)**.** *Soit $(A_n)_{n\geq 1}$ une suite d'événements :*

1. *Si $\sum_{n=1}^{\infty} P(A_n) < \infty$ alors $P(\limsup_n A_n) = 0$.*
2. *Si $\sum_{n=1}^{\infty} P(A_n) = \infty$ et la SUITE $(A_n)_{n\geq 1}$ EST INDEPENDANTE, alors $P(\limsup_n A_n) = 1$.*

Preuve : Le point 1. est facile. Par ex, $E(\sum_{n=1}^{\infty} 1_{A_n}) = \sum_{n=1}^{\infty} P(A_n) < \infty$ donc

$$1 = P(\sum_{n=1}^{\infty} 1_{A_n} < \infty) \leq P(1_{A_n} \to 0) \leq P(\limsup_n 1_{A_n} = 0) = P((\limsup_n A_n)^c).$$

Pour le point 2., on utilise l'indépendance pour calculer

$$P(\cap_{n=k}^{k+K} A_n^c) = \prod_{n=k}^{k+K}(1 - P(A_n)) = \exp(\sum_{n=k}^{k+K} \ln(1 - P(A_n))) \leq \exp(-\sum_{n=k}^{k+K} P(A_n)) \to_{K\to\infty} 0$$

car $\ln(1 + x) \leq x$ car $\lim_{K\to\infty} \sum_{n=k}^{k+K} P(A_n) = \infty$. Donc comme $\cap_{n=k}^{k+K} A_n^c$ est une suite décroissante en K, on déduit :

$$P(\cap_{n=k}^{\infty} A_n^c) = \lim_{K\to\infty} P(\cap_{n=k}^{k+K} A_n^c) = 0$$

d'où par union dénombrable $P(\liminf_n A_n^c) = P(\cup_k \cap_{n=k}^{\infty} A_n^c) \leq \sum_k P(\cap_{n=k}^{\infty} A_n^c) = 0$.

∎

En application, on donne la réciproque de la loi des grands nombres.

Proposition 24. *Soit $(X_n)_{n\geq 1}$ une suite de variables indépendantes de même loi, telle que*

$$P(\{\omega : \frac{X_1(\omega) + ... + X_n(\omega)}{n} \text{ converge}\}) = 1,$$

alors $E(|X_1|) < \infty$.

Preuve : On prouve la contraposée, on suppose donc $E(|X_1|) = \infty$.

Or par la proposition 8,

$$E(|X|) = \int_0^\infty dt P(|X| > t) \leq \sum_{n=0}^\infty \int_n^{n+1} dt P(|X| > n) = \sum_{n=0}^\infty P(|X| > n)$$

car pour $t \in [n, n+1] P(|X| > t) \leq P(|X| > n)$.

Donc comme les X_i sont i.d, $\sum_{n=0}^\infty P(|X_n| > n) = \infty$ donc comme elles sont aussi indépendantes, par Borel-Cantelli, $P(\limsup_n \{\frac{|X_n|}{n} > 1\}) = 1$. En utilisant la remarque précédent le lemme, $P(\limsup_n \frac{|X_n|}{n} \geq 1) \geq P(\limsup_n \{\frac{|X_n|}{n} > 1\}) = 1$. Donc p.s. $\limsup_n \frac{|X_n|}{n} \geq 1$ donc $P(\lim_n \frac{|X_n|}{n} = 0) = 0$ Or si S_n/n converge (avec $S_n = X_1 + ... + X_n$) alors $\frac{X_n}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n-1} \frac{(n-1)}{n} \to 0$. Donc $P(\{\omega : \frac{X_1(\omega)+...+X_n(\omega)}{n} \text{ converge}\}) = 0$. ■

Chapitre 2

Notions de Convergence de variables aléatoires.

1 Convergences de v.a. sur le même espace : convergences L^p, en probabilité et presque sûr.

Les trois notions de convergences que l'on a déjà rencontrées sont les suivantes. Elles nécessitent toutes que les variables aléatoires soient définies sur le **même espace** Ω.

Pour des variables à valeurs $\mathbb{R}^d$, on dit que $X = (X_1, ..., X_d) \in L^p$ si $X_i \in L^p$ pour tout i. On note $|.|$ la distance euclidienne sur $\mathbb{R}^d$ (mais peu importe la distance considérée par équivalence des normes)

Définition 25. Soient $X_n : \Omega \to \mathbb{R}^d, X : \Omega \to \mathbb{R}^d$ des variables aléatoires réelles.

1. Si $X_n, X \in L^p$, X_n **converge dans** L^p **vers** X ($p \in [1, \infty]$), noté $X_n \stackrel{L^p}{\to}_{n\to\infty} X$, si
$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = 0,$$
2. X_n **converge en probabilité vers** X , noté $X_n \stackrel{P}{\to}_{n\to\infty} X$, si pour tout $\epsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$
3. X_n **converge presque sûrement vers** X , noté $X_n \stackrel{p.s.}{\to}_{n\to\infty} X$, si
$$P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Nous décrivons maintenant les relations entre ces types de convergences.

Proposition 25. *Si $X_n \stackrel{L^p}{\to}_{n\to\infty} X$ alors $X_n \stackrel{P}{\to}_{n\to\infty} X$ et pour tout $1 \leq q < p$, $X_n \stackrel{L^q}{\to}_{n\to\infty} X$.*

De même, si $X_n \stackrel{p.s.}{\to}_{n\to\infty} X$ alors $X_n \stackrel{P}{\to}_{n\to\infty} X$.

Preuve : L'inégalité de Markov dit $P(|X_n - X| > \epsilon) = P(|X_n - X|^p > \epsilon^p) \leq \frac{\mathbf{E}(|X_n - X|^p)}{\epsilon^p} \to 0$. De plus, si$X_n \stackrel{p.s.}{\to}_{n\to\infty} X$, $1_{|X_n - X| \geq \epsilon} \to 0 p.s.$ donc par convergence dominée $P(|X_n - X| \geq \epsilon) = \mathbf{E}(1_{|X_n - X| \geq \epsilon}) \to 0$. ■

Lemme 26 (de Borel-Cantelli)**.** *Pour que $X_n \xrightarrow{p.s.}_{n\to\infty} X$ il suffit que pour tout $\epsilon > 0$,*

$$\sum_{n=0}^{\infty} P(|X_n - X| > \epsilon) < \infty.$$

En particulier, si $X_n \xrightarrow{P}_{n\to\infty} X$ il existe une sous-suite X_{n_k} tel que $X_{n_k} \xrightarrow{p.s.}_{k\to\infty} X$.

Preuve : On applique Borel-Cantelli aux événements $A_n = \{|X_n - X| > \epsilon\}$, on obtient $P(\limsup_n A_n) = 0$ donc $P(\liminf_n A_n^c) = P(\exists N \forall n \geq N |X_n - X| \leq \epsilon) = 1$. Comme il suffit de prendre une intersection dénombrable sur $\epsilon = 1/p$, on obtient finalement $P\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = P(\forall p \exists N \forall n \geq N |X_n - X| \leq \frac{1}{p}) = 1$.

La première conséquence vient de ce qu'une série convergente a un terme général convergent vers 0. La seconde conséquence, vient que de la suite $P(|X_n - X| > \epsilon)$ tendant vers 0 o peut extraire une sous-suite $P(|X_{n_k} - X| > \epsilon) \leq \frac{1}{2^k}$ qui est donc sommable. ■

Exemple 18. X_n de loi $P_{X_n} = \frac{1}{n^k}\delta_{n^{k+1}} + (1 - \frac{1}{n^k})\delta_0$ converge en proba vers 0, pas dans L^1 et converge p.s. vers 0 ssi $k > 1$.

1.1 Supplément [facultatif] sur la convergence en probabilité

Proposition 27. *Sur $\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{T}, P)$ l'ensemble des v.a. à valeur $\mathbb{R}^d$, on introduit la distance*

$$d(X,Y) = \mathbf{E}(\min(1, |X - Y|)).$$

$(\mathcal{L}^0_{\mathbb{R}^d}(\Omega, \mathcal{T}, P), d)$ est un espace métrique complet et $d(X_n, X) \to 0$ si et seulement si $X_n \xrightarrow{P}_{n\to\infty} X$.

On a la réciproque partiel caractérisant la convergence en probabilité en terme de convergence p.s.

Proposition 28. *$X_n \xrightarrow{P}_{n\to\infty} X$ si et seulement si de toute suite n' d'entiers on peut extraire une sous-suite (n'_k) tel que $X_{n'_k} \xrightarrow{p.s.}_{n\to\infty} X$. En particulier la convergence p.s. n'est pas métrisable.*

1.2 Supplément [facultatif] sur la convergence L^1

Définition 26. Une famille $(X_i)_{i\in I}$ de variable de $L^1(\Omega)$ est dite **uniformément intégrable** (u.i.) si

$$\lim_{c\to\infty} \sup_{i\in I} \int_{|X_i|>c} |X_i| dP = 0.$$

Proposition 29. *Si X_n une suite de v.a. de L^1, alors ($X_n \xrightarrow{P}_{n\to\infty} X$ et (X_n) u.i.) si et seulement si ($X \in L^1$ et $X_n \xrightarrow{L^1}_{n\to\infty} X$).*

2 Convergence en loi, Théorème de Paul Lévy

On note $C_b(\mathbb{R}^d)$ l'espace de Banach (e.v.n complet) des fonctions continues bornées sur $\mathbb{R}^d$ muni de la norme de la convergence uniforme : $||\phi|| = \sup_{x\in\mathbb{R}^d} |\phi(x)|$.

Définition 27. Soit $Y_n : \Omega_n \to \mathbb{R}^d$ une suite de variables aléatoires et μ une probabilité sur $\mathbb{R}^d$. On dit que Y_n **converge en loi vers** μ, notée $Y_n \xrightarrow{\mathcal{L}}_{n\to\infty} \mu$, si

$$\forall \varphi \in C_b(\mathbb{R}^d), \quad \mathbf{E}[\varphi(Y_n)] \to_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x) d\mu(x).$$

Si de plus $Y : \Omega \to \mathbb{R}^d$ une autre v.a., Y_n converge en loi vers Y, notée $Y_n \xrightarrow{\mathcal{L}}_{n\to\infty} Y$, si Y_n converge en loi vers P_Y.

La convergence en loi ne dépend que des lois de Y_n et c'est pour cela que les espaces probabilisés Ω_n n'ont pas besoin d'êtres égaux. On relie d'abord la convergence en loi avec la convergence la plus faible déjà rencontrée, la convergence en probabilité :

Proposition 30. *Si $X_n \xrightarrow{P}_{n\to\infty} X$ alors $X_n \xrightarrow{\mathcal{L}}_{n\to\infty} X$.*

PREUVE : Par l'absurde, soit φ tel que $\mathbf{E}[\varphi(X_n)] \not\to_{n\to\infty} \mathbf{E}[\varphi(X)]$, on prend donc $\epsilon > 0$ et une sous-suite X_{n_k} telle que $|\mathbf{E}[\varphi(X_{n_k}) - \varphi(X)]| \geq \epsilon$. Or si $X_{n_k} \xrightarrow{P}_{k\to\infty} X$ on aurait une sous-sous-suite $X_{n_{k_p}} \xrightarrow{p.s.}_{p\to\infty} X$ par le lemme de Borel-Cantelli 26 mais, comme φ est continue cela implique $\varphi(X_{n_{k_p}}) \xrightarrow{p.s.}_{p\to\infty} \varphi(X)$ et par convergence dominée (par la constante $||\varphi||$) cela impliquerait $\mathbf{E}[\varphi(X_{n_{k_p}})] \to_{p\to\infty} \mathbf{E}[\varphi(X)]$. La contradiction recherchée. ∎

On cherche ensuite des définitions équivalentes. On note $C_c(\mathbb{R}^d)$ les fonctions continues à support compact (c'est à dire nulles en dehors d'un compact).

Proposition 31. *1. Soit $X_n : \Omega_n \to \mathbb{R}^d$ une suite de variables aléatoires et μ une probabilité sur $\mathbb{R}^d$. Alors $X_n \xrightarrow{\mathcal{L}}_{n\to\infty} \mu$, si et seulement si*

$$\forall \varphi \in C_c(\mathbb{R}^d), \quad \mathbf{E}[\varphi(X_n)] \to_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x) d\mu(x).$$

2. (cas des v.a. entières) Si $Y_n : \Omega_n \to \mathbb{Z}^d$ et μ une probabilité sur $\mathbb{Z}^d$ alors $Y_n \xrightarrow{\mathcal{L}}_{n\to\infty} \mu$, si et seulement si

$$\forall x \in S, \quad P(Y_n = x) \to_{n\to\infty} \mu(\{x\}).$$

Exemple 19. Limite en loi de X_n de loi binomiale $\mathcal{B}(n, p = \frac{\lambda}{n})$.

PREUVE : 1. L'implication directe est évidente puisque $C_c(\mathbb{R}^d) \subset C_b(\mathbb{R}^d)$. Pour la réciproque, on prend $h_n \in C_c(\mathbb{R}^d)$ suite croissante à valeur $[0,1]$ tendant vers 1 (par exemple $h_n(x) = \max(1 - d(x, B(0,n)), 0)$ qui vaut 1 sur la boule $B(0,n)$.) comme $h_m\phi$ est à support compact $\mathbf{E}((h_m\phi)(X_n) \to_{n\to\infty} \int_{\mathbb{R}^d} \varphi(x) h_m(x) d\mu(x)$.

Or par positivité de $1 - h_m$:

$$|\mathbf{E}((1-h_m)(X_n)\varphi(X_n))| \leq ||\varphi|| E((1-h_m)(X_n))$$

et de même

$$|\int (1-h_m)\varphi d\mu| \leq ||\varphi|| \int (1-h_m) d\mu.$$

Comme d'habitude, on utilise l'inégalité triangulaire :

$$\begin{aligned}\left|\mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d} \varphi(x) d\mu(x)\right| &\leq \left|\mathbf{E}[(h_m\varphi)(X_n)] - \int_{\mathbb{R}^d} h_m\varphi(x) d\mu(x)\right| \\ &\quad + |\mathbf{E}((1-h_m)(X_n)\varphi(X_n))| + |\int (1-h_m)\varphi d\mu|\end{aligned}$$

donc comme le premier terme tend vers 0, en utilisant la sous additivité de la lim sup et les deux bornes précédentes pour les 2 derniers termes, on a :

$$\limsup_{n\to\infty}\left|\mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d}\varphi(x)d\mu(x)\right| \le \limsup_{n\to\infty}||\varphi||\left(E((1-h_m)(X_n)) + \int(1-h_m)d\mu\right).$$

L'utilisation de la limsup permet la convergence de X_n de nouveau sur la borne car $(E((1-h_m)(X_n)) = 1 - E(h_m(X_n)) \to 1 - \int h_m d\mu = \int(1-h_m)d\mu$. Donc

$$\limsup_{n\to\infty}\left|\mathbf{E}[\varphi(X_n)] - \int_{\mathbb{R}^d}\varphi(x)d\mu(x)\right| \le 2||\varphi||\int(1-h_m)d\mu \to_{m\to\infty} 0,$$

et la dernière convergence par convergence monotone est fait alors que le premier terme ne dépend pas de m et le second pas de n.

Pour 2. $P(Y_n = x) = \mathbf{E}[1_{\{x\}}(Y_n)] = \mathbf{E}(h_\epsilon(Y_n))$ pour une fonction h_ϵ continue égale à 1 sur x et 0 en dehors de $B(x,\epsilon)$ donc 0 sur les autres points de $\mathbb{Z}^d$ si $\epsilon \le 1/2$). Donc pour l'implication directe $P(Y_n = x) \to \mu(h_\epsilon) \to_{\epsilon\to 0} \mu(\{x\})$ car en faisant tendre $\epsilon \to 0$, par convergence monotone on obtient la limite souhaitée. Pour la réciproque Il suffit de noter pour h a support compacte $E(h(Y_n)) = \sum_{x\in\mathbb{Z}^d, h(x)\neq 0} h(x)P(Y_n = x)$ qui comme la somme est finie tend vers $\sum_{x\in\mathbb{Z}^d, h(x)\neq 0} h(x)\mu(\{x\})$. Ceci montre que Y_n converge en loi vers $\sum_{x\in\mathbb{Z}^d}\mu(\{x\})\delta_x = \mu$ car μ est une probabilité sur $\mathbb{Z}^d$. (si on ne savait pas cela, on ne pourrait exclure une perte de masse)

■

Le résultat principal est la caractérisation en terme de fonctions caractéristiques :

Théorème 32 (Lévy). *Soit $Y_n : \Omega_n \to \mathbb{R}^d$ une suite de variables aléatoires et $Y : \Omega \to \mathbb{R}^d$ une autre v.a., alors Y_n converge en loi vers Y si et seulement si :*

$$\forall t \in \mathbb{R}^d, \Phi_{Y_n}(t) \to_{n\to\infty} \Phi_Y(t).$$

Exemple 20. (X_i) i.i.d $\mathcal{N}(0,1)$ limite en loi de $\frac{1}{n^{5/2}}\sum_{k=1}^n k^2 X_k$.

Exemple 21. Soit (X_n) i.i.d de loi de bernoulli $\mathcal{B}(1/2)$. Mq $\sum_{k=1}^n \frac{1}{2^k}X_k \xrightarrow{\mathcal{L}}_{n\to\infty} \mathscr{U}([0,1])$.

Remarque 7. Comme la convergence est partout absolue, la limite U redonne $X_k = Ent(2^k U) - 2Ent(2^{k-1}U)$ sont indépendants pour U uniforme. La mesure de Lebesgue donne donc une suite de v.a. de Bernoulli indépendante. et on construit à partir de là n'importe quelle suite de v.a. indépendantes.

Démonstration. Comme $x \to e^{itx}$ est une fonction continue bornée, l'implication suit directement de la définition. Réciproquement, soit $h \in C_c(\mathbb{R}^d)$, on a vu dans la preuve du théorème d'injectivité de la transformée de Fourier que

$$\mathbf{E}((h * g_\sigma)(X)) = \int_{\mathbb{R}^{2d}} dxdv \frac{h(x)}{(2\pi)^d} exp(-\frac{\sigma^2||v||^2}{2} - i\langle x, v\rangle))\Phi_Y(v)$$

et de même pour Y_n comme h est à support compacte $h(x)exp(-\frac{\sigma^2||v||^2}{2})$ est intégrable par rapport à la mesure de Lebesgue et donne une domination (vu $|\Phi_{Y_n}(v)| \le 1$). Or par hypothèse $\Phi_{Y_n}(v) \to \Phi_Y(v)$ donc la convergence dominée :

$$\mathbf{E}((h * g_\sigma)(Y_n)) \to \mathbf{E}((h * g_\sigma)(Y)).$$

Or, d'après la preuve du lemma 21, $(h * g_\sigma)$ converge uniformément vers h car h est uniformément continue (car à support compact et par le Théorème de Heine) donc pour ϵ fixé on a $||(h*g_\sigma)-h|| \leq \epsilon$

D'où $\limsup_{n\to\infty} |\mathbf{E}(h(Y_n)) - \mathbf{E}(h(Y))| \leq 2\epsilon + \limsup_{n\to\infty} |\mathbf{E}((h * g_\sigma)(Y_n)) - \mathbf{E}((h * g_\sigma)(Y))| = 2\epsilon \to_{\epsilon\to 0} 0$. Donc $\mathbf{E}(h(Y_n)) \to \mathbf{E}(h(Y))$ pour tout h continue à support compact et, par la proposition précédente, alors Y_n converge en loi vers Y. □

2.1 Version forte du Théorème de Paul Lévy [Preuve facultative]

Dans la version forte, on n'a pas besoin de supposer que la limite des fonctions caractéristiques est une fonction caractéristique, on obtient ce résultat. On a aussi le résultat utile pour les transformées de Laplace.

Théorème 33 (Lévy)**.** *1. Soit $Y_n : \Omega_n \to I\!R^d$ une suite de vecteurs aléatoires, on suppose que*

$$\forall t \in I\!R^d, \Phi_{Y_n}(t) \to_{n\to\infty} \varphi(t)$$

et que sa limite φ est continue en $t = 0$.
Alors, Y_n converge en loi vers une probabilité μ sur $\mathcal{B}(I\!R^d)$ et $\varphi(t) = \int_{I\!R^d} e^{i\langle x,t\rangle} d\mu(x)$.

2. Soit $Y_n : \Omega_n \to I\!R_+$ une suite de variables aléatoires positives, on suppose que la transformée de Laplace $\mathscr{L}_{Y_n}(t) = \mathbf{E}[exp(-Y_n t)]$ converge :

$$\forall t \in I\!R_+, \mathscr{L}_{Y_n}(t) \to_{n\to\infty} h(t)$$

et que sa limite h est continue en $t = 0$.
Alors, Y_n converge en loi vers une probabilité μ sur $\mathcal{B}([0,\infty[)$ dont la transformée de Laplace est $h(t) = \int_0^\infty e^{-tx} d\mu(x)$.

Exemple 22. Si $Y_n \sim \mathcal{N}(0, n)$, $\Phi_{Y_n}(t) \to 1_{\{0\}}(t)$ mais Y_n ne converge pas en loi, $\frac{Y_n}{1+|Y_n|} \xrightarrow{\mathcal{L}}_{k\to\infty} \frac{\delta_1+\delta_{-1}}{2}$ (Y_n perd donc moitié de sa masse en $+\infty$, moitié de sa masse en $-\infty$,)

La preuve est assez technique et donc FACULTATIVE (la suite de cette section, définitions comprises est donc facultative). L'idée est en 2 étapes, de montrer que la condition de continuité implique qu'aucune masse ne peut s'échapper vers l'infini. On rend ceci précis avec la notion de tension (suite tendue). Ensuite, on voit comment la tension implique une notion de compacité pour la convergence faible et on conclut.

Définition 28. Une suite $X_n : \Omega \to \mathbb{R}^d$ de vecteurs aléatoires est **tendue** si pour tout $\epsilon > 0$ il existe K compact tel que, pour tout n : $P(X_n \in K) \geq 1 - \epsilon$.

Lemme 34. *Pour toute v.a. $X : \Omega \to I\!R$, resp. tout v.a. positive $Y : \Omega \to I\!R_+$,*

$$P(|X| \geq \frac{2}{u}) \leq \frac{1}{u}\int_{-u}^{u}(1 - \mathrm{Re}\,\Phi_X(t))dt, \quad P(Y \geq \frac{1}{u}) \leq \frac{e}{u}\int_0^u (1 - \mathscr{L}_Y(t))dt.$$

En conséquence, si une suite satisfait l'hypothèse du Thm 33.1 ou 2., elle est tendue.

Preuve : Un calcul donne $\int_{-u}^{u} du(1 - e^{itx})dt = 2u - \frac{2\sin(ux)}{x}$ donc en utilisant Fubini (et introduisant la partie réelle sans risque puisque l'expression est déjà réelle) :

$$\frac{1}{u}\int_{-u}^{u} du(1 - \mathrm{Re}\,\Phi_X(t))dt \geq 2\int_{\mathbb{R}}\left(1 - \frac{\sin(ux)}{ux}\right)dP_X(x) \geq 2\int_{|x|u\geq 2}\left(1 - \frac{\sin(ux)}{ux}\right)dP_X(x) \geq P(|X| \geq \frac{2}{u})$$

car $1 - \frac{\sin(ux)}{ux} \geq 1/2$ si $|x|u \geq 2$ et est toujours positif. Pour l'application aux coordonnées de la suite Y_n, on obtient pour $u = 2/M$:$\limsup_{n\to\infty} P(|(Y_n)_i| \geq M) \leq \limsup_{n\to\infty} \frac{M}{2} \int_{-2/M}^{2/M} (1 - \mathrm{Re}\,\Phi_{Y_n}(0, ..., t_i, 0, ..., 0))dt_i = \frac{M}{2} \int_{-2/M}^{2/M} (1 - \mathrm{Re}\,\varphi(0, ..., t_i, 0, ..., 0))dt_i$, en utilisant le théorème de convergence dominée (par une constante) pour le calcul de la limite. Pour ϵ fixé dans la définition de la tension, on trouve, M tel que pour tout i, l'intégrande (donc l'intégrale) est uniformément bornée par $\epsilon/2d$, donc par définition de la limsup, on trouve k tel que $\forall n \geq k, P(\exists i, |(Y_n)_i| \geq M)) \leq d(\epsilon/d) = \epsilon$. En choisissant M plus grand on réalise de même pour le nombre fini des premières coordonnées (un nombre fini de v.a. est toujours tendue dans $\mathbb{R}^d$ par convergence dominée $P(\exists i, |(Y_n)_i| \geq M))) \to_{M\to\infty} 0$). On conclut vu que la boule de rayon M pour la norme infini est compacte.

Dans le cas positif, $\frac{1}{u} \int_0^u du(1 - e^{-tx})dt = \frac{e^{-ux}-1+ux}{ux} \geq e^{-1} 1_{ux\geq 1}$ (par un tableau de variation) et le raisonnement est le même. ■

Le résultat clé qui donne de la compacité est le suivant (voir Barbe Ledoux p 128 pour une preuve). Il est surtout basé sur un résultat d'analyse fonctionnelle : le Théorème de Banach-Alaoglu sur la compacité dans les espaces duaux (dont on verra un cas particulier pour les espaces L^p dans la section 1.3 du chapitre 5, elle-même facultative).

Théorème 35 (Prokhorov). *Soit $X_n : \Omega \to \mathbb{R}^d$ une suite tendue de vecteurs aléatoires. Alors il existe une sous-suite X_{n_k} et une mesure de probabilité μ telle que $X_{n_k} \xrightarrow{\mathcal{L}}_{k\to\infty} \mu$.*

Preuve du Thm 33 : 1. Par le théorème et le lemme, toute sous-suite de Y_n a une sous-suite convergeant en loi vers une mesure de proba. Par le sens simple de l'implication du premier Thm de Lévy, sa transformée de Fourier est φ. Donc on a maintenant l'hypothèse du sens difficile de ce premier théorème pour conclure la convergence en loi vers cette proba μ de transformée de Fourier φ.

2. On montre de même la tension par le lemme, on en déduit que toute sous-suite de Y_n a une sous-suite convergeant en loi vers une mesure de proba μ. Par la réciproque simple, la transformée de Laplace de cette proba est h. Donc toutes les sous-suites ont la même limite en loi (on utilise théorème d'inversion pour la transfo de Laplace, cf TD 2 pour déduction à partir de Fourier). Prenons f continue bornée $E(f(Y_n))$ a une sous-suite convergente par compacité dans $\mathbb{R}$, et toutes ces sous-sous-suite convergent vers la même limite $\int f d\mu$, on en déduit aisément par l'absurde que la suite de départ doit converger vers cette limite (sans quoi on construit une sous-suite loin de cette limite qui n'a donc pas de sous-sous-suite convergeant vers cette limite.) Cette convergence donne la convergence en loi. ■

2.2 Suppléments sur la convergence en loi.

On a aussi les caractérisations équivalentes suivantes (admises) de la convergence en loi.

Théorème 36. 1. *Soit $Y_n : \Omega_n \to \mathbb{R}$ une suite de variables aléatoires et μ une probabilité. Y_n converge en loi vers μ si et seulement si pour tout*

$$\forall x \in \mathbb{R}, \mu(\{x\}) = 0 \Rightarrow \lim_{n\to\infty} P(Y_n \leq x) = \mu(]-\infty, x]).$$

2. *(Lemme de Skohokhod)[facultatif] Soient $Y_n : \Omega_n \to \mathbb{R}^d$ une suite de vecteurs aléatoires et Y un vecteur aléatoire, $Y_n \xrightarrow{\mathcal{L}}_{n\to\infty} Y$ si et seulement si il existe un espace probabilisé $(\Omega, \mathcal{T}, P)$, $X_n, X : \Omega \to \mathbb{R}^d$ tels que $P_{X_n} = P_{Y_n}, P_X = P_Y$ et $X_n \xrightarrow{\mathcal{L}}_{n\to\infty} X$.*

Dans le premier énoncé, il est important que l'on ne considère les limites que pour les points tels que $\mu(\{x\}) = 0$.

Exemple 23. Par exemple, Soit $P_{Y_n} = \mathcal{N}(0, 1/n)$, on a vu que $Y_n \xrightarrow{\mathcal{L}}_{n\to\infty} \delta_0$. Mais $P(Y_n \le 0) = 1/2$ par symétrie qui ne converge pas vers $\delta_0(]-\infty, 0]) = 1$ car 0 est un point tel que $\delta_0(\{0\}) = 1 \neq 0$.

Exemple 24. $x_n \to x$ si et seulement si $x_n \xrightarrow{\mathcal{L}}_{n\to\infty} \delta_x$ (moins évident avec Fourier...)

Idées de Preuve : 1. On peut prendre Y de loi μ. Pour l'implication directe, on prend $f_m(x) = 1$ si $x \le y - 1/m$, 0 si $x \ge y$ continue linéaire entre les deux. de même $g_m(x) = 1$ si $x \le y$, 0 si $x \ge y + 1/m$ continue linéaire entre les deux. on a $\mathbf{E}(f_m(Y_n)) \to \mathbf{E}(f_m(Y))$ $\mathbf{E}(g_m(Y_n)) \to \mathbf{E}(g_m(Y))$.

Puis on utilise le bon comportement de la limsup et liminf avec les inégalités et $f_m \le 1_{]-\infty,y]} \le g_m$ d'où

$$\limsup_{n\to\infty} P(Y_n \le y) \le \limsup_{n\to\infty} \mathbf{E}(g_m(Y_n)) = \mathbf{E}(g_m(Y)) \to_{m\to\infty} P(Y \le y)$$
$$\liminf_{n\to\infty} P(Y_n \le y) \ge \liminf_{n\to\infty} \mathbf{E}(f_m(Y_n)) = \mathbf{E}(f_m(Y)) \to_{m\to\infty} P(Y < y)$$

les limites en m par convergence monotone ou dominée (par une constante). Les deux limites sont égales si $P(Y = y) = \mu(\{y\}) = 0$.

Pour la réciproque, on utilise que 2 n'utilise que cette conséquence, et on utilise que convergence p.s. implique convergence en loi.

Pour 2, $d = 1$ on utilise un modèle canonique concret (sur [0,1]) utilisant l'inverse de la fonction de répartition. ∎

3 Une application : le théorème central limite

Pour le cas vectoriel du théorème central limite, on a besoin de la définition d'un vecteur gaussien que l'on étudiera plus au prochain chapitre. Pour $X = (X_1, ...X_d)$ on note $\mathbf{Cov}(X)$ la matrice de covariance $(\mathbf{Cov}(X))_{i,j} = Cov(X_i, X_j)$. On rappelle qu'une matrice symétrique réelle est positive si ces valeurs propres sont toutes positives.

Définition 29. Soit $C \in M_d(\mathbb{R})$ une matrice symétrique positive. Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ avec $X_i \in L^2$ est appelé **vecteur gaussien centré de covariance** C, dite loi $\mathcal{N}(0, C)$, si

$$\forall t \in \mathbb{R}^d, \Phi_X(t) = \mathbf{E}[e^{i\langle t,X\rangle}] = exp(-\frac{\langle t, Ct\rangle}{2}).$$

On rappelle que $\langle t, s\rangle = \sum_i t_i s_i$ est le produit scalaire et donc que $\langle t, Ct\rangle = \sum_{ij} t_i C_{ij} t_j$. On remarquera que cela coïncide avec la loi normale usuelle dans le cas $d = 1$.

On verra l'existence des vecteurs gaussiens au prochain chapitre et une construction simple. Sinon, l'application du théorème de Levy fort dans le théorème ci-dessous donnera aussi l'existence dès que C sera la matrice de covariance d'une loi de probabilité. (Et on verra encore au prochain chapitre que toute matrice réelle symétrique positive peut être obtenue). L'utilisation du théorème de Paul Lévy va être basée sur un développement limité des fonctions caractéristiques.

Lemme 37. *Soit un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ avec $X_i \in L^2$. Alors Φ_X est de classe $\mathcal{C}^2$ et son développement de Taylor en 0 est :*

$$\Phi_X(t) = 1 + i\sum_{j=1}^{d} t_j E[X_j] - \frac{1}{2}\sum_{k,j=1}^{d} t_k t_j E[X_k X_j] + o(t^2).$$

PREUVE : $\varphi : (t,\omega) \to e^{it\langle t, X(\omega)\rangle}$ est intégrable en ω et $\mathcal{C}^\infty$ en t, et ces deux premières dérivées sont :

$$\frac{\partial \varphi}{\partial t_j} = iX_j e^{it\langle t, X(\omega)\rangle}, \quad \frac{\partial^2 \varphi}{\partial t_k \partial t_j} = -X_k X_j e^{it\langle t, X(\omega)\rangle},$$

qui sont dominés par

$$|\frac{\partial \varphi}{\partial t_j}| \leq |X_j|, |\frac{\partial^2 \varphi}{\partial t_k \partial t_j}| \leq |X_k X_j|$$

qui sont intégrables car $X_i \in L^2$ et en particulier par Cauchy-Schwartz, $\mathbf{E}[|X_j X_k|] \leq ||X_j||_2 ||X_k||_2$. Donc d'après le théorème de dérivation avec condition de domination $\Phi_X(t) = \int \varphi(t,\omega) dP(\omega)$ est $\mathcal{C}^2$ et ces dérivées sont :

$$\frac{\partial \Phi_X}{\partial t_j}(t) = \mathbf{E}[iX_j e^{it\langle t, X\rangle}], \quad \frac{\partial^2 \Phi_X}{\partial t_k \partial t_j}(t) = -\mathbf{E}(X_k X_j e^{it\langle t, X\rangle}).$$

L'énoncé vient donc de la formule de Taylor-Young en 0. ■

Théorème 38. *Soit (X_n) une suite de vecteurs aléatoires $X : \Omega \to I\!R^d$ indépendants de même loi et dans L^2. Alors*

$$\frac{(X_1 + ... + X_n - n\mathbf{E}[X_1])}{\sqrt{n}} \xrightarrow{\mathcal{L}}_{n\to\infty} \mathcal{N}(0, \mathbf{Cov}(X_1)).$$

PREUVE : En remplaçant X_i par $X_i - \mathbf{E}[X_i]$ on peut supposer et on suppose $\mathbf{E}[X_1] = 0$. On calcule la fonction caractéristique en utilisant d'abord l'indépendance (cf lemme 18) puis le fait que les X_i ont même loi.

$$\Phi_{\frac{X_1+...+X_n}{\sqrt{n}}}(t) = \mathbf{E}\left[exp(i\langle t, \frac{X_1 + ... + X_n}{\sqrt{n}}\rangle)\right] = \prod_{j=1}^{n} \mathbf{E}\left[exp(i\langle t, \frac{X_j}{\sqrt{n}}\rangle)\right] = \left(\Phi_{X_1}(\frac{t}{\sqrt{n}})\right)^n.$$

En utilisant le dL du lemme précédent, on obtient une suite (complexe) $\epsilon_n \to 0$ telle que

$$\Phi_{X_1}(\frac{t}{\sqrt{n}}) = \left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n} + \frac{\epsilon_n}{n}\right) = \left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n}\right)(1 + \frac{\epsilon_n}{n}\left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n}\right)^{-1}).$$

Si on pose $\eta_n = \epsilon_n \left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n}\right)^{-1} \to 0$, vérifions que $(1 + \frac{\eta_n}{n})^n \to 1$ (ce qui est bien connu pour le cas réel en utilisant le log mais peut être moins dans le cas η_n complexe.)= On utilise la formule du binôme :

$$|(1 + \frac{\eta_n}{n})^n - 1| \leq \sum_{k=1}^{n} C_n^k \frac{|\eta_n^k|}{n^k} = |(1 + \frac{|\eta_n|}{n})^n - 1| \leq |\eta_n|(1 + \frac{|\eta_n|}{n})^{n-1} \to 0,$$

la deuxième inégalité venant du théorème des accroissement fini pour $x > 1$,

$$|f(x) - f(1)| = |(1 + x)^n - 1| \leq |x - 1| \sup f'(x) = |x - 1| n(1 + x)^{n-1}.$$

En conclusion

$$\Phi_{\frac{X_1+...+X_n}{\sqrt{n}}}(t) \sim_{n\to\infty} \left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n}\right)^n = \exp[n \ln\left(1 - \frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2n}\right)] \to_{n\to\infty} exp(-\frac{\langle t, \mathbf{Cov}(X_1)t\rangle}{2}).$$

Comme la limite est la fonction caractéristique de la loi gaussienne centrée de covariance $\mathbf{Cov}(X_1)$ le théorème de Paul Lévy conclut. ■

Exemple 25. Mq $u_n = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \to 1/2$.

Chapitre 3

Vecteurs Gaussiens.

1 Définition générale et transformée de Fourier

Définition 30. Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est un **vecteur gaussien** si toute combinaison linéaire $\sum_{i=1}^d \lambda_i X_i$ avec $\lambda_i \in \mathbb{R}$ est de loi normale sur $\mathbb{R}$.

(nécessairement $\mathcal{N}(\sum_{i=1}^d \lambda_i \mathbf{E}(X_i), \sum_{i,j=1}^d \lambda_i \lambda_j Cov(X_i, X_j))$.)

Théorème 39. *Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est un* ***vecteur gaussien*** *si et seulement si sa fonction caractéristique est*

$$\Phi_X(t) = exp(i\langle m, t\rangle - \frac{1}{2}\langle t, Ct\rangle)$$

et alors $m = E(X), C = \mathbf{Cov}(X)$. Dans ce cas on note $X \sim \mathcal{N}(m, C)$

La notation et la définition sont donc cohérentes avec la définition 29 d'un vecteur gaussien centré intervenant dans le TCL.

Preuve : Par définition $\Phi_X(t) = \mathbf{E}[e^{i\langle X,t\rangle}] = \Phi_{\langle X,t\rangle}(1)$ or par le lemme 12 (et injectivité de la transformée de Fourier), $\langle X, t\rangle = \sum_{i=1}^n t_i X_i$ est de loi $\mathcal{N}(\sum_{i=1}^d t_i \mathbf{E}(X_i), \sum_{i,j=1}^d t_i t_j Cov(X_i, X_j))$ si et seulement si sa transformée de Fourier est $\Phi_{\langle X,t\rangle}(l) = exp(i\langle m, lt\rangle - \frac{l^2}{2}\langle t, Ct\rangle)$ soit la valeur dans la caractérisation. ∎

Corollaire 40. *Deux vecteurs gaussiens ont la même loi si et seulement si ils ont même espérance et même matrice de covariance.*

Corollaire 41. *Si un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est gaussien, alors pour toute application linéaire $A : \mathbb{R}^d \to \mathbb{R}^m$ et tout vecteur $n \in \mathbb{R}^m$, $Y = n + A(X) : \Omega \to \mathbb{R}^m$ est un vecteur gaussien. Plus précisément, si $X \sim \mathcal{N}(m, C)$ alors $Y \sim \mathcal{N}(n + A(m), ACA^T)$.*

On rappelle que $(A^T)_{ij} = A_{ji}$ désigne la transposée de A.

Preuve : Pour tout $\lambda_1, ..., \lambda_m$, $\sum_{i=1}^m \lambda_i Y_i = \sum_{i=1}^m \lambda_i m_i + \sum_{i=1}^m \sum_{j=1}^d \lambda_i A_{ij} X_j = c + \sum_{j=1}^d \mu_j X_j$. avec $\mu_j = \sum_{i=1}^m \lambda_i A_{ij}$ c'est donc la translaté d'une v.a. gaussienne donc une v.a. gaussienne de $\mathbb{R}$. Comme les λ_i sont arbitraires, Y est gaussien. Il ne reste qu'à calculer son espérance et sa variance. Par linéarité $\mathbf{E}(Y) = n + A\mathbf{E}(X)$ et par bilinéarité pour la covariance

$$\mathbf{Cov}(Y_i, Y_j) = \sum_{i=1}^m \sum_{j=1}^m A_{ik} Cov(X_k, X_l) A_{jl} = (ACA^T)_{ij}$$

∎

Exemple 26. Soit (X_1, X_2, X_3) un vecteur gaussien de loi $\mathcal{N}(m, \Gamma)$ avec $m = (4, 1, 0)$ et $\Gamma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$. Loi de $Y = (X_2, 3X_3 - 2X_2)$?

Exemple 27. Soit $(X_1, X_2, X_3, Y_1, Y_2, Y_3)$ de loi $\mathcal{N}(0, I_6)$ quelle est la loi de $Z = (\frac{\sum_{i=1}^3 X_i Y_i}{\sqrt{\sum_{i=1}^3 Y_i^2}}, Y_1, Y_2, Y_3)$? (attention la transformation n'est PAS linéaire et on ne peut pas appliquer le résultat précédent.) Montrer que $Z \sim \mathcal{N}(0, I_4)$.

Exemple 28. (cf TD) si $(X, Y) \sim \mathcal{N}(0, I_2)$, $P_{X/Y}(dz) = \frac{1}{\pi(1+z^2)} dz$ est la loi de Cauchy.

On va donner une construction canonique d'un vecteur $\mathcal{N}(m, C)$ en utilisant la forme du corollaire précédent.

2 Construction et critère d'absolue continuité

On appelle *variable gaussienne standard* notée γ_d une variable $\mathcal{N}(0, I_d)$ sur $\mathbb{R}^d$. On rappelle que pour une matrice positive $C = ODO^T$ dont on a une diagonalisation ($D = diag(\lambda_1, ..., \lambda_d)$ diagonale à coefficient positif, O orthogonale) la racine carrée de C est $\sqrt{C} = O\sqrt{D}O^T$ avec $D = diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_d})$. On a donc $\sqrt{C}$

Théorème 42. *Si $X = (X_1, ..., X_d)$ variable gaussienne standard γ_d, $m \in \mathbb{R}^d$ et $C \in M_d(\mathbb{R})$ une matrice positive. Soit $B = \sqrt{C}$ la racine carrée de C. Alors $Y = m + BX$ est un vecteur gaussien de loi $\mathcal{N}(m, C)$.*

Cela donne une construction d'un vecteur $\mathcal{N}(m, C)$. La preuve est une application directe du dernier corollaire.

Théorème 43. *Soit $Y : \Omega \to \mathbb{R}^d$ un vecteur gaussien de loi $\mathcal{N}(m, C)$, alors Y admet une densité par rapport à la mesure de Lebesgue si et seulement si C est définie positive (comme C est toujours positive ssi $det(C) \neq 0$) et alors $P_Y(dx) = f_Y(x)dx$ avec sa densité :*

$$f_Y(x) = \frac{1}{(2\pi)^{d/2}\sqrt{det(C)}} exp(-\frac{1}{2}\langle C^{-1}(x - m), (x - m)\rangle).$$

De plus si C n'est pas définie positive, Y est p.s. à valeur dans $m + Im(C)$.

PREUVE : Si $det(C) \neq 0$, on reprend la notation du théorème précédent, et il suffit de montrer que $Y = m + BX$ a la densité cherchée par changement de variable. En effet, par la forme de la transformée de Fourier de X, les coordonnées de X sont indépendantes et la formule pour la densité découle du calcul des transformées de Fourier une variable lemme 12 et du théorème de caractérisation de l'indépendance comme mesure produit :

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} exp(-\frac{||x||_2^2}{2}).$$

On fait donc le changement de variables $y = m + Bx$ le jacobien est B, $x = B^{-1}(y - m)$ et B^{-1} existe puisque $det(B) = \sqrt{det(C)}$, on obtient donc pour h mesurable positive par transfert puis le changement de variables ci-dessus :

$$\mathbf{E}(h(Y)) = \int_{\mathbb{R}^d} h(m+Bx)\frac{1}{(2\pi)^{d/2}}exp(-\frac{||x||_2^2}{2})dx = \int_{\mathbb{R}^d} h(y)\frac{1}{(2\pi)^{d/2}}exp(-\frac{||B^{-1}(y-m)||_2^2}{2})\frac{1}{det(B)}dy.$$

Ce qui donne par identification la densité une fois que l'on a remarqué que $||B^{-1}(y-m)||_2^2 = \langle B^{-1}(y-m), B^{-1}(y-m)\rangle = \langle C^{-1}(y-m), (y-m)\rangle$ (car $(B^{-1})^T B^{-1} = B^{-2} = C^{-1}$ et $det(B) = \sqrt{det(C)}$.

Comme C est une matrice carré elle est inversible si et seulement si elle est surjective donc il suffit de prouver le dernier résultat pour avoir la contraposée de la réciproque (si C non inversible, vu son support ayant mesure de Lebesgue 0 comme tout e.v. de plus petite dimension, elle ne peut pas avoir de densité.) Mais comme $Im(C) = Im(B)$, le modèle concret donne l'image de valeur p.s. $m + Im(B)$(formellement on prend n'importe quelle fonction h supportée à l'extérieure, par transfert l'intégrale de h est $\mathbf{E}(h(m+BX))$ et $h(m+BX) = 0$ par hypothèse sur le support). ■

On peut décrire explicitement la mesure gaussienne dans tous les cas mais le résultat est moins important donc facultatif.

Corollaire 44. *(Facultatif) Si $X = (X_1, ..., X_d) \sim \mathcal{N}(m, C)$ variable gaussienne avec $m \in \mathbb{R}^d$ et $C \in M_d(\mathbb{R})$ une matrice positive. Soit I l'injection de $\mathbb{R}^m \simeq Im(C) \subset \mathbb{R}^d$ telle que $I(e_i)$ base orthonormale de $Im(C)$ et $D_{ij} = \langle CI(e_i), I(e_j)\rangle$ la matrice induite. On suppose $m \geq 1$. Alors pour tout h mesurable positive*

$$\mathbf{E}(h(X)) = \int_{\mathbb{R}^m} dy_1...dy_m \frac{1}{(2\pi)^{m/2}\sqrt{det(D)}} h(I(y)+m)exp(-\frac{\langle D^{-1}y, y\rangle}{2}).$$

Preuve : En identifiant $Im(C)$ à un $\mathbb{R}^m$ $Y_i = \langle I(e_i), X - m\rangle$ donne en effet un vecteur gaussien centré $(Y_1, ..., Y_m)$ de covariance D, qui est la restriction/projection orthogonale de C à $Im(C)$ est de déterminant non nul (car l'orthogonal de $Im(C)$ est $Ker(C^T) = Ker(C)$, donc D restriction à $Im(C)$ est injective (le noyau est dans l'orthogonal) donc bijective par le théorème de rang). Y nous ramène donc au cas précédent d'une densité. ■

3 Indépendance

Le résultat suivant est important. Au contraire des variables générales, pour laquelle l'indépendance n'est pas caractérisée par l'annulation de la covariance, l'indépendance dans un vecteur gaussien est caractérisée de cette façon. On va voir que c'est une conséquence simple de la forme de la fonction caractéristique.

Théorème 45. *Soit $(X_1, ...X_n, Y_1, ...Y_m) : \Omega \to \mathbb{R}^{m+n}$ un vecteur gaussien. Alors $X = (X_1, ..., X_n)$ (ou $\sigma(X_1, ..., X_n)$) est indépendant de $Y = (Y_1, ..., Y_m)$ (ou $\sigma(Y_1, ..., Y_m)$) si et seulement si*

$$\forall i \in [1, n], \forall j \in [1, m], \quad Cov(X_i, Y_j) = 0.$$

Il est crucial que l'ensemble des vecteurs $(X_1, ...X_n, Y_1, ...Y_m)$ soit un vecteur gaussien, et PAS seulement $(X_1, ..., X_n)$ et $(Y_1, ..., Y_m)$! !

Preuve : L'implication, indépendance entraîne covariance nulle, a déjà été notée $E(X_iY_j) = E(X_i)E(Y_j)$ si X_i, Y_j indépendantes. Pour la réciproque, comme $(X_1, ...X_n, Y_1, ...Y_m)$ gaussien on peut utiliser la formule de la fonction caractéristique et l'annulation de la partie des bloc reliant les X et les Y :

$$\begin{aligned}\Phi_{(X_1,...X_n,Y_1,...Y_m)}(t_1, ..., t_n, s_1, ...s_m) &= exp(i\langle \mathbf{E}(X), t\rangle + i\langle \mathbf{E}(Y), s\rangle - \frac{1}{2}\langle (t,s), \mathbf{Cov}(X,Y)(t,s)\rangle) \\ &= exp(i\langle \mathbf{E}(X), t\rangle + i\langle \mathbf{E}(Y), s\rangle - \frac{1}{2}\langle t, \mathbf{Cov}(X)t\rangle - \frac{1}{2}\langle s, \mathbf{Cov}(Y)s\rangle) \\ &= \Phi_X(t)\Phi_Y(s)\end{aligned}$$

Par la caractérisation 16, on obtient l'indépendance souhaitée. ■

Exemple 29. On voit que Y_1, Y_2 de l'ex 26 sont indépendantes par le calcul de leur covariance.

4 Mesures, espaces et processus gaussiens

Une famille de variables aléatoires $(X_i)_{i\in I}$ est souvent appelé processus. C'est particulièrement le cas quand $I = \mathbb{N}$, ou $I = \mathbb{R}, I = \mathbb{R}_+$ (il y a alors une intuition temporelle liée à l'ordre qui est l'origine de la terminologie processus).

Définition 31. Un processus $X_t, t \in I$ est un **processus gaussien** (centré) si pour tout $t_1, ..., t_n \in I$ $(X_{t_1}, ..., X_{t_n})$ est un vecteur gaussien (centré).

Un sous-espace vectoriel H de $L^2(\Omega, \mathcal{T}, P)$ est un **espace gaussien** (centré), si $(X_1, ..., X_n)$ est un vecteur gaussien (centré) pour tout $X_1, ..., X_n \in H$ (c'est-à-dire si pour tout $\lambda_1, ..., \lambda_n \in \mathbb{R}$, $\sum \lambda_i X_i$ est une variable gaussienne).

Exemple 30. Une suite (X_n) de variables $\mathcal{N}(0,1)$ i.i.d est un processus gaussien.

Exemple 31. Si $X_i, i \in I$est un processus gaussien, $Vect(X_i, i \in I)$ est un espace gaussien. Un espace gaussien H est lui même un processus gaussien (indicé par lui même) $(X)_{X\in H}$.

Théorème 46. *Pour tout espace gaussien H, la fermeture $\overline{H}^{L^2}$ est encore un espace gaussien.*

PROOF : Il suffit de voir que si $Y = (Y_1, ..., Y_m) = lim(X_1^{(n)}, ...X_m^{(n)})$ dans L^2 et que les $X^{(n)} = (X_1^{(n)}, ...X_m^{(n)})$ sont des vecteurs gaussiens, il en est de même de $(Y_1, ..., Y_m)$. En effet, la moyenne $E(X^{(n)}) \to m$ et la matrice de covariance $C_{ij}^{(n)} = Cov(X_i^{(n)}, X_j^{(n)}) \to C$ d'après la convergence dans L^2 (et par Cauchy-Schwartz pour le second). De plus la convergence L^2 implique la convergence en loi, donc par le thm de Paul Lévy, les fonctions caractéristiques convergent

$$\Phi_Y(t) = \lim_{n\to\infty} \Phi_{X^{(n)}}(t) = \lim_{n\to\infty} exp(i\langle t, E(X^{(n)})\rangle - \frac{\langle C^{(n)}t, t\rangle}{2}) = exp(i\langle t, m\rangle - \frac{\langle Ct, t\rangle}{2}),$$

donc par la caractérisation en terme de transformée de Fourier $Y \sim \mathcal{N}(m, C)$.

■

Définition 32. On appelle covariance d'un processus $X_t, t \in I$, la fonction sur I^2, $C(t,s) = Cov(X_t, X_s)$.

Théorème 47 (Kolmogorov)**.** *(facultatif) Pour toute matrice de covariance C sur I, c'est à dire fonction sur I^2 telle que pour tout $t_1, ..., t_n \in I$ $(C(t_i, t_j))_{i,j\in[1,n]}$ est une matrice symétrique positive. Il existe un unique processus gaussien centré $X_t, t \in I$ tel que $C(t,s) = Cov(X_t, X_s)$.*

Enfin, on parle parfois de mesure gaussienne pour généraliser à un espace de Banach séparable le cas des mesures associés aux vecteurs gaussiens de $\mathbb{R}^n$. On verra une mesure gaussienne et un processus gaussien particulier au chapitre 5, le mouvement brownien.

Définition 33. Une mesure γ sur la tribu des boréliens d'un espace de Banach E est appelée **mesure gaussienne** si pour toute forme linéaire continue $L \in E^*$, sa mesure image $L_*\gamma(B) = \gamma(L^{-1}(B))$ sur $\mathbb{C} = \mathbb{R}^2$ est une mesure gaussienne.

Chapitre 4

Martingales

1 D'autres rappels sur les espaces L^p

Tous les espaces L^p sont implicitement des espaces de (classes d'équivalences p.s. de) fonctions à valeur COMPLEXE !

1.1 L'espace de Hilbert $L^2(\Omega)$

La formule

$$\langle f, g\rangle = \mathbf{E}[\overline{f}g]$$

définit un produit scalaire tel que $||f||_2^2 = \langle f, f\rangle$. On peut donc appliquer à $L^2(\Omega)$ le résultat vu en MASS 51 (Topologie et convexité) sur les espaces de Hilbert, dont le théorème de projection sur un convexe fermé. On rappelle ici l'énoncé du cas d'un sous-espace vectoriel fermé. Ce sera la base que l'on généralisera aux autres espaces L^p avec l'espérance conditionnelle dans la section suivante.

Théorème 48. *Soit $K \subset L^2(\Omega, \mathcal{T}, P)$ un sous espace vectoriel fermé. Pour tout $f \in L^2(\Omega, \mathcal{T}, P)$ il existe un unique $u = P_K(f)$ tel que*

$$||f - u||_2 = \inf_{v\in K} ||f - g||_2.$$

De plus c'est l'unique vecteur $u \in K$ tel que

$$\forall v \in K, \quad \mathbf{E}(\overline{v}(f - u)) = 0$$

c'est à dire tel que $f - u$ est orthogonal à K ou encore

$$\forall v \in K, \qquad \mathbf{E}(\overline{v}f) = \mathbf{E}(\overline{v}P_K(f)). \qquad (PC)$$

*P_K est une application linéaire bornée appelée **projection orthogonale sur** K.*

On en déduit maintenant le calcul de l'espace vectoriel des formes linéaires continues sur L^2 (voir sous-section 1.2 pour des rappels).

Théorème 49. *Soit ϕ une forme linéaire continue sur $L^2(\Omega, \mathcal{T}, P)$ alors il existe un unique $f \in L^2(\Omega, \mathcal{T}, P)$ tel que*

$$\forall v \in L^2(\Omega, \mathcal{T}, P), \phi(v) = \mathbf{E}(fv).$$

De plus, on a l'expression duale pour la norme :

$$||f||_2 = \sup_{||v||_2\leq 1} |\mathbf{E}(fv)|.$$

Preuve : Soit $K = \phi^{-1}(0)$ le noyau de ϕ. Si $K = L^2(\Omega, \mathcal{T}, P)$ alors $f = 0$ convient. On suppose donc $K \neq L^2(\Omega, \mathcal{T}, P)$. Soit donc $g_0 \in L^2(\Omega, \mathcal{T}, P)$ et $g = \frac{g_0 - P_K(g_0)}{p}||g_0 - P_K(g_0)||_2$ un vecteur de norme 1 et orthogonal à H. Comme ϕ est une forme linéaire, on s'attend que H et g engendrent L^2, sorte de généralisation du théorème du rang. En effet, soit $v \in L^2(\Omega, \mathcal{T}, P)$, $w = v - \frac{\phi(v)}{\phi(g)}g$ vérifie $\phi(w) = \phi(v) - \frac{\phi(v)}{\phi(g)}\phi(g) = 0$ donc $w \in H$ et $v = \lambda g + w$ avec $\lambda = \frac{\phi(v)}{\phi(g)}$.

On montre donc que $f = \phi(g)\overline{g}$ convient en montrant l'égalité sur v quelconque précédent :

$$\mathbf{E}(fv) = \phi(g)\langle g, v\rangle = \phi(g)\langle g, \lambda g + w\rangle = \phi(g)\lambda||g||_2^2 = \phi(g)\lambda = \phi(v).$$

L'égalité des normes vient de Hölder, cf proposition 50 ci-dessous. ■

1.2 Dualité dans les espaces L^p [Facultatif]

On rappelle du cours MASS 31 (prop 23 p 12) que les formes linéaires continues sur un espace vectoriel normé coïncident avec les formes linéaires lipschitziennes. Pour E un e.v.n, on note E^* l'espace des formes linéaires continues avec pour norme la constante de lipschitzianité :

$$||\phi||_{E^*} = \sup\{|\phi(x)|; ||x||_E \leq 1\}.$$

On appelle E^* l'espace de Banach dual de E (il est facile de voir que cet espace vectoriel normé est toujours complet).

La conséquence facile de Hölder que l'on utilisera fréquemment est la proposition suivante :

Proposition 50. *Soit $p \in [1, \infty]$, q tel que $1/p + 1/q = 1$ le coefficient conjugué, alors $L^q(\Omega, P) \subset (L^p(\Omega, P))^*$ en identifiant $g \in L^q(\Omega, P)$l'application linéaire à $f \mapsto \int fgdP$ et en particulier :*

$$||g||_q = \sup\{\left|\int fgdP\right|; ||f||_p \leq 1\}.$$

Preuve : Par Hölder, $fg \in L^1$ donc l'intégrale est définie et

$$|\int fgdP| \leq ||fg||_1 \leq ||f||_p||g||_q$$

d'où $||g||_q$ est plus grand que le sup de l'énoncé. Mais, pour $p \in]1, \infty[$, si on prend $f = \overline{g}|g|^{q-2}/||g||_q^{q-1}$ on a $|f|^p = |g|^{p(q-1)}/||g||_q^{p(q-1)} = |g|^q/||g||_q^q$ car $p(q-1) = qp(1-1/q) = q$, donc $f \in L^p$ et $||f||_p^p = E(|f|^p) = ||g||_q^q/||g||_q^q = 1$ d'où le sup est supérieur à $\left|\int fgdP\right| = \int |g|^q dP/||g||_q^{q-1} = ||g||_q$. On déduit donc l'égalité énoncée. En conséquence on déduit que l'identification est bien injective (si la forme linéaire est nulle, le sup aussi, donc $||g||_q = 0$ donc $g = 0$ p.s. c'est à dire dans L^q.)

Si $p = 1, q = \infty$, g est dans tout L^r et en prenant le $f = \overline{g}|g|^{r-2}/||g||_r^{r-1}$ précédant qui est de norme L^1 inférieure à 1, on voit que le sup est inférieur à $||g||_r$ donc à la limite quand $r \to \infty$ à savoir $||g||_\infty$.

Si $p = \infty, q = 1$, il suffit de prendre $f = 1_{g\neq 0}\frac{\overline{g}}{|g|}$ de sorte que $fg = |f|$ et la norme $||f||_\infty \leq 1$. ■

On a même le théorème suivant (on notera que $p < \infty$ contrairement au cas de la proposition précédente) :

Théorème 51. *Dans l'identification précédente, $L^q(\Omega, P) = (L^p(\Omega, P))^*$, pour $p \in [1, \infty[$, q tel que $1/p + 1/q = 1$.*

Preuve : (FACULTATIF) Une première preuve classique utilise le théorème de Radon-Nikodym qui est au programme du cours de Th de la mesure (cf. par exemple le cours de Probabilités de Philippe Barbe et Michel Ledoux). Il existe aussi une preuve par l'uniforme convexité dans le livre d'Haim Brezis d'analyse fonctionnelle pour $p \neq 1$ et avec une preuve directe n'utilisant que le cas $p = 2$ (cas Hilbert simple) pour le cas $p = 1$. On donne ici une méthode d'analyse fonctionnelle plus abstraite.

Le **cas** $p = 2$ **a été traité dans la partie espace de Hilbert.**

On donne maintenant une **preuve du cas** $p = 1$ Soit $\phi \in L^1(\Omega)^*$, il faut montrer qu'elle vient d'un élément de $L^\infty(\Omega)$ comme dans la précédente proposition. D'abord on définit T application linéaire continue sur $L^2(\Omega)$ (en fait à valeur dans son dual identifié à lui même) par :

$$\langle Tx, y\rangle = \phi(\overline{x}y)$$

vu que $\overline{x}y \in L^1(\Omega)$ par Hölder et on a

$$||T|| := \sup\{||Tx||_2, ||x||_2 \leq 1\} = \sup\{|\langle Tx, y\rangle|, ||x||_2 \leq 1, ||y||_2 \leq 1\} \leq ||\phi||_{L^1(\Omega)^*}.$$

La première égalité est la définition de la norme des applications linéaires bornées, la deuxième est le résultat de dualité du cas $p = 2$, la troisième utilise Hölder et la définition de la norme du dual. Notons que si $z \in L^\infty(\Omega)$,

$$\langle Tzx, y\rangle = \phi(\overline{zx}y) = \langle Tx, \overline{z}y\rangle = \langle zTx, y\rangle$$

la deuxième relation en utilisant la commutativité des espaces de fonction relation $\overline{zx}y = \overline{xz}y$ et la seconde la définition du produit scalaire $\langle Tx, \overline{z}y\rangle = \int \overline{Tx}\overline{z}ydP$. donc on déduit si m_z est la multiplication par $z \in L^\infty$, $Tm_z = m_zT$. Montrons que $T = m_x$ pour $x \in L^\infty$.

En effet, soit $x = T(1) \in L^2$ l'image de la fonction constante 1. On peut supposer $||T|| \leq 1$.

Pour $g \in L^\infty$ avec $||g||_1 \leq 1$,

$$|\int T(1)gdP| = |\int (|g|^{1/2}T)(1)g|g|^{-1/2}dP| = |\int T(|g|^{1/2})g|g|^{-1/2}dP| \leq |||g|^{1/2}||_2||g|g|^{-1/2}||_2 = ||g||_1 = 1$$

où on a utilisé à la deuxième égalité la commutation avec $m_{|g|^{1/2}}$. On montre maintenant $||T(1)||_{q+1} \leq 1$ par récurrence sur $q \geq 1$, on connait $q = 1$ par hypothèse sur la norme de T. En prenant $g = \overline{T(1)}|T(1)|^{q-1}1_{|T(1)|\leq R}$ de sorte que $||g||_1$ borné par 1 indépendamment de R par hyp. de rec. au rang $q-1$ et on déduit $\int T(1)gdP = \int |T(1)|^{q+1}1_{|f(1)|\leq R}dP \leq 1$ et par convergence monotone $R \to \infty$ on obtient le résultat $\int |T(1)|^{q+1}dP \leq 1$. En prenant la limite $q \to \infty$, on obtient donc $T(1) \in L^\infty$ de norme inférieure à 1. On peut maintenant conclure, pour $g \in L^\infty$,, $T(g) = gT(1) = m_{T(1)}(g)$ et par densité cela s'étend à $g \in L^2$ donc $T = m_{T(1)}$ et

$$\phi(\overline{z}y) = \int \overline{T(1)}\overline{z}ydP$$

comme tout élément $X \in L^1(\Omega)$ se met sous la forme $\overline{z}y$ pour $z = |X|^{1/2}, y = X|X|^{-1/2} \in L^2(\Omega)$ on déduit que ϕ est l'élément du dual associé à $\overline{T(1)}$ par la proposition précédente.

Terminons par une **preuve du cas** $p > 1$ **utilisant les cas** $p = 1, 2$**.** On commence par montrer que $L^p(\Omega)^* \subset L^1(\Omega)$. Si $p \leq 2$ c'est évident par l'inclusion $L^p(\Omega)^* \subset [L^2(\Omega)]^* = L^2(\Omega)$ (la première inclusion est l'application de restriction qui est injective par densité de $L^2 \subset L^p$). Si $p > 2$ pour $x \in L^\infty$, $\phi \in (L^p)^*$,

$$|\phi(x)|^p \leq \int |x|^pdP \leq \int |x|^2||x||_\infty^{p-2}dP \leq ||x||_2^2||x||_\infty^{p-2}.$$

Par l'inégalité d'Young (cas particulier d'Holder utilisé dans sa preuve) $|ab| \leq a^P/P + b^Q/Q$ utilisé avec $1/P + 1/Q = 1, P = p/2, Q = p/(p-2)$, $a = ||x||_2^{1/P}/\epsilon^{1/Q}, b = (\epsilon||x||_\infty)^{1/Q}$, on obtient :

$$|\phi(x)| \leq \epsilon||x||_\infty + \frac{1}{\epsilon^{P/Q}}||x||_2.$$

En incluant $\{(x,x), x \in L^p(\Omega)^*\} \subset L^\infty(\Omega) \times L^2(\Omega)$ avec norme $||(x,y)|| = \epsilon||x||_\infty + \frac{1}{\epsilon^{P/Q}}||y||_2$ on étend par Hahn Banach ϕ à $L^\infty(\Omega) \times L^2(\Omega)$ donnant un élément de $(\phi_1, \phi_2) \in L^\infty(\Omega)^* \times L^2(\Omega)$ avec $||\phi_1|| \leq \epsilon, ||\phi_2|| \leq \frac{1}{\epsilon^{P/Q}}$ comme donc $||\phi - \phi_2||_{L^\infty(\Omega)^*} \leq \epsilon$ et $\phi_2 \in L^1(\Omega)$. Or par le cas $p = 1$, $(L^1(\Omega))^{**} = L^\infty(\Omega)^*$ et il contient par un résultat général $L^1(\Omega)$ comme espace fermé isométriquement (comme tout espace de Banach est inclus isométriquement comme espace fermé dans son bidual). Comme le résultat précédent indique $\phi \in \overline{L^2(\Omega)}^{(L^1(\Omega))^{**}}$, on déduit $\phi \in L^1(\Omega)$ comme voulu.

Soit donc y l'image dans L^1 de ϕ (on revient au cas général $p \in]1, \infty[$) soit donc pour $x \in L^\infty$, $\phi(x) = \int yxdP$ Or $||\phi||_{(L^p)^*} = \sup\{|\phi(x)|, ||x||_p \leq 1, x \in L^\infty\} = \sup\{|\int yxdP|, ||x||_p \leq 1, x \in L^\infty\}$

Or comme $y1_{|y| \leq R} \subset L^q$ on peut utiliser la proposition précédente pour dire (en utilisant aussi la densité de L^∞ dans L^p :

$$||y1_{|y|\leq R}||_q = \sup\{|\int y1_{|y|\leq R}xdP|, ||x||_p \leq 1, x \in L^\infty\} \leq ||\phi||_{(L^p)^*}$$

car $||1_{|y|\leq R}x||_p \leq ||x||_p \leq 1$. En prenant la limite $R \to \infty$ par convergence monotone, on déduit $y \in L^q$ comme on voulait.

■

1.3 Compacité faible-* dans L^p [Facultatif]

Cette partie est facultative mais sera essentielle dans les preuves des résultats de convergences des martingales.

Contrairement à la dimension finie, la boule unité de l'espace vectoriel normé $L^p(\Omega, \mathcal{T}, P)$ n'est pas compact ce qui rendrait beaucoup de résultats plus compliqués si on ne pouvait utiliser une autre métrique que celle de la norme pour remédier au problème.

Pour cela on suppose $\mathcal{T} = \sigma(A_n, n \in \mathbb{N})$, on dit que $\mathcal{T}$ est **dénombrablement engendrée**. C'est par exemple le cas si $\mathcal{T} = \sigma(X_n, n \in \mathbb{N})$ pour une suite de variable aléatoire car cette tribu est engendrée par $X_n^{-1}(]-\infty, q])q \in \mathbb{Q}, n \in \mathbb{N}$ qui est un ensemble dénombrable.

On fixe $1 < p \leq \infty$ pour que $L^p(\Omega, \mathcal{T}, P)$ soit un espace dual. On pose alors $A = \{\sum_{i=1}^n \lambda_i 1_{A_i}, \lambda_i \in \mathbb{Q}, n \in \mathbb{N}\}$ c'est un ensemble dénombrable de fonctions étagées qui est facilement dense dans $L^q(\Omega, \mathcal{T}, P)$, $q \in [1, \infty[$. On énumére $A = \{f_n, n \in \mathbb{N}\}$.

On définit donc la distance sur $L^p(\Omega, \mathcal{T}, P)$:

$$d(h, g) = \sum_{n=0}^{\infty} \frac{1}{2^n} \min(1, |\mathbf{E}[(h-g)f_n]|).$$

Lemme 52. *On fixe $p \in]1, \infty]$ et q l'exposant conjugué tel que $1/p + 1/q = 1$. d est une distance sur $L^p(\Omega, \mathcal{T}, P)$ et pour une suite $g_n \in L^p(\Omega, \mathcal{T}, P), g \in L^p(\Omega, \mathcal{T}, P)$ alors on a l'équivalence entre*

1. *g_n est bornée dans $L^p(\Omega, \mathcal{T}, P)$ et $d(g_n, g) \to 0$*
2. *$\forall f \in L^q(\Omega, \mathcal{T}, P), \mathbf{E}[g_n f] \to \mathbf{E}[gf]$.*

En particulier, cette convergence ne dépend pas du choix de A et on dit que g_n ***converge vers*** *g* ***pour la topologie faible-* de*** *$L^p(\Omega, \mathcal{T}, P)$, notée $g_n \overset{*-L^p}{\to}_{n\to\infty} g$ ou plus simplement $g_n \overset{*-L^p}{\to}_{n\to\infty} g$.*

Preuve : D'abord, d est bien définie car $\frac{1}{2^n}\min(1, |\mathbf{E}[(h-g)f_n]|) \leq \frac{1}{2^n}$ donne la convergence. Si $g, h, l \in L^p$, $|\mathbf{E}[(h-g)f_n]| \leq |\mathbf{E}[(h-l)f_n]| + |\mathbf{E}[(l-g)f_n]|$ donc $\min(1, |\mathbf{E}[(h-g)f_n]|) \leq \min(1, |\mathbf{E}[(h-l)f_n]|) + \min(1, |\mathbf{E}[(l-g)f_n]|)$ donc en sommant on obtient l'inégalité triangulaire $d(g,h) \leq d(g,l) + d(l,h)$. La symétrie est évidente et si $d(f,g) = 0$ pour tout n $\mathbf{E}[(h-g)f_n] = 0$ donc par densité pour tout $f \in L^q, \mathbf{E}[(h-g)f] = 0$ donc $h = g$ dans $(L^q)^* = L^p$ par le théorème 51, mais on peut même simplement utiliser la conséquence de Holder, la proposition 50 :

$$||(h-g)||_p = \sup\{|\mathbf{E}[f(h-g)]| \,;\, ||f||_q \leq 1\} = 0.$$

Pour l'équivalence des convergences, soit g_n bornée avec $d(g_n, g) \to 0$. Fixons $C = \sup\{||g_n||_p, n \in \mathbb{N}, ||g||_p\}$

Soit $f \in L^q(\Omega, \mathcal{T}, P)$ et $\epsilon \in]0,1|$ et par densité soit f_m tq $||f_m - f||_q \leq \epsilon/3C$.

On prend N tel que, pour $n \geq N$, $d(g_n, g) \leq \epsilon/32^m$ donc $\frac{1}{2^m}\min(1, |\mathbf{E}[(g_n - g)f_m]|) \leq d(g_n, g) \leq \epsilon/32^m$ donc $\min(1, |\mathbf{E}[(g_n-g)f_m]|) \leq \epsilon/3$ et donc comme le min n'est pas 1 : $|\mathbf{E}[(g_n-g)f_m]| \leq \epsilon/3$,

Il suffit alors d'appliquer l'inégalité triangulaire et Hölder :

$$|\mathbf{E}[(g_n-g)f]| \leq |\mathbf{E}[(g_n-g)(f-f_m)]| + |\mathbf{E}[(g_n-g)f_m]| \leq (||g_n||_p + ||g||_p)||f-f_m||_q + \epsilon/3 \leq 2C\epsilon/3C + \epsilon/3 = \epsilon.$$

Donc comme ϵ arbitraire, $|\mathbf{E}[(g_n - g)f] \to 0$ et on a le second énoncé.

Réciproquement supposons $\mathbf{E}[(g_n - g)f] \to 0$ pour tout $f \in L^q(\Omega, \mathcal{T}, P)$ en particulier pour $f = f_m$ et donc si $\epsilon \in]0,1[$ on peut fixer N tel que $\sum_{n=N+1}^{\infty} \frac{1}{2^n} \leq \epsilon/2$ ce qui ramène la borne sur $d(g_n, g)$ a une borne sur un nombre fini de termes ce qui est possible, donc on prend n_0 tel que pour $n \geq n_0$ et $m \leq N$, $|\mathbf{E}[(g_n-g)f]| \leq \epsilon/2(N+1)$, de sorte que

$$d(g_n, g) \leq \sum_{m=0}^{N} \frac{1}{2^m}|\mathbf{E}[(g_n-g)f]| + \epsilon/2 \leq \sum_{m=0}^{N} \epsilon/2(N+1) + \epsilon/2 = \epsilon.$$

On a donc $d(g_n, g) \to 0$ et il reste à voir que g_n est bornée dans L^p.

Il faut utiliser une résultat non-évident conséquence du Théorème de Banach-Steinhaus qui dit que si $B \subset (L^q)^*$ est borné si et seulement si il est faiblement borné, c'est-à-dire, pour tout $f \in L^q$, $\{b \in B, b(f)\}$ est borné. Ici, pour tout n , $f \in L^q$, $\mathbf{E}(g_n f)$ converge donc est bornée, donc $\{g_n\}$ est faiblement bornée dans $(L^q)^* = L^p$ donc bornée dans L^p.

■

Le résultat fondamental est le suivant :

Théorème 53. *Soit $p \in]1, \infty]$ et $\mathcal{T}$ une tribu dénombrablement engendrée comme ci-dessus, la boule unité de $L^p(\Omega, \mathcal{T}, P)$ est compacte pour la distance d.*

Preuve : On montre que tout suite g_n bornée (disons par 1) dans $L^p(\Omega, \mathcal{T}, P)$ admet une sous-suite convergente. Par Hölder, $(\mathbf{E}(g_n f_m))_n$ est bornée pour tout m donc par compacité dans $\mathbb{C}$ admet $(\mathbf{E}(g_n f_0))_n$ admet une sous-suite $(\mathbf{E}(g_{\varphi_1(n)} f_0))_n$ convergente. Par récurrence on construit $(\varphi_m(n))_n$ extraite de $(\varphi_{m-1}(n))_n$ telle que $(\mathbf{E}(g_{\varphi_m(n)} f_i))_n$ est convergente pour $i \leq m-1$. (Par axiome des choix dépendant, un point subtil) On considère l'extraction diagonale $g_{\varphi_n(n)}$ comme $(\mathbf{E}(g_{\varphi_n(n)} f_i))_n$ est extraite de $(\mathbf{E}(g_{\varphi_m(n)} f_i))_n$ quand restreinte à n assez grand, elle converge. Donc pour tout i, $\mathbf{E}(g_{\varphi_n(n)} f_i))_n \to \lambda(f_i)$.

Comme $f \to \mathbf{E}(g_{\varphi_n(n)} f)$ est une application linéaire sur A, il en est de même par limite de λ. De plus $\sup\{|\mathbf{E}(g_{\varphi_n(n)} f)|, f \in A ||f||_q \leq 1\} \leq ||g_{\varphi_n(n)}||_p \leq 1$, donc par limite $\sup\{|\lambda(f)|, f \in A||f||_q \leq 1\} \leq 1$, donc λ s'étend en un élément du dual $(L^q)^* = L^p$ donc il existe $g \in L^p$ telle que $\lambda(f) = \mathbf{E}(gf), f \in A$ et donc on a vu au lemme précédent que cela disait $g_{\varphi_n(n)} \overset{*-L^p}{\to}_{n\to\infty} g$ d'où la suite extraite convergente cherchée. ■

2 Espérance conditionnelle

En L2, vous avez vu la notion de probabilité conditionnelle $P(A|B) = P(A \cap B)/P(B)$ de deux événements. Cela donne une probabilité $P(.|B)$ il est donc facile d'intégrer par rapport à cette mesure pour obtenir un nombre. On veut maintenant pouvoir considérer à la place d'un événement supposé connu, avoir une tribu d'événements. Ce sera l'objet de base pour pouvoir formaliser un processus avec notre connaissance qui augmente avec le temps (cf. exemple $\mathcal{F}_n = \sigma(X_1, ..., X_n)$ déjà rencontré pour la loi du $0 - 1$.) Le prochain résultat veut donc définir une espérance "sachant une tribu" qui sera une v.a. mesurable par rapport à cette tribu et devra avoir les propriétés de l'espérance. On va s'appuyer sur la projection orthogonale dans L^2 de la section précédente et l'étendre au cas des variables intégrables.

Théorème 54. *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et $\mathcal{A} \subset \mathcal{T}$ une sous-tribu. Soit $X \in L^1(\Omega, \mathcal{T}, P)$ une v.a. réelle, il existe une unique (p.s.) v.a. réelle $\mathcal{A}$ mesurable notée $\mathbf{E}(X|\mathcal{A}) \in L^1(\Omega, \mathcal{A}, P) \subset L^1(\Omega, \mathcal{T}, P)$ vérifiant la propriété caractéristique :*

$$\forall f \in L^\infty(\Omega, \mathcal{A}, P), \mathbf{E}(fX) = \mathbf{E}(f\mathbf{E}(X|\mathcal{A})) \qquad (PC).$$

Remarque 8. (PC) est équivalente à :

$$\forall A \in \mathcal{A}, \mathbf{E}(1_A\ X) = \mathbf{E}(1_A\ \mathbf{E}(X|\mathcal{A})) \qquad (PC')$$

comme on va voir dans la preuve. (PC') suffit à voir l'unicité et (PC) implique (PC') car $1_A \in L^\infty(\Omega, \mathcal{A}, P)$. L'existence de la variable avec la propriété la plus forte implique donc que la v.a. vérifiant (PC') qui est unique doit aussi vérifier (PC) (sinon on raisonne comme pour la construction de l'intégrale pour aller au delà des fonctions étagées.

Preuve :

Unicité Soit g, h $\mathcal{A}$-mesurables vérifiant (PC'). On pose pour $\epsilon > 0$, $A = A_\epsilon = \{\omega, (g\omega) - h(\omega)) \geq \epsilon\} = (g - h)^{-1}([\epsilon, \infty[)$ qui est donc $\mathcal{A}$-mesurable comme $h - g$, donc par (PC') $E(1_A h) = E(1_A f) = E(1_A g)$ soit en utilisant $1_A(g - h) \geq \epsilon 1_A$:

$$0 = E(1_A(g - h)) \geq \epsilon P(A)$$

donc $P(A) = 0$ pour tout ϵ donc par union croissante

$$P(g > h) = \lim_{\epsilon \to 0} P(A_\epsilon) = 0.$$

Par symétrie $P(h > g) = 0$ et donc $g = hp.s.$

Existence Si $X \in L^2(\Omega, \mathcal{T}, P)$, on pose $\mathbf{E}(X|\mathcal{A}) = P_{L^2(\Omega,\mathcal{A},P)}(X)$ la projection sur le sous-espace fermé $L^2(\Omega, \mathcal{A}, P)$. Par définition elle vérifie pour (PC) avec $f \in L^2(\Omega, \mathcal{A}, P)$ donc a fortiori avec $f \in L^\infty(\Omega, \mathcal{A}, P)$.

Montrons que pour $X \in L^2(\Omega, \mathcal{T}, P)$,

$$|\mathbf{E}(X|\mathcal{A})| \leq \mathbf{E}(|X||\mathcal{A}), p.s.$$

Vu que $|\mathbf{E}(X|\mathcal{A})| = \max \mathbf{E}(X|\mathcal{A}), \mathbf{E}(-X|\mathcal{A}))$, il suffit de voir et par linéarité il suffit de voir $\mathbf{E}(|X| - X|\mathcal{A}) \geq 0$ p.s. Mais si $A = \{\omega : \mathbf{E}(|X| - X|\mathcal{A}) < -\epsilon\}$, $\epsilon > 0$ qui est $\mathcal{A}$ mesurable, $\mathbf{E}(1_A\ (|X| - X)) = \mathbf{E}(1_A\ \mathbf{E}(|X| - X|\mathcal{A}))$ la première espérance est positive comme espérance de fonctions positives la seconde est négative par définition de A, donc elle est nulle. Or $\mathbf{E}(1_A\ \mathbf{E}(|X| - X|\mathcal{A})) \leq -\epsilon P(A)$

donc $P(A) = 0$ En prenant $\epsilon = 1/n$ aussi par union dénombrable $P(\mathbf{E}(|X| - X|\mathcal{A}) < 0) = 0$ ce qu'il fallait démontrer.

En intégrant l'inégalité que l'on vient de montrer, on obtient :

$$||\mathbf{E}(X|\mathcal{A})||_1 \leq \mathbf{E}(\mathbf{E}(|X||\mathcal{A})) = \mathbf{E}(|X|) = ||X||_1$$

par la propriété caractéristique. On étend donc $\mathbf{E}(.|\mathcal{A}) : L^1(\Omega, \mathcal{T}, P) \to L^1(\Omega, \mathcal{A}, P)$ par continuité et par densité de $L^2(\Omega, \mathcal{T}, P) \subset L^1(\Omega, \mathcal{T}, P)$. (On verra plus loin une construction dans le cas positif qui évite cette prolongation par continuité.). (PC) est étendu par densité vu que $L^\infty \subset (L^1)^*$ (et on a vu qu'on avait même égalité). ■

Exemple 32. Si $\mathcal{A} = \{\Omega, \emptyset\}$ est la tribu triviale :

$$\mathbf{E}(X|\mathcal{A}) = \mathbf{E}(X).$$

Plus généralement on a le même résultat si $\sigma(X)$ et $\mathcal{A}$ sont indépendantes (quel que soit $\mathcal{A}$).

Exemple 33. Si A est un événement, $\mathcal{A} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$ tq $P(A) \neq 0, P(A^c) \neq 0$ et si $B \in \mathcal{T}$,

$$\mathbf{E}(1_B|\mathcal{A}) = P(B|A)1_A + P(B|A^c)1_{A^c}.$$

Plus généralement, on a :

$$\mathbf{E}(X|\mathcal{A}) = \frac{\mathbf{E}(X1_A)}{P(A)}1_A + \frac{\mathbf{E}(X1_{A^c})}{P(A^c)}1_{A^c}.$$

On commence par donner les propriétés fondamentales similaires à celles de l'intégrale et on donne ensuite les exemples fondamentaux (cas discrets, continus et gaussiens).

Dans les propriétés suivantes, on sous-entend les égalités/inégalités *p.s.* puisque les égalités sont entre objets de L^1, définis comme classes d'équivalence p.s.

Théorème 55. *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, $\mathcal{A} \subset \mathcal{T}$ une sous-tribu alors et $X, Y \in L^1(\Omega, \mathcal{T}, P)$*

1. *(Linéarité)* $\mathbf{E}(aX + bY + c|\mathcal{A}) = a\mathbf{E}(X|\mathcal{A}) + b\mathbf{E}(Y|\mathcal{A}) + c$.
2. *(monotonie)* $X \leq Y \Rightarrow \mathbf{E}[X|\mathcal{A}] \leq \mathbf{E}[Y|\mathcal{A}]$ *. En particulier,* $X \geq 0 \Rightarrow \mathbf{E}[X|\mathcal{A}] \geq 0$.
3. *(Jensen) Si* $\phi : I\!R \to I\!R$ *est convexe* X *v.a. réelle, et* $\phi(X) \in L^1(\Omega, \mathcal{T}, P)$ *alors*

$$\phi(\mathbf{E}(X|\mathcal{A})) \leq \mathbf{E}(\phi(X)|\mathcal{A}).$$

4. *(Cas L^p) Si* $p \in [1, \infty]$, *si* $X \in L^p(\Omega, \mathcal{T}, P)$, *alors* $\mathbf{E}(X|\mathcal{A}) \in L^p(\Omega, \mathcal{A}, P)$ *et*

$$||\mathbf{E}(X|\mathcal{A})||_p \leq ||X||_p.$$

5. *(Cas L^2) Si* $X \in L^2$, $E(X|A) = P_{L^2(\Omega,\mathcal{A},P)}(X)$.
6. *(Cas positif) Si* $Z \geq 0$,

$$\mathbf{E}(Z|\mathcal{A}) = \lim_{n\to\infty} \mathbf{E}(\min(Z, n)|\mathcal{A})$$

limite croissante à valeur $[0, \infty]$ *est l'unique v.a. positive* $\mathcal{A}$ *mesurable vérifiant (PC'), ou par*

$$\forall f \geq 0, \sigma(f) \subset \mathcal{A}, \mathbf{E}(fX) = \mathbf{E}(f\mathbf{E}(X|\mathcal{A})) \qquad (PC'').$$

7. *(Convergence monotone TCM) Si* $Z_n \geq 0$, Z_n *suite de v.a.r croissante et tend simplement vers* Z *alors* $\mathbf{E}(Z_n|\mathcal{A}) \to \mathbf{E}(Z|\mathcal{A})$*p.s. en croissant.*

8. *(Lemme de Fatou) Si* $Z_n \geq 0$ $\mathbf{E}(\liminf Z_n|\mathcal{A}) \leq \liminf \mathbf{E}(Z_n|\mathcal{A})$ *p.s.*
9. *(Convergence dominée TCD) Si* $|Z_n| \leq Y$, $\mathbf{E}[Y] < \infty$ *et* $Z_n \to Z$*p.s. alors*

$$\mathbf{E}(Z_n|\mathcal{A}) \to \mathbf{E}(Z|\mathcal{A}) \text{ p.s et dans } L^1.$$

10. *(Modularité) Si* Y *est* $\mathcal{A}$ *mesurable et soit* $Y, X \geq 0$ *soit* $X, XY \in L^1$ *alors* $E(XY|\mathcal{A}) = YE(X|A)$.
11. *(Conditionnement successif) Si* $\mathcal{B} \subset \mathcal{A}$, $\mathbf{E}(\mathbf{E}(X|\mathcal{A})|\mathcal{B}) = \mathbf{E}(X|\mathcal{B})$.
12. *(Indépendance) Si* $X, T : \Omega \to I\!R^m$ *indépendantes de* $Z : \Omega \to I\!R^d$ *alors* $\mathbf{E}(X|\sigma(Z,T)) = \mathbf{E}(X|T)$.

PREUVE : 2. et 5. on été vérifiées durant la construction. Toutes les égalités vont venir de (PC).

- Pour 1. soit $f \in L^\infty(\Omega, \mathcal{A}, P)$, par linéarité de l'intégrale et application de (PC) à chaque membre !

$$\mathbf{E}(f(a\mathbf{E}(X|\mathcal{A}) + b\mathbf{E}(Y|\mathcal{A}) + c)) = \mathbf{E}(fa\mathbf{E}(X|\mathcal{A})) + \mathbf{E}(fb\mathbf{E}(Y|\mathcal{A})) + \mathbf{E}(fc) = \mathbf{E}(faX + fbY + fc)$$

ce qui donne le résultat par (PC).

-De même pour 11. si $f \in L^\infty(\Omega, \mathcal{B}, P) \subset L^\infty(\Omega, \mathcal{A}, P)$, deux applications successives de (PC) pour les deux espérances conditionnelles donnent :

$$\mathbf{E}\left[f\mathbf{E}(\mathbf{E}(X|\mathcal{A})|\mathcal{B})\right] = \mathbf{E}\left[f\mathbf{E}(X|\mathcal{A})|\right] = \mathbf{E}\left[fX\right]$$

ce qui donne l'égalité par (PC).

-Pour montrer 3., on utilise que le graphe de ϕ convexe est l'intersection des demi-plans de pentes rationnelles dessous lui. Si $E_\phi = \{(a,b) \in \mathbb{Q}^2 : \forall x \in \mathbb{R} \phi(x) \geq ax + b\}$, on a donc

$$\phi(x) = \sup_{(a,b)\in E_\phi} (ax + b).$$

Par 2. $\mathbf{E}(\phi(X)|\mathcal{A}) \geq \mathbf{E}(aX + b|\mathcal{A})$ p.s pour tout $(a,b) \in E_\phi$ c'est à dire sur $\Omega_{a,b}$ avec $P(\Omega_{a,b}) = 1$ Donc si $\Omega' = \cap_{a,b\in E_\phi}\Omega_{a,b}$ on a $P(\Omega') = 1$ et

$$\forall \omega \in \Omega', \ \mathbf{E}(\phi(X)|\mathcal{A}) \geq \sup_{(a,b)\in E_\phi} \mathbf{E}(aX + b|\mathcal{A}) = \phi(\mathbf{E}(aX + b|\mathcal{A})),$$

donc l'identité est vrai p.s.

- Pour 4. comme $\phi(x) = |x|^p$, $p \in]1, \infty[$ est convexe, on applique 3. et on prend l'espérance

$$\mathbf{E}(|\mathbf{E}(X|\mathcal{A})|^p) \leq \mathbf{E}(\mathbf{E}(|X|^p|\mathcal{A})) = \mathbf{E}(|X|^p) = ||X||_p^p < \infty$$

l'égalité venant de (PC) avec $f = 1$. Cela donne l'énoncé et l'égalité.

- Pour 6., (PC) usuelle appliquée à $\min(f, n)$ et le TCM 2 fois donne (PC'')

$$\mathbf{E}(fX) = \lim_{n\to\infty} \mathbf{E}(\min(f,n)\min(X,n)) = \lim_{n\to\infty} \mathbf{E}(\min(f,n)\mathbf{E}(\min(X,n)|\mathcal{A})) = \mathbf{E}(f\mathbf{E}(X|\mathcal{A})).$$

L'unicité utilise le même $A_\epsilon = \{g - h \geq \epsilon\}$ que dans le théorème, pour g, h vérifiant (PC') $E(h1_{A_\epsilon}1_{h<n}) = E(g1_{A_\epsilon}1_{h<n}) \geq E(h1_{A_\epsilon}1_{h<n}) + \epsilon P(A_\epsilon \cap \{h < n\})$ donc comme les deux termes sont finis $\epsilon P(A_\epsilon \cap \{h < n\}) = 0$ en faisant tendre $n \to \infty$, $P(A_\epsilon) = 0$ et on conclut comme avant.

- Pour 7. $E(Z_n|\mathcal{A})$ tend p.s. vers Y par monotonie il suffit de voir que Y vérifie (PC") par TCM des intégrales. Pour f positive, (c'est crucial pour garder monotonie), $\mathcal{A}$ mesurable, $\mathbf{E}(fY) = \lim_{n\to\infty}\mathbf{E}(fE(Z_n|\mathcal{A})) = \lim_{n\to\infty}\mathbf{E}(fZ_n) = \mathbf{E}(fZ)$.

-8.9 se démontrent alors comme en théorie de la mesure. $\inf_{k\geq n} Z_k \to \liminf Z_n$ en croissant donc on peut utiliser 7. et il suffit de remarquer que $\mathbf{E}[\inf_{k\geq n} Z_k|\mathcal{A}] \leq \inf_{k\geq n}\mathbf{E}[Z_k|\mathcal{A}]$ en utilisant 2. et en utilisant la dénombrabilité pour avoir le résultat p.s.

Donc

$$\mathbf{E}[\liminf Z_n|\mathcal{A}] = \lim \mathbf{E}[\inf_{k\geq n} Z_k|\mathcal{A}] \leq \lim \inf_{k\geq n}\mathbf{E}[Z_k|\mathcal{A}] = \liminf \mathbf{E}[Z_n|\mathcal{A}].$$

-Pour 9. vu $Z_n - Y, Z_n + Y \geq 0$ on applique Fatou

$$\mathbf{E}(Z - X|\mathcal{A}) = \mathbf{E}(\liminf(Z - X_n)|\mathcal{A}) \leq \mathbf{E}(Z|A) - \limsup \mathbf{E}(X_n|\mathcal{A})$$

$$\mathbf{E}(Z + X|\mathcal{A}) = \mathbf{E}(\liminf(Z + X_n)|\mathcal{A}) \leq \mathbf{E}(Z|A) + \liminf \mathbf{E}(X_n|\mathcal{A})$$

donc en soustrayant le terme en Z, $\mathbf{E}(X|\mathcal{A}) \leq \liminf \mathbf{E}(X_n|\mathcal{A}) \leq \limsup \mathbf{E}(X_n|\mathcal{A}) \leq \mathbf{E}(X|\mathcal{A})$ p.s. ce qui donne l'égalité p.s. et la convergence p.s. la convergence L^1 vient du TCD vu que $\mathbf{E}(Z|A)$ sert de domination.

- Pour 10, le cas positif est évident par (PC") et le cas L^1 s'en déduit en décomposant en parties positives et négatives.

- Pour 12 $\mathbf{E}(X|T)$ est $\sigma(Z,T)$ mesurable, et par le lemme de Doob-Dynkin une fonction $Y \in L^\infty(\Omega, \sigma(Z,T))$ s'écrit $Y = h(Z,T)$ avec h borélienne bornée. Or de même $\mathbf{E}(X|T) = g(T)$ avec $g \in L^1(\mathbb{R}^m, P_T)$ donc par transfert, indépendance, Fubini et (PC) :

$$\begin{aligned}\mathbf{E}(\mathbf{E}(X|T)h(Z,T)) &= \int dP_T(t)\int dP_Z(z)g(t)h(z,t)\\ &= \int dP_Z(z)\mathbf{E}(\mathbf{E}(X|T)h(z,T)) = \int dP_Z(z)\mathbf{E}(Xh(z,T))\\ &= \int dP_{(X,T)}(x,t)\int dP_Z(z)xh(z,t) = \mathbf{E}(Xh(Z,T))\end{aligned}$$

ce qui conclut par (PC).

■

Remarque 9. Pour des v.a. $X_1, ..., X_n$ Y positive ou intégrable, on note

$$\mathbf{E}(Y|X_1, ..., X_n) := \mathbf{E}(Y|\sigma(X_1, ..., X_n)).$$

2.1 Cas discret

Il y a deux façons de le décrire avec les systèmes presque complets d'événements (definition 6) et les variables aléatoires.

Proposition 56. *1. Soit $(A_n)_{n\geq 1}$ un système presque complet d'événements et $\mathcal{A} = \sigma(A_n, n \geq 1)$ alors pour $X \in L^1(\Omega, \mathcal{T}, P)$:*

$$\mathbf{E}(X|\mathcal{A}) = \sum_{n:P(A_n)\neq 0}\frac{\mathbf{E}(X1_{A_n})}{P(A_n)}1_{A_n}.$$

2. Si Y v.a. à valeur dans un ensemble dénombrable E et $X \in L^1(\Omega, \mathcal{T}, P)$, soit $\phi(y) = \frac{E(X1_{Y=y})}{P(Y=y)}$ si $P(Y = y) \neq 0$ et $\phi(y) = 0$ sinon, alors

$$\mathbf{E}(X|Y) = \phi(Y).$$

2.2 Cas Continue

Théorème 57. *Soit $(X_1, ...X_n, Y_1, ...Y_m)$ un vecteur aléatoire de densité p sur $\mathbb{R}^{n+m}$. Soit $h : \mathbb{R}^n \to [0, \infty[$ mesurable positive et g définie par*

$$g(y) = \frac{1}{q(y)} \int_{\mathbb{R}^n} h(x)p(x, y)dx$$

si $q(y) := \int_{\mathbb{R}^n} p(x, y)dx$ la densité marginale est non nulle et $g(y) = 0$ (ou n'importe quelle valeur) si $q(y) = 0$ alors :

$$E[h(X)|Y] = g(Y).$$

Preuve : Il suffit de vérifier la propriété caractéristique (PC") du cas positif en appliquant le transfert. Par le lemme de Doob-Dynkin (proposition 7), prendre F $\sigma(Y)$ mesurable positive revient à prendre $f : \mathbb{R}^n \to \mathbb{R}_+$ borélienne positive et $F = f(Y)$. Par transfert, pour appliquer (PC) on calcule donc :

$$\begin{aligned}
\mathbf{E}(f(Y)g(Y)) &= \int_{\mathbb{R}^m} f(y)g(y)q(y)dy \\
&= \int_{q(y)\neq 0} dyf(y)q(y)\frac{1}{q(y)} \int_{\mathbb{R}^n} h(x)p(x, y)dx \\
&= \int_{\mathbb{R}^m} dyf(y) \int_{\mathbb{R}^n} h(x)p(x, y)dx \\
&= \mathbf{E}(f(Y)h(X)).
\end{aligned}$$

car $\int_{\mathbb{R}^n} h(x)p(x, y)dx = 0$ si $q(y) = 0$ (qui implique $p(x, y) = 0$ pour presque tout x) ce qui a permis de ramener l'intégrale à une intégrale sur $\mathbb{R}^m$ qui est de nouveau la valeur donnée par transfert. Par PC cas positif cela conclut à l'identité.

■

Remarque 10. La donnée de $E(h(X)|Y)$ pour tout h positive mesurable est souvent appelée donnée de la loi conditionnelle de X sachant Y.

2.3 Cas Gaussien

Pour les vecteurs gaussiens, on a deux simplifications, d'abord le calcul de l'espérance conditionnelle est ramené à un problème d'algèbre linéaire :

Théorème 58. *Soit $(X, Y_1, ..., Y_m)$ un vecteur gaussien centré (i.e.$\mathbf{E}(X) = \mathbf{E}(Y_j) = 0$) et $Y = (Y_1, ..., Y_m)$, alors*

$$\mathbf{E}(X|Y) = \sum_{j=1}^{m} \lambda_j Y_j$$

est la projection orthogonale sur $Vect(Y_1, ..., Y_m) \subset L^2(\Omega, \sigma(Y), P)$.

Cette projection est donc sur un espace beaucoup plus petit que celle a priori pour tous les vecteurs L^2 à savoir sur $L^2(\Omega, \sigma(Y), P)$.
Preuve :
Soit λ_j avec $\sum_{j=1}^m \lambda_j Y_j$ la projection orthogonale sur $Vect(Y_1, ..., Y_m)$. Voyons la propriété caractéristique pour la projection orthogonale sur $L^2(\Omega, \sigma(Y), P)$, qu'on sait être $\mathbf{E}(X|Y)$. On veut donc voir que $X - \sum_{j=1}^m \lambda_j Y_j$ est orthogonale à $L^2(\Omega, \sigma(Y), P)$.

Or $Z = X - \sum_{j=1}^m \lambda_j Y_j$ orthogonale à $Vect(Y_1, ..., Y_m)$ donc $Cov(Z, Y_j) = E(ZY_j) = 0$ (car les vecteurs sont centrés). De plus $(Z, Y_1, ..., Y_m)$ est un vecteur gaussien comme application linéaire de $(X, Y_1, ..., Y_m)$ donc par le critère d'indépendance 45, $\sigma(Z)$ et $\sigma(Y_1, ..., Y_m)$ sont indépendants, donc pour tout $f \in L^2(\Omega, \sigma(Y), P)$, $E(Zf) = E(Z)E(f) = 0$. Donc Z est bien orthogonal à tout $L^2(\Omega, \sigma(Y), P)$. Donc $Z = \mathbf{E}(X|Y)$.

$\blacksquare$

Exemple 34. Soit (X_1, X_2, X_3) un vecteur gaussien de loi $\mathcal{N}(0, \Gamma)$ et $\Gamma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix}$. Montrer que $\mathbf{E}(X_2|X_1, X_3) = \frac{X_1+X_3}{2}$.

Le résultat suivant donne en fait la "loi conditionnelle".

Théorème 59. *Si $(X_1, ..., X_n, Y_1, ..., Y_m)$ un vecteur gaussien centré, $Y = (Y_1, ..., Y_m)$. Soit $h : \mathbb{R}^n \to [0, \infty[$ mesurable positive. Soit C la matrice de covariance conditionnelle*

$$C_{ij} = \mathbf{E}\left[(X_i - \mathbf{E}(X_i|Y))(X_j - \mathbf{E}(X_j|Y))\right]$$

et soit $g : \mathbb{R}^n \to [0, \infty[$ définie par

$$g(m) = \mathbf{E}(h(Z(m))),$$

avec $Z(m) \sim \mathcal{N}(m, C)$. Alors

$$\mathbf{E}(h(X)|Y) = g(\mathbf{E}(X|Y)).$$

Exemple 35. Dans l'exemple précédent, $Var(X_2 - \mathbf{E}(X_2|X_1, X_3)) = (-1/2, 1, -1/2)\Gamma(-1/2, 1, -1/2)^T = (-1/2, 1, -1/2)(0, 3/2, 0)^T = 3/2$. Donc

$$E(X_2^4|X_1, X_3) = 27/4 + \left(\frac{X_1 + X_3}{2}\right)^4 + 9\left(\frac{X_1 + X_3}{2}\right)^2.$$

Preuve : On note g_h le g de l'énoncé. Par la caractérisation du cas positif et le lemme de Doob-Dynkin, il suffit de montrer pour $f : \mathcal{T}^m \to [0, \infty[$ mesurable :

$$\mathbf{E}(f(Y)g_h(\mathbf{E}(X|Y))) = \mathbf{E}(f(Y)h(X))$$

Si $\mathbf{E}(f(Y)) = 0$ $f(Y) = 0 p.s.$ donc les deux cotés sont nuls. On suppose le contraire, soit

$$\mu(h) = \frac{\mathbf{E}(f(Y)h(X))}{\mathbf{E}(f(Y))}, \nu(h) = \frac{\mathbf{E}(f(Y)g_h(\mathbf{E}(X|Y)))}{\mathbf{E}(f(Y))}.$$

$\mu(1) = 1$ et $g_1 = 1$ donc $\nu(1) = 1$. $\mu(B) = \mu(1_B), \nu(B) = \nu(1_B)$ définissent des mesures de probabilités sur $\mathcal{B}(\mathbb{R}^n)$ et il est facile de voir que $\mu(h) = \int h d\mu, \nu(h) = \int h d\nu$.

Donc pour voir $\mu(h) = \nu(h)$ il suffit de voir $\mu = \nu$ et par le théorème d'injectivité de la transformée de Fourier il suffit de voir pour tout $t \in \mathbb{R}^n$:

$$\mathbf{E}(f(Y)E(e^{i\langle t, Z(\mathbf{E}(X|Y))\rangle}))) = \mathbf{E}(f(Y)e^{i\langle t, X\rangle}).$$

Soit encore

$$E(e^{i\langle t, X\rangle}|Y) = E(e^{i\langle t, Z(\mathbf{E}(X|Y))\rangle}) = exp(i\langle t, \mathbf{E}(X|Y)\rangle - \frac{\langle Ct, t\rangle}{2})$$

la dernière identité par la transformée de Fourier du vecteur gaussien $Z(m)$. Or $(\langle t, X\rangle, Y_1, ..., Y_m)$ est gaussien, et $\langle t, \mathbf{E}(X|Y)\rangle = \mathbf{E}(\langle t, X\rangle|Y)$ par linéarité de l'espérance conditionnelle donc par la preuve du résultat précédent $\langle t, X\rangle - \langle t, \mathbf{E}(X|Y)\rangle$ est indépendant de $Y_1, ..., Y_m$, donc

$$E(e^{i\langle t,X\rangle}|Y) = E(e^{i\langle t,X\rangle-\langle t,\mathbf{E}(X|Y)\rangle}|Y)exp(i\langle t, \mathbf{E}(X|Y)\rangle = E(e^{i\langle t,X\rangle-\langle t,\mathbf{E}(X|Y)\rangle})exp(i\langle t, \mathbf{E}(X|Y)\rangle,$$

par la modularité (Thm 55.10) et la relation à l'indépendance (ex 60). En utilisant la transformée de Fourier d'une variable gaussienne et $\langle Ct, t\rangle = Var(\langle t, X\rangle - \langle t, \mathbf{E}(X|Y)\rangle)$, on obtient le résultat d'égalité des transformées de Fourier qu'il fallait démontrer. ■

2.4 Cas Indépendant

Proposition 60. *Soient $X : \Omega \to I\!R^n, Y\Omega \to I\!R^m$ des v.a. indépendantes et soit $\Phi : I\!R^{n+m} \to I\!R_+$ mesurable et alors $\varphi(x) = \mathbf{E}[\Phi(x, Y)] \in I\!R_+$, est mesurable et*

$$\mathbf{E}[\Phi(X, Y)|X] = \varphi(X) \ p.s.$$

Preuve : La mesurabilité vient de Fubini-Tonelli. On est dans le cas positif où on utilise (PC") pour l'espérance conditionnelle positive. Soit $Z \geq 0$ $\sigma(X)$- mesurable, par le lemme de Doob-Dynkin 7, $Z = h(X)$ avec $h : R^n \to \mathbb{R}_+$.

Or par transfert et indépendance

$$\begin{aligned}\mathbf{E}[\Phi(X, Y)Z] = \mathbf{E}[\Phi(X, Y)h(X)] &= \int dP_X(x) \int dP_Y(y)\Phi(x, y)h(x) \\ &= \int dP_X(x)\mathbf{E}(\Phi(x, Y))h(x) = \int dP_X(x)\varphi(x)h(x) = \mathbf{E}(\varphi(X)Z)\end{aligned}$$

ce qui établit par (PC") $\mathbf{E}[\Phi(X, Y)|X] = \varphi(X)$. ■

3 Martingales et temps d'arrêt : définitions et exemples

3.1 Martingales

On fixe $(\Omega, \mathcal{T}, P)$ un espace probabilisé.

Définition 34. Une *filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est une suite croissante de sous-tribus de $\mathcal{T}$.

Intuitivement $\mathcal{F}_n$ est la tribu des événements connus à l'instant n. A partir de maintenant **toutes les tribus $\mathcal{F}_n, \mathcal{T}$ sont supposés dénombrablement engendrées**, même si ce n'est pas strictement nécessaire pour les énoncés, seulement pour simplifier les preuves.

Exemple 36. Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variable aléatoire, on considère souvent la filtration engendrée par ce processus, définie par $\mathcal{F}_n = \sigma(X_0, ..., X_n)$.

Réciproquement on a une notion correspondant à un processus connu à l'instant n pour une filtration.

Définition 35. Une suite $(X_n)_{n\in\mathbb{N}}$ est dite *adaptée* à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si pour tout n ,X_n est $\mathcal{F}_n$-mesurable.

Exemple 37. Une suite $(X_n)_{n\in\mathbb{N}}$ est toujours adaptée à la filtration engendrée par le processus $(\mathcal{F}_n = \sigma(X_0, ..., X_n))_{n\in\mathbb{N}}$ et elle est adaptée à $\mathcal{F}_n$ si et seulement si $\sigma(X_0, ..., X_n) \subset \mathcal{F}_n$

On fixe à présent une filtration $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$.

Définition 36. Un processus $(X_n)_{n\in\mathbb{N}}$ adapté pour $\mathcal{F}$ et dans L^1 (c'est-à-dire $\mathbf{E}(|X_n|) < \infty$) est :

1. une *martingale* si
$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n,$$
2. une *sous-martingale* si
$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) \geq X_n,$$
3. une *sur-martingale* si
$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) \leq X_n.$$

Exemple 38 (Martingales fermées). Si $X \in L^1$, alors $X_n = \mathbf{E}(X|\mathcal{F}_n)$ est une martingale. Une martingale de ce type est dite fermée.

Exemple 39 (Marches aléatoires). Si $X_i \in L^1$, i.i.d. et $S_n = X_0 + ... + X_n$ Alors (S_n) est une martingale pour la filtration $\mathcal{F}_n = \sigma(X_0, ..., X_n)$ (resp. une sous-martingale,resp. une sur-martingale) si et seulement si $\mathbf{E}[X_0] = 0$ (resp. $\mathbf{E}[X_0] \geq 0$, resp. $\mathbf{E}[X_0] \leq 0$).

Exemple 40 (Martingales exponentielles). Si X_i i.i.d. avec $e^{X_i} \in L^1$, $S_n = X_0 + ... + X_n$ et

$$Y_n = exp(S_n - nl)$$

avec $e^l = \mathbf{E}(e^{X_1})$. Alors (Y_n) est une martingale pour la filtration $\mathcal{F}_n = \sigma(X_0, ..., X_n)$.

Lemme 61. *Si (X_n) est une martingale alors pour tout $m \geq n$ on a*

$$\mathbf{E}(X_m|\mathcal{F}_n) = X_n.$$

Exemple 41 (Transformée de Martingale/ Intégrale stochastique discrète). Si $H_n \in L^\infty(\Omega, \mathcal{F}_{n-1})$ on dit que (H_n) est un processus prévisible. Alors si X_n est adapté, on pose $(H.X)_0 = 0$ et

$$(H.X)_n := \sum_{k=1}^{n} H_k(X_k - X_{k-1}).$$

Si (X_n) est une martingale, c'est aussi le cas de $(H.X)$ car $(H.X)_n = (H.X)_{n-1} + H_n(X_n - X_{n-1})$ et

$$\mathbf{E}((H.X)_n|\mathcal{F}_{n-1}) = (H.X)_{n-1} + H_n\mathbf{E}((X_n - X_{n-1})|\mathcal{F}_{n-1}) = (H.X)_{n-1}$$

comme H_n est $\mathcal{F}_{n-1}$-mesurable et en utilisant le théorème 55.(10) (modularité). De même si (X_n) est une sous-martingale et $H_n \geq 0$, on déduit du même calcul que $(H.X)$ est une sous-martingale.

3.2 Temps d'arrêt

Définition 37. Soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration. Une variable aléatoire $T : (\Omega, \mathcal{T}) \to \mathbb{N} \cup \{\infty\}$ est un *temps d'arrêt* si pour tout $n \in \mathbb{N}, \{T \leq n\} \in \mathcal{F}_n$.. On note alors la tribu en T

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

Une fonction constante $T = n$ est un temps d'arrêt et alors $\mathcal{F}_T = \mathcal{F}_n$ est une notation consistante.

Exemple 42 (Temps d'atteinte). Soit A un ensemble mesurable dans $\mathbb{R}^d$ et $X_n : \Omega \to \mathbb{R}^d$ un processus adapté, alors

$$T = \inf\{n \in \mathbb{N}, X_n \in A\}$$

est un temps d'arrêt. En effet on a $\{T \leq n\} = \{\exists m \leq n X_m \in A\} \in \mathcal{F}_n$car $\sigma(X_0, ..., X_n) \subset \mathcal{F}_n$.

Proposition 62 (1^{er} Théorème d'arrêt). *Si (X_n) est une martingale et T un temps d'arrêt alors $(X_{n\wedge T})_{n\in\mathbb{N}}$ est une martingale.*

Si de plus $T \in L^\infty(\Omega)$, alors $X_T \in L^1(\Omega)$ et on a égalité $\mathbf{E}[X_0] = \mathbf{E}[X_T]$.

Si (X_n) est seulement une sous-martingale et $S \leq T$ des temps- d'arrêt BORNÉS (c'est-à-dire dans $L^\infty(\Omega)$) . Alors

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

PREUVE : Pour $n \geq 1$ on pose pour $S \leq T$, $H_n = 1_{\{S<n\leq T\}} = 1_{\{S\leq n-1\}} - 1_{\{T\leq n-1\}} \geq 0$ qui est prévisible c'est-à-dire $\mathcal{F}_{n-1}$-mesurable. Par l'exemple 41, si (X_n) est une martingale (resp. une sous-martingale) il en est de même de

$$(H.X)_n = \sum_{k=1}^{n} 1_{\{S<k\leq T\}}(X_k - X_{k-1}) = \sum_{S<k\leq n\wedge T} (X_k - X_{k-1}) = X_{n\wedge T} - X_{n\wedge S}.$$

Donc dans le cas martingale avec $S = 0$ on trouve $(X_{n\wedge T})$ martingale. Si T est borné p.s. par N, alors $X_T = X_0 + (H.X)_N$ et on déduit $X_T \in L^1$ et $\mathbf{E}((H.X)_N) = \mathbf{E}(\mathbf{E}((H.X)_N|\mathcal{F}_0)) = \mathbf{E}((H.X)_0) = 0$ d'où l'égalité $\mathbf{E}[X_0] = \mathbf{E}[X_T]$. Dans le cas sous martingale, on a déjà de même $X_T, X_S \in L^1$ et $(H.X)_N = X_T - X_S$ et par la propriété de sous-martingale

$$\mathbf{E}(X_T - X_S) = \mathbf{E}((H.X)_N) = \mathbf{E}(\mathbf{E}((H.X)_N|\mathcal{F}_0)) \geq \mathbf{E}((H.X)_0) = 0.$$

■

3.3 Exemple en finance : Modèle de Cox, Ross et Rubinstein

Nous référons au livre *Cours de Calcul stochastique appliquée à la finance* de Lamberton et Lapeyre pour plus de détails sur ce modèle (version discrète du modèle de Black et Scholes du prochain chapitre) et surtout pour un traitement beaucoup plus complet des notions utilisées en finance et que nous n'introduisons que sur le cas du modèle.

En finance, on considère souvent un placement sans risque à intérêt fixe dont la valeur à l'instant n est $Z_n = Y_0(1 + r)^n$ si il vaut Y_0 à l'instant 0 et gagne un taux d'intérêt r à chaque pas de temps. On considère également un placement risqué de valeur Y_n à l'instant n.

On suppose que

$$Y_{n+1} = T_n Y_n$$

avec $T_n : \Omega \to \{1 + a, 1 + b\}$ une variable de Bernoulli telle que $0 < 1 + a < 1 + b$. a, b sont donc les taux d'intérêt gagnés possibles (qui peut être négatif ou inférieur à r) et $T_n - 1$ le taux d'intérêt aléatoire donc risqué. On considère la filtration engendrée $\mathcal{F}_n = \sigma(T_1, ..., T_{n-1})$.

1. Comme le placement à taux d'intérêt r est toujours disponible, on regarde $X_n = Y_n/Z_n$ que l'on appelle *valeur actualisée* à l'instant n. On cherche à savoir quand X_n est une martingale. Or $X_{n+1} = T_n Y_n/Z_n(1 + r) = X_n T_n/(1 + r)$ et X_n est $\mathcal{F}_n$ mesurable donc

$$\mathbf{E}(X_{n+1}|\mathcal{F}_n) = X_n \frac{\mathbf{E}(T_n|\mathcal{F}_n)}{1 + r}.$$

Donc X_n est une martingale si et seulement si $\mathbf{E}(T_n|\mathcal{F}_n) = 1 + r$ pour tout n.

2. On parle d'*arbitrage* si il y a possibilité d'avoir un gain strictement positif avec une mise nulle en empruntant/achetant éventuellement l'actif sans risque. Par exemple, si $r \geq b$ on vend un actif risqué (à découvert) pour une valeur Y_0 et on achète avec le gain de l'actif sans risque. A la date N, on récupère l'intérêt $Y_0(1+r)^N$ sur le marché sans risque et on paie la valeur Y_N de l'actif risqué soit un gain $Y_0(1+r)^N - Y_N \geq Y_0[(1+r)^N - (1+b)^N]$. On a donc gagné un montant strictement positif sans mise, donc sans risque de perdre de l'argent. De même, on montre que si $r \leq a$, il y a possibilité d'arbitrage (en empruntant sur le marché sans risque de quoi acheter l'actif risqué). Un marché financier est dit *viable* si il n'y a pas de possibilité d'arbitrage. Ici, cela impose $r \in]a, b[$.
3. On suppose maintenant $r \in]a, b[$.

   Lemme 63. *X_n est une martingale si et seulement si les T_n sont i.i.d de loi*
   $$P_{T_n} = \frac{b-r}{b-a}\delta_{1+a} + \frac{r-a}{b-a}\delta_{1+b}.$$

   Démonstration. Si les T_n sont i.i.d comme indiqué, il suffit de calculer vu que T_n indépendant de $\mathcal{F}_n$,
   $$\mathbf{E}(T_n|\mathcal{F}_n) = \mathbf{E}(T_n) = \frac{b-r}{b-a}(1+a) + \frac{r-a}{b-a}(1+b) = 1 + \frac{a(b-r)+b(r-a)}{b-a} = 1+r,$$
   comme il fallait d'après le point 1. Réciproquement, si pour tout n $\mathbf{E}(T_n|\mathcal{F}_n) = 1+r$, vu que $T_n = 1 + a1_{\{T_n=1+a\}} + b1_{\{T_n\neq 1+a\}} = 1 + b + (a-b)1_{\{T_n=1+a\}}$ est de Bernoulli, on voit
   $$r - b = \mathbf{E}((a-b)1_{\{T_n=1+a\}}|\mathcal{F}_n).$$
   Donc pour tout $A \in \mathcal{F}_n$, par (PC) pour l'espérance conditionnelle et le calcul ci-dessus
   $$P(T_n = 1+a|A) = \frac{\mathbf{E}(1_A 1_{\{T_n=1+a\}})}{P(A)} = \frac{\mathbf{E}(1_A \mathbf{E}(1_{\{T_n=1+a\}}|\mathcal{F}_n))}{P(A)} = \frac{(r-b)\mathbf{E}(1_A)}{(a-b)P(A)} = \frac{(r-b)}{(a-b)},$$
   donc $\{T_n = 1+a\}, \{T_n = 1+b\}$ sont indépendants de A donc de $\mathcal{F}_n$ comme voulu, soit T_n indépendant de $\mathcal{F}_n$ et est de loi souhaitée.

   $\square$

4. En finance, si l'actif actualisé T_n est une martingale pour une certaine loi de probabilité, il est facile de voir qu'il n'y a pas de stratégie d'arbitrage et que le marché est (donc dit) viable. Si en plus, comme ci-dessus, il y a une unique loi (ici pour les T_n) qui permet d'obtenir une martingale, on dit que le marché est *complet*, et on utilise cette (unique) loi pour calculer des prix et des stratégies (de couverture pour les marchés dérivés).
5. Les fameux "produits dérivés" intervenant en finance, parmi les plus simples, sont les options européennes d'achat et de vente construites sur l'actif risqué Y_n. Une option d'achat (ou en anglais call) est le droit (mais non l'obligation) d'acheter l'actif à l'instant N au prix K (l'adjectif européen désigne les options d'achat-vente à terme, ici l'instant final N). Elle rapporte donc la différence $C_N = (Y_N - K)_+ = \max(0, Y_N - K)$ (puisque si $Y_N - K \neq 0$ on n'a pas intérêt à acheter l'actif comme on a le droit avec l'option). De même, une option de vente (ou put) au prix K rapporte $P_N = (K - Y_N)_+ = \max(0, K - Y_N)$. On appelle valeur de l'option à l'instant n, l'actif dont la valeur actualisée est une martingale fermée de valeur à l'instant N, la valeur (actualisée) de l'option (C_N ou P_N), soit :
   $$C_n = Z_n \mathbf{E}(\frac{(Y_N - K)_+}{Z_N}|\mathcal{F}_n) = (1+r)^{n-N}\mathbf{E}((Y_N - K)_+|\mathcal{F}_n), \quad P_n = (1+r)^{n-N}\mathbf{E}((K - Y_N)_+|\mathcal{F}_n).$$

Les probabilités pour lesquels sont calculés ces valeurs (dans un marché viable et complet) est celle pour laquelle le prix de l'actif actualisé est une martingale.

6. Calculons la "relation de parité call-put" :

$$\begin{aligned} C_n - P_n &= (1+r)^{n-N}\mathbf{E}((S_N-K)_+ - (K-S_N)_+|\mathcal{F}_n) = (1+r)^{n-N}\mathbf{E}(S_N - K|\mathcal{F}_n) \\ &= (1+r)^n\mathbf{E}(X_N|\mathcal{F}_n) - (1+r)^{n-N}K = (1+r)^n X_n - (1+r)^{n-N}K \\ &= S_n - K(1+r)^{n-N}. \end{aligned}$$

On a utilisé que $X_n = S_n(1+r)^{-n}$ est une martingale par hypothèse et la relation élémentaire $(S_N-K)_+ - (K-S_N)_+ = S_N - K$.

7. Calculons finalement la valeur explicite de C_n dans ce modèle c'est à dire

$$C_n = (1+r)^{n-N}\mathbf{E}((Y_nT_n...T_{N-1} - K)_+|\mathcal{F}_n).$$

Pour appliquer la proposition 60 puisque Y_n et $T_n...T_{N-1}$ sont indépendants et $\ln(T_n...T_{N-1})$ est une somme de v.a. de Bernoulli indépendantes de loi $p\delta_{\ln(1+a)} + (1-p)\delta_{\ln(1+b)}$ avec $p = \frac{b-r}{b-a}$ de sorte que $\frac{1}{\ln(\frac{1+b}{1+a})}[\ln(T_n...T_{N-1}) - (N-n)\ln(1+a)]$ est de loi binomiale $\mathcal{B}(N-n, 1-p)$ (et vaut j quand $T_n...T_{N-1}$ vaut $(1+b)^j(1+a)^{N-n-j}$, $N \geq n+1$) et donc on calcule par transfert

$$c(n,s) := \mathbf{E}((sT_n...T_{N-1} - K)_+) = \sum_{k=0}^{N-n} C_{N-n}^j p^j(1-p)^{N-n-j}(s(1+a)^j(1+b)^{N-n-j} - K)_+.$$

On obtient finalement la formule explicite $C_n = (1+r)^{n-N}c(n, Y_n)$.

Exemple 43 (Modèle simple de portefeuille d'actions). Il est utile de considérer le cas symétrique $r = \frac{a+b}{2}$. Dans ce cas, on appelle r taux d'actualisation et $\sigma = \frac{b-a}{2}$ le coefficient de volatilité. On pose $\epsilon_n = \frac{T_n-(1+r)}{\sigma} \in \{-1, 1\}$. On a donc

$$Y_n = (1 + r + \sigma\epsilon_n)Y_{n-1}.$$

On a donc vu que $X_n = (1+r)^{-n}Y_n$ est une martingale si et seulement si les ϵ_n sont i.i.d avec $P_{\epsilon_n} = \frac{1}{2}(\delta_1 + \delta_{-1})$. En supposant ce cadre, on voit que Y_n est une martingale si et seulement si $r = 0$, une sous-martingale (resp. surmartingale) si et seulement si $r > 0$ (resp. $r < 0$). Noter que la condition $1 + a > 0$ revient à $|\sigma| < 1 + r$.

On illustrera par la suite les théorèmes sur les martingales sur l'exemple de Y_n (dans le cas ϵ_n i.i.d comme ci-dessus).

4 Théorème de convergence des martingales

Le théorème suivant résume les résultats fondamentaux de convergences :

Théorème 64. *Soit (X_n) un processus adapté pour la filtration $(\mathcal{F}_n)$*

1. *Si (X_n) est une martingale, bornée dans $L^p(\Omega)$ pour $p \in]1, \infty[$, c'est-à-dire :*

$$\sup_n \mathbf{E}(|X_n|^p) < \infty,$$

alors X_n converge p.s. et dans L^p vers $X_\infty \in L^p$ et $||X_\infty||_p = \sup_{n\in I\!N}||X_n||_p$. De plus, (X_n) est une martingale fermée :

$$X_n = \mathbf{E}(X_\infty|\mathcal{F}_n).$$

2. *Si (X_n) est une sous-martingale et*

$$\sup_{n\in\mathbb{N}} \mathbf{E}(\max(0, X_n)) < \infty$$

(par exemple si (X_n) est une martingale bornée dans L^1) alors X_n converge p.s. vers une variable aléatoire $X_\infty \in L^1(\Omega)$.

3. *Si (X_n) est une martingale uniformément intégrale, c'est-à-dire*

$$\lim_{C\to\infty} \sup_n \mathbf{E}(1_{\{|X_n|>C\}}|X_n|) = 0,$$

alors X_n converge p.s. et dans L^1 vers X_∞ et (X_n) est une martingale fermée :

$$X_n = \mathbf{E}(X_\infty|\mathcal{F}_n).$$

Exemple 44. On reprend l'exemple 43 $Y_n = \prod_{k=1}^n (1 + r + \sigma\epsilon_k)$ avec ϵ_k i.i.d. $1/2(\delta_1 + \delta_{-1})$. On a vu que c'est une sur-martingale pour $r \leq 0$ et comme $1 + r - \sigma > 0$ elle est positive. Donc $-Y_n$ est alors une sous-martingale et $\max(0, -Y_n) = 0$ est borné donc $-Y_n$ converge p.s. par le (2) du théorème dans le cas $r \neq 0$.

De plus par indépendance $\mathbf{E}(Y_n) = [\mathbf{E}((1 + r + \sigma\epsilon_k))]^n = (1 + r)^n$. Donc si $r < 0$, Y_n converge p.s. et dans L^1 vers 0 (car $\mathbf{E}(Y_n) = \mathbf{E}(|Y_n|)$).

Exemple 45. Dans l'exemple précédent on considère $X_n = (1 + r)^{-n} Y_n$ qui est toujours une martingale positive donc converge p.s. En prenant le log $log(X_n) = \sum_{k=1}^n \log(1 + r + \sigma\epsilon_k) - \log(1 + r)$ et $\mathbf{E}(\log(1+r+\sigma\epsilon_k) - \log(1+r)) = \frac{1}{2}(\log(1+r+\sigma) + \log(1+r-\sigma) - 2\log(1+r)) = \frac{1}{2}\log(1 - \frac{\sigma^2}{(1+r)^2}) = c < 0$ vu $|\sigma| < 1 + r$. Donc par la loi des grands nombres (pour des variables L^∞ donc L^2 i.i.d), $\frac{log(X_n)}{n} \to c < 0$ donc $log(X_n) \to -\infty$ p.s. et donc la limite p.s. de X_n est 0. On obtient donc aussi la limite 0 pour Y_n dans le cas $r = 0$.

4.1 Martingales rétrogrades et application à la loi des Grands nombres

Définition 38. Un processus $(X_n)_{n\in\mathbb{N}}$ adapté à une suite décroissante $(\mathcal{F}_n)$ de sous-tribus et dans L^1 (c'est-à-dire $\mathbf{E}(|X_n|) < \infty$) est :

1. une *martingale rétrograde* si pour $m \leq n$

$$\mathbf{E}(X_m|\mathcal{F}_n) = X_n,$$

2. une *sous-martingale rétrograde* si pour $m \leq n$

$$\mathbf{E}(X_m|\mathcal{F}_n) \geq X_n,$$

3. une *sur-martingale rétrograde* si pour $m \leq n$

$$\mathbf{E}(X_m|\mathcal{F}_n) \leq X_n.$$

Théorème 65. *Toute martingale rétrograde bornée dans L^p pour $1 \leq p < \infty$ converge p.s. et dans L^p.*

Pour obtenir une application, on montre :

Lemme 66. *Soit Z_n une suite i.i.d dans L^1 et $S_n = Z_1 + ... + Z_n$, $\mathcal{F}_n = \sigma(S_k, k \geq n)$. Alors $\frac{S_n}{n}$ est une martingale rétrograde pour la filtration $\mathcal{F}_n$.*

Démonstration. Montrons que $\mathbf{E}(Z_1|\mathcal{F}_n) = \frac{S_n}{n}$. D'abord $\sigma(S_n, ...S_{n+m}) = \sigma(S_n, Z_{n+1}, ..., Z_{n+m})$ avec Z_k, S_n indépendants de $Z_{n+1}, ...Z_{n+m}$ pour $k \leq n$ donc par le théorème 55 (12)

$$\mathbf{E}(Z_k|\sigma(S_n, Z_{n+1}, ..., Z_{n+m})) = \mathbf{E}(Z_k|\sigma(S_n)) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}(Z_k|\sigma(S_n)) = \frac{S_n}{n}$$

en utilisant le caractère i.i.d pour voir $\mathbf{E}(Z_k|\sigma(S_n)) = \mathbf{E}(Z_1|\sigma(S_n))$ pour $k \leq n$. En utilisant $\mathcal{G}_m = \sigma(S_n, Z_{n+1}, ..., Z_{n+m})$ est une filtration croissante et $\mathbf{E}(Z_1|\mathcal{G}_m)$ une martingale terminée, on obtient qu'elle converge (par la proposition 67) vers $\mathbf{E}(Z_k|\sigma(S_n, Z_{n+m}m \geq 1)) = \mathbf{E}(Z_k|\mathcal{F}_n)$. D'où $\mathbf{E}(Z_k|\mathcal{F}_n) = \frac{S_n}{n}, k \leq n$. En faisant la moyenne on obtient pour $m \leq n$

$$\mathbf{E}(\frac{S_m}{m}|\mathcal{F}_n) = \frac{S_n}{n}.$$

□

Par le théorème de convergence des martingales rétrogrades, on déduit donc que $\frac{S_n}{n}$ converge p.s. et dans L^1 vers une limite Y. Mais la variable limite est mesurable pour la tribu assymptotique donc par la loi du $0-1$ est p.s. constante, et par convergence L^1 on a $Y = \mathbf{E}(Y) = \lim_n \mathbf{E}(\frac{S_n}{n}) = \mathbf{E}(X_1)$. Donc on a déduit la loi forte des grands nombres.

4.2 Preuves facultatives des théorèmes de convergences

Les preuves seront inhabituelles pour les probabilistes et proviennent de l'étude des martingales à valeur espace de Banach. Nous les devons à un cours (non publié) de Gilles Pisier.

On commence par montrer une propriété intuitive de l'espérance conditionnelle sur des suites de tribus croissantes ou décroissantes.

Proposition 67.
- *Soit $(\mathcal{F}_n)$ une filtration (croissante). Tout martingale fermée $X_n = \mathbf{E}(X|\mathcal{F}_n)$ avec $X \in L^p(\Omega, \mathcal{T})$ pour $p \in [1, \infty[$ converge dans $L^p(\Omega, \mathcal{T})$ vers $\mathbf{E}(X|\mathcal{F})$ avec $\mathcal{F} = \sigma(\bigcup_n \mathcal{F}_n)$.*
- *Soit $(\mathcal{F}_n)$ une filtration décroissante. Pour $f \in L^p(\Omega, \mathcal{T})$, pour $p \in [1, \infty[$ $\mathbf{E}(f|\mathcal{F}_n)$ converge dans $L^p(\Omega, \mathcal{T})$ vers $\mathbf{E}(f|\mathcal{F}_\infty)$ avec $\mathcal{F}_\infty = \cap_n \mathcal{F}_n$.*

PREUVE : – Dans le premier cas, on montre que $X_n \overset{L^p}{\to}_{n\to\infty} \mathbf{E}(X|\mathcal{F}_\infty)$. Soit $\epsilon > 0$.

On remarque d'abord que $L = \cup_n L^p(\Omega, \mathcal{F}_n)$ est dense dans $L^p(\Omega, \mathcal{F}_\infty)$.

En effet, l'ensemble des 1_A, $A \in \mathcal{F}_n$ est inclus dans L. De plus $\mathcal{B} = \{A \in \mathcal{T}, \forall \epsilon > 0, \exists B \in \cup_n \mathcal{F}_n, ||1_A - 1_B||_1 = P(A \Delta B) \leq \epsilon\}$ est une tribu (exo) qui contient $\cup_n \mathcal{F}_n$ et donc contient la tribu engendrée $\mathcal{F}$. Donc $\overline{L \cap B_{L^\infty}(0,1)}^{L^p} = \overline{L \cap B_{L^\infty}(0,1)}^{L^1}$ contient les $1_A, A \in \mathcal{F}_\infty$. Or par définition, l'espace vectoriel engendré par les fonctions étagées est dense dans $L^p(\Omega, \mathcal{F}_\infty)$ donc L aussi.

Soit donc $g \in L^p(\Omega, \mathcal{F}_N)$ avec $||g - \mathbf{E}(X|\mathcal{F}_\infty)|| \leq \epsilon$. Or pour $n \geq N$ $\mathcal{F}_N \subset \mathcal{F}_n$ donc $\mathbf{E}(g|\mathcal{F}_n) = g$ et de même $X_n = \mathbf{E}(X|\mathcal{F}_n) = \mathbf{E}(\mathbf{E}(X|\mathcal{F}_\infty)|\mathcal{F}_n)$ par conditionnement successif (théorème 55 (11)).

Donc par inégalité triangulaire, on obtient pour $n \geq N$:

$$\begin{aligned}
||X_n - \mathbf{E}(X|\mathcal{F}_\infty)||_p &= ||\mathbf{E}(\mathbf{E}(X|\mathcal{F}_\infty)|\mathcal{F}_n) - \mathbf{E}(X|\mathcal{F}_\infty)||_p \\
&\leq ||\mathbf{E}(\mathbf{E}(X|\mathcal{F}_\infty) - g|\mathcal{F}_n)||_p + ||\mathbf{E}(g|\mathcal{F}_n) - g||_p + ||g - \mathbf{E}(X|\mathcal{F}_\infty)||_p \\
&\leq 2||g - \mathbf{E}(X|\mathcal{F}_\infty)||_p \leq 2\epsilon
\end{aligned}$$

en utilisant la contractivité de l'espérance conditionelle en norme L^p.

– Dans le deuxième cas des filtrations rétrogrades, si $p=2$ les espces $(L^2(\Omega,\mathcal{F}_n)\ominus L^2(\Omega,\mathcal{F}_{n+1}))_{n\in\mathbb{N}}$ (complémentaire orthogonale de $L^2(\Omega,\mathcal{F}_{n+1})$ dans $L^2(\Omega,\mathcal{F}_n)$) sont orthogonaux, donc, par le théorème de Pythagore, pour $n\geq l$, on a :

$$||\mathbf{E}(f|\mathcal{F}_l)||_2^2 = ||\mathbf{E}(f|\mathcal{F}_{n+1})||_2^2 + \sum_{k=l}^{n} ||\mathbf{E}(f|\mathcal{F}_k) - \mathbf{E}(f|\mathcal{F}_{k+1})||_2^2$$

donc la série $\sum_k ||\mathbf{E}(f|\mathcal{F}_k) - \mathbf{E}(f|\mathcal{F}_{k+1})||_2^2$ converge donc la suite $\mathbf{E}(f|\mathcal{F}_l)$ est de Cauchy dans L^2 car $||\mathbf{E}(f|\mathcal{F}_l) - \mathbf{E}(f|\mathcal{F}_{n+1})||_2^2 = \sum_{k=l}^{n} ||\mathbf{E}(f|\mathcal{F}_k) - \mathbf{E}(f|\mathcal{F}_{k+1})||_2^2$. Donc $\mathbf{E}(f|\mathcal{F}_l)$ converge vers f_∞ qui est dans $\mathcal{F}_l$ pour tout l par décroissance de la filtration de sorte que $\mathbf{E}(f|\mathcal{F}_n)$ est asymptotiquement dans $\mathcal{F}_l$. Or, par (PC), on a pour $g\in L^2(\mathcal{F}_\infty)$

$$\mathbf{E}(f_\infty g) = \lim_{n\to\infty} \mathbf{E}(\mathbf{E}(f|\mathcal{F}_n)g) = \mathbf{E}(fg)$$

ce qui donne par (PC) $f_\infty = \mathbf{E}(f|\mathcal{F}_\infty)$. Pour $p\leq 2$, on raisonne par densité et pour $f\in L^p$, on prend $g\in L^2$ avec $||g-f||_p\leq\epsilon$ de sorte que par contractivité de l'espérance conditionnelle :

$$\limsup_n ||\mathbf{E}(f|\mathcal{F}_n) - \mathbf{E}(f|\mathcal{F}_\infty)||_p \leq 2||f-g||_p + \limsup_n ||\mathbf{E}(g|\mathcal{F}_n) - \mathbf{E}(g|\mathcal{F}_\infty)||_2 \leq 2\epsilon.$$

Pour $p\geq 2$, par Hölder et le cas hilbertien on a dans le cas $f\in L^\infty$:

$$||\mathbf{E}(f|\mathcal{F}_n) - \mathbf{E}(f|\mathcal{F}_\infty)||_p^p \leq ||\mathbf{E}(f|\mathcal{F}_n) - \mathbf{E}(f|\mathcal{F}_\infty)||_2^2 ||\mathbf{E}(f|\mathcal{F}_n) - \mathbf{E}(f|\mathcal{F}_\infty)||_\infty^{p-2} \to 0$$

et on conclut comme avant par densité (ici de L^∞ dans L^p).

■

Le deuxième ingrédient est l'inégalité maximale de Doob

Proposition 68 (Inégalité maximale de Doob)**.** *Soit (X_n) une sous-martingale et soit*

$$X_n^* = \sup_{0\leq k\leq n} X_k,$$

alors pour tout $a>0$:

$$aP(X_n^*\geq a) \leq \mathbf{E}(X_n 1_{\{X_n^*\geq a\}}) \leq \mathbf{E}(\max(0,X_n)).$$

En particulier, on a :

$$aP(\sup_{n\geq 0} X_n \geq a) \leq \sup_n \mathbf{E}(\max(0,X_n)).$$

Preuve : Soit le temps d'atteinte (exemple 42) $T=\inf\{n\in\mathbb{N} X_n\geq a\}$ de sorte que

$$A := \{X_n^*\geq a\} = \{\exists k\in[0,n] X_k\geq a\} = \{T\leq n\}.$$

En appliquant le premier théorème d'arrêt 62 au temps d'arrêt $T\wedge n\leq n$ on obtient $\mathbf{E}(X_{T\wedge n})\leq\mathbf{E}(X_n)$. Or comme $X_{T\wedge n}1_{A^c} = X_n 1_{A^c}$ et sur A, par définition de T, on a $X_{T\wedge n}1_A = X_T 1_A \geq a1_A$, on obtient

$$X_{T\wedge n} = X_{T\wedge n}(1_A + 1_{A^c}) \geq a1_A + X_n 1_{A^c}$$

Donc on déduit

$$\mathbf{E}(X_n 1_A) = \mathbf{E}(X_n) - \mathbf{E}(X_n 1_{A^c}) \geq \mathbf{E}(X_{T\wedge n}) - \mathbf{E}(X_n 1_{A^c}) \geq aP(A)$$

Or $X_n\leq\max(0,X_n)$ donc $\mathbf{E}(X_n 1_A)\leq\mathbf{E}(\max(0,X_n))$ est évident. En prenant la limite $n\to\infty$ dans la borne $aP(X_n^*\geq a)\leq\sup_k\mathbf{E}(\max(0,X_k))$, on obtient la deuxième inégalité. ■

En conséquence, on obtient le :

Lemme 69. *Une martingale* (X_n) *qui converge dans* L^1 *converge aussi p.s.*

PREUVE : Comme (X_n) est de Cauchy dans L^1, on déduit que pour $\epsilon > 0$, on a N tel que pour $p, q \geq N$, $||X_p - X_q|| \leq \epsilon$. On pose $Y_n = 1_{\{n \geq N\}}(X_n - X_N)$. Si $n \geq N$, X_N est $\mathcal{F}_n$ mesurable donc

$$\mathbf{E}(Y_{n+1}|\mathcal{F}_n) = \mathbf{E}(X_{n+1} - X_N|\mathcal{F}_n) = \mathbf{E}(X_{n+1}|\mathcal{F}_n) - X_N = X_n - X_N = Y_n$$

et de plus $n < N$ $Y_{n+1} = Y_n = 0$ donc (Y_n) est une martingale et donc aussi $(|Y_n|)$ est une sous-martingale car par Jensen conditionnelle $\mathbf{E}(|Y_{n+1}||\mathcal{F}_n) \geq |\mathbf{E}(Y_{n+1}|\mathcal{F}_n)| = |Y_n|$. En appliquant l'inégalité de Doob, on obtient pour $t > 0$

$$tP(\sup_{n \geq N} |X_n - X_N| \geq t) = tP(\sup_n |Y_n| \geq t) \leq \sup_{n \geq N} \mathbf{E}(|X_n - X_N|) \leq \epsilon.$$

Or $l = \lim_{N\to\infty} \sup_{p,q \geq N} |X_p - X_q| \leq 2 \sup_{p \geq N} |X_p - X_N|$ pour tout N. Donc $P(l \geq 2t) \leq P(\sup_{p \geq N} |X_p - X_N| \geq t) \leq \epsilon/t$ et comme ϵ est arbitraire, $P(l \geq 2t) = 0$ pour tout t donc $l = 0$ p.s. donc (X_n) est p.s. de Cauchy, donc converge p.s.

■

Preuve du Théorème 64 (1). Comme (X_n) est une martingale bornée dans L^p et que la boule unité de L^p est weak-* compact, on extrait $X_{n_k} \overset{*-L^p}{\to}_{n\to\infty} X$. Montrons que X_n est une martingale fermée $X_n = \mathbf{E}(X|\mathcal{F}_n)$ en appliquant la propriété caractéristique (PC) Soit $f \in L^\infty(\Omega, \mathcal{F}_n) \subset L^q(\Omega, \mathcal{F}_n)$, avec q l'exposant conjugué de p :

$$\mathbf{E}(Xg) = \lim_{k\to\infty} \mathbf{E}(X_{n_k} g) = \lim_{k\to\infty} \mathbf{E}(\mathbf{E}(X_{n_k}|\mathcal{F}_n)g) = \mathbf{E}(X_n g)$$

car $\mathbf{E}(X_{n_k}|\mathcal{F}_n) = X_n$ pour k assez grand. On a utilisé que la convergence faible-* donne exactement la convergence des produits scalaires $\mathbf{E}(X_{n_k} g) \to \mathbf{E}(Xg)$ et maintenant (PC) donne en effet la conclusion $X_n = \mathbf{E}(X|\mathcal{F}_n)$. Comme $X \in L^p$, on peut appliquer la proposition 67 pour montrer que X_n converge dans L^p d'où la limite $X_\infty = X$ et on a établi la relation de martingale fermée souhaitée $X_n = \mathbf{E}(X_\infty|\mathcal{F}_n)$. Comme $||\mathbf{E}(X_\infty|\mathcal{F}_n)||_p \leq ||X_\infty||_p$ et qu'on a convergence L^p donc par inégalité triangulaire inverse $||\mathbf{E}(X_\infty|\mathcal{F}_n)||_p \to ||X_\infty||_p$, on en déduit la relation entre les normes $||X_\infty||_p = \sup_n ||\mathbf{E}(X_\infty|\mathcal{F}_n)||_p$. La convergence p.s. se déduit alors de la convergence L^1 et du lemme précédent. □

Pour le cas uniformément intégrable, la preuve est la même en utilisant le critère suivant de compacité. (cf. annexe pour une preuve des implications que l'on va utiliser)

Théorème 70 (Dunford-Pettis)**.** *Soit une suite* (X_n) *dans* L^1*. On a l'équivalence entre*

1. (X_n) *est uniformément intégrable*
2. (X_n) *admet une sous-suite* (X_{n_k}) *ayant pour limite faible* $X \in L^1$*, c'est-à-dire :*

$$\forall f \in L^\infty(\Omega), \quad \mathbf{E}((X_{n_k} - X)f) \to 0.$$

3. (X_n) *est bornée dans* L^1 *et pour tout* $\epsilon > 0$*, il existe* $\eta > 0$ *tel que si* $A \in \mathcal{T}$ *vérifie* $P(A) \leq \eta$ *alors pour tout* n, $\mathbf{E}(1_A|X_n|) \leq \epsilon$.

Preuve du Théorème 64 (3). En extrayant comme dans le (1), on montre que X_n est fermée donc converge dans L^1 et donc aussi p.s. par les résultats précédents.

□

Le lemme suivant va nous permettre de nous ramener au cas uniformément intégrable.

Lemme 71. *Si (X_n) est une martingale bornée dans L^1, alors si $T_t = \inf\{n : |X_n| > t\}$ alors $(X_{n\wedge T_t})$ est une martingale uniformément intégrable.*

Preuve : On a déjà vu que le 1er théorème d'arrêt 62 implique que $(X_{n\wedge T_t})$ est une martingale. On note maintenant T pour T_t.

Comme $(|X_n|)$ est une sous-martingale (voir preuve du lemme 69) on a $\mathbf{E}(|X_n||\mathcal{F}_k) \geq |X_k|$ donc par (PC) vu $\{T = k\} = \{T \leq k\} \cap \{T \leq k-1\}^c \in \mathcal{F}_k$ on a pour $k \leq n$:

$$\mathbf{E}(|X_n|1_{\{T=k\}}) = \mathbf{E}(1_{\{T=k\}}\mathbf{E}(|X_n||\mathcal{F}_k)) \geq \mathbf{E}(|X_k|1_{\{T=k\}}) = \mathbf{E}(|X_T|1_{\{T=k\}})$$

donc en sommant sur k

$$\mathbf{E}(|X_T|1_{\{T\leq n\}}) \leq \mathbf{E}(|X_n|1_{\{T\leq k\}}) \leq \sup_n \mathbf{E}(|X_n|).$$

En prenant le suprémum sur n, on déduit $|X_T|1_{\{T<\infty\}} \in L^1(\Omega)$.

De plus, $\sup_n |X_{n\wedge T}| = \max(1_{\{T\leq n\}}|X_T|, \sup_{n<T}|X_n|)$ et $\sup_{n<T}|X_n| \leq t$ par définition de T et $1_{T\leq n}|X_T| \leq |X_T|1_{\{T<\infty\}}$ d'où on déduit $\sup_n |X_{n\wedge T}| \leq Z := |X_T|1_{\{T<\infty\}} + t \in L^1(\Omega)$. Or, par définition, la famille de un élément $\{Z\}$ est u.i. donc par le théorème 84 (1) $\Rightarrow$ (3) on déduit que, soit $\epsilon > 0$, on a $\eta > 0$ tel que si $P(A) \leq \eta$, on a $\mathbf{E}(Z1_A) \leq \epsilon$ de sorte que $\mathbf{E}(|X_{n\wedge T}|1_A) \leq \epsilon$ pour tout n et donc par la réciproque, $(X_{n\wedge T})$ est bien uniformément intégrale. ■

Preuve du Théorème 64 (2) (Cas martingale). On prend une martingale (X_n) bornée $(X_{n\wedge T_t})$ est donc u.i. par le lemme, donc converge p.s. par le (3) du Thm. Or par l'inégalité de Doob pour la sous-martingale $(|X_n|)$, on a :

$$tP(T_t < \infty) = tP(\sup_n |X_n| > t) \leq \sup_n \mathbf{E}(|X_n|) =: C.$$

Or on conclut donc (vu que sous $\{T_t < \infty\}$, $X_{n\wedge T_t} - X_n \to 0$ de façon stationnaire)

$$P(X_n\, converge) \geq P(X_{n\wedge T}\, converge\, et\, T_t < \infty) \geq P(X_{n\wedge T_t}\, converge) - P(T_t < \infty) \geq 1 - \frac{C}{t} \to_{t\to\infty} 1$$

Notons X_∞ la limite p.s. Par Fatou, on a

$$E[|X_\infty|] \leq \liminf_n E[|X_n|] \leq \sup_n E[|X_n|] < \infty.$$

□

Pour obtenir le cas sous-martingale, il suffit d'utiliser la fameuse décomposition de Doob des sous-martingales :

Proposition 72 (Décomposition de Doob)**.** *Soit (X_n) une sous-martingale il existe Y_n, Z_n avec $X_n = Y_n + Z_n$, Y_n martingale et Z_n croissant prévisible (c'est-à-dire Z_n $\mathcal{F}_{n-1}$-mesurable).*

Preuve : On pose $Z_0 = 0$, et pour $n \geq 1$,

$$Z_n = \sum_{k=1}^{n} \mathbf{E}(X_k - X_{k-1}|\mathcal{F}_{k-1})$$

qui est par définition $\mathcal{F}_{n-1}$-mesurable et croissant vu par la propriété des sous-martingales $\mathbf{E}(X_k - X_{k-1}|\mathcal{F}_{k-1}) \geq 0$. Si on pose $Y_n = X_n - Z_n$, on obtient bien une martingale :

$$\mathbf{E}(Y_n|\mathcal{F}_{n-1}) = \mathbf{E}(X_n - Z_n|\mathcal{F}_{n-1}) = \mathbf{E}(X_n - X_{n-1} - \mathbf{E}(X_n - X_{n-1}|\mathcal{F}_{n-1})|\mathcal{F}_{n-1}) + X_{n-1} - Z_{n-1} = Y_{n-1}.$$

■

Preuve du Théorème 64 (2) (Cas sous-martingale). Comme $\mathbf{E}(X_n) = \mathbf{E}(\max(0, X_n)) - \mathbf{E}(\max(0, -X_n)) \geq \mathbf{E}(X_0)$, on obtient

$$\sup_n \mathbf{E}(\max(0, -X_n)) \leq -\mathbf{E}(X_0) + \sup_n \mathbf{E}(\max(0, X_n))$$

d'où

$$\sup_n \mathbf{E}(|X_n|) \leq \sup_n \mathbf{E}(\max(0, -X_n)) + \sup_n \mathbf{E}(\max(0, X_n)) < \infty.$$

Donc la sous-martingale est en fait bornée dans L^1. On prend la décomposition de Doob ci -dessus et $\mathbf{E}(X_n) = \mathbf{E}(Z_n) + \mathbf{E}(X_0)$ vu $\mathbf{E}(Y_n) = \mathbf{E}(Y_0) = \mathbf{E}(X_0)$. Donc par convergence monotone : $E(\sup_n Z_n) = \sup E(Z_n) = \sup E(X_n - X_0) < \infty$. Donc en particulier Z_n borné dans L^1 et converge p.s. (par croissance) et dans L^1 (par l'application du TCM ci-dessus) et donc Y_n est aussi borné dans L^1. Donc par le cas martingale Y_n converge p.s., d'où, par somme, aussi X_n. $\square$

On peut aussi conclure la preuve beaucoup plus simple du théorème de convergence des martingales rétrogrades.

Preuve du Théorème 65 . Comme $X_n = \mathbf{E}(X_1|\mathcal{F}_n)$, on a par la proposition 67 la convergence dans L^p et il suffit de voir la convergence p.s. En remplaçant X_1 par $X_1 - \mathbf{E}(X_1|\mathcal{F}_\infty)$ on est réduit au cas de convergence p.s. vers 0.

Pour $n, k > 0$, $M_j = X_{n+k-j}1_{j\leq k} + X_n 1_{j>k}$ est une martingale ordinaire pour la filtration $\mathcal{F}_{\min(-n,-n-k+j)}$ et donc l'inégalité de Doob donne pour $t > 0$

$$tP(\sup_{n\leq l\leq n+k} |X_l| \geq t) = tP(\sup_{0\leq j\leq k} |M_j| \geq t) \leq \mathbf{E}(|M_k|) = \mathbf{E}(|X_n|)$$

soit en faisant tendre $k \to \infty$

$$tP(\sup_{l\geq n} |X_l| \geq t) \leq \mathbf{E}(|X_n|).$$

Or par la convergence L^1, on a $\mathbf{E}(|X_n|) \to 0$. Donc $\sup_{l\geq n} |X_l| \to 0$ en proba, donc par le lemme de Borel Cantelli, une sous-suite $\sup_{l\geq n_k} |X_l|$ converge p.s. vers 0 et donc $X_n \to 0$ p.s. comme souhaité. $\square$

5 Théorème d'arrêt optionnel de Doob

Nous montrons maintenant la version complète du premier théorème d'arrêt de Doob :

Théorème 73 (1er théorème d'arrêt de Doob)**.** *Si (X_n) est une martingale (resp. une sur-martingale) et S, T des temps d'arrêt bornés $S, T \in L^\infty(\Omega)$, avec $S \leq T$ alors $X_T, X_S \in L^1(\Omega)$ et on a :*

$$\mathbf{E}[X_T|\mathcal{F}_S] = X_S \quad (\mathit{resp.}\ \mathbf{E}[X_T|\mathcal{F}_S] \leq X_S)$$

Preuve : On a vu à la preuve de la proposition 62 que $X_T, X_S \in L^1(\Omega)$ Or si $A \in \mathcal{F}_S$, $A \cap \{S = j\} \in \mathcal{F}_j$ donc pour $k \geq j$ en notant $S = S \wedge T$

$$\mathbf{E}(1_{A\cap\{S=j\}}(X_{k\wedge T} - X_S)) = \mathbf{E}(1_{A\cap\{S=j\}}(X_{k\wedge T} - X_{j\wedge T})) = 0\ (resp \leq 0)$$

en utilisant que $(X_{n\wedge T})$ est une martingale. En sommant sur j (entre 1 et une borne fini N pour S) en notant $\sum_{j=1}^N 1_{A\cap\{S=j\}} = 1_A$, on obtient $\mathbf{E}(1_A(X_{k\wedge T} - X_S)) = 0$ donc par (PC') $\mathbf{E}((X_{k\wedge T} - X_S)|\mathcal{F}_S) = 0(resp \leq 0)$.

Or, pour un borélien B, $X_S^{-1}(B) \cap \{S = j\} = X_j^{-1}(B) \cap \{S = j\}$ est $\mathcal{F}_j$-mesurable donc X_S est $\mathcal{F}_S$-mesurable. On a donc obtenu $\mathbf{E}(X_{k\wedge T}|\mathcal{F}_S) = X_S(resp \leq X_S)$. En prenant pour k une borne pour T on obtient le résultat.

■

Exemple 46. Soit Y_n la martingale de l'exemple 43 dans le cas $r = 0$. Quelle que soit le temps d'arrêt borné T, on trouve $\mathbf{E}(Y_T) = \mathbf{E}(Y_0)$.

Exemple 47. Le Théorème précédent ne peut pas s'appliquer pour tout temps d'arrêt non borné. Par exemple si $X_n = Y_1 + ... + Y_n$ somme de bernoulli $\frac{1}{2}(\delta_1 + \delta_{-1})$ i.i.d. et $T = \inf\{n \in \mathbb{N} X_n = 1\}$ est un temps d'arrêt et il est facile de voir que $P(T < \infty) = 1$, mais $X_T = 1$ Donc $E(X_T) = 1 \neq E(X_0) = 0$.

Les deux résultats suivant donnent des hypothèses sur T et (X_n) pour avoir égalité.

Théorème 74 (2eme théorème d'arrêt de Doob)**.** *Si (X_n) est une martingale uniformément intégrable de limite X_∞ et T un temps d'arrêt. On note*

$$X_T = X_\infty 1_{\{T=\infty\}} + \sum_{k=0}^{\infty} X_k 1_{\{T=k\}}$$

alors $X_T \in L^1(\Omega)$ et on a :

$$\mathbf{E}[X_\infty|\mathcal{F}_T] = X_T$$

de sorte que $\mathbf{E}[X_T] = \mathbf{E}[X_0]$.

Preuve : La preuve va utiliser fortement le découpage $1 = 1_{\{T=\infty\}} + \sum_{k=0}^{\infty} 1_{\{T=k\}}$, $|X_T| = |X_\infty| 1_{\{T=\infty\}} + \sum_{k=0}^{\infty} |X_k| 1_{\{T=k\}}$

On vérifie $X_T \in L^1$ en calculant

$$\mathbf{E}[|X_T|] = \mathbf{E}[|X_\infty| 1_{\{T=\infty\}} + \sum_{k=0}^{\infty} \mathbf{E}[|X_k| 1_{\{T=k\}}] \leq \mathbf{E}[|X_\infty| 1_{\{T=\infty\}} + \sum_{k=0}^{\infty} \mathbf{E}[|X_\infty| 1_{\{T=k\}}] = \mathbf{E}[|X_\infty|]$$

où on a utilisé $|X_k| = |\mathbf{E}[X_\infty|\mathcal{F}_k]| \leq \mathbf{E}(|X_\infty||\mathcal{F}_k)$ et (PC) pour obtenir $\mathbf{E}(|X_k| 1_{\{T=k\}}) \leq \mathbf{E}(\mathbf{E}(|X_\infty||\mathcal{F}_k) 1_{\{T=k\}})$ $\mathbf{E}[|X_\infty| 1_{\{T=k\}}]$. De même pour $A \in \mathcal{F}_T$, on a

$$\mathbf{E}[1_A X_\infty] = \sum_{k \in \mathbb{N} \cup \{\infty\}} \mathbf{E}[1_{A \cap \{T=k\}} X_\infty] = \sum_{k \in \mathbb{N} \cup \{\infty\}} \mathbf{E}[1_{A \cap \{T=k\}} X_k] = \mathbf{E}[1_A X_T].$$

Ceci conclut par (PC').

■

Théorème 75. *[Théorème d'arrêt optionnel de Doob] Si (X_n) est une sur-martingale et T un temps d'arrêt. On suppose l'une des hypothèses suivantes :*

1. *T est borné*
2. *$\mathbf{E}(T) < \infty$ et il existe $C > 0$, tel que pour tout n*

$$\mathbf{E}(|X_{n+1} - X_n| \; |\mathcal{F}_n) \leq C.$$

3. *(X_n) est positive*

Alors $1_{\{T<\infty\}} X_T = \sum_{k=0}^{\infty} X_k 1_{\{T=k\}} \in L^1(\Omega)$ et on a $\mathbf{E}[1_{\{T<\infty\}} X_T] \leq \mathbf{E}(X_0)$.

Preuve : Le cas 1. est le premier théorème d'arrêt. En particulier dans tous les cas $\mathbf{E}[X_{n\wedge T}] \leq \mathbf{E}[X_0]$

Dans le cas 2. on trouve une condition de domination

$$|X_{n\wedge T}| \leq M := |X_0| + \sum_{k=1}^{T-1} |X_{k+1} - X_k| = |X_0| + \sum_{k=1}^{\infty} |X_{k+1} - X_k| 1_{T>k}.$$

On note en utilisant Fubini-Tonelli et (PC) :

$\mathbf{E}[M] = \mathbf{E}[|X_0|] + \sum_{k=1}^{\infty} \mathbf{E}[\mathbf{E}[|X_{k+1} - X_k||\mathcal{F}_k] 1_{T>k}] \leq \mathbf{E}[|X_0|] + C\sum_{k=1}^{\infty} \mathbf{E}[1_{T>k}] = \mathbf{E}[|X_0|] + C\mathbf{E}(T) < \infty$.

On a donc la condition de domination pour appliquer TCD à $X_{n\wedge T} \to 1_{\{T<\infty\}} X_T$ on obtient $1_{\{T<\infty\}} X_T \in L^1$ et l'égalité $E[1_{\{T<\infty\}} X_T] = \lim_n E[X_{n\wedge T}] \leq E[X_0]$.

Dans le cas 3., on peut appliquer Fatou à $1_{\{T<\infty\}} X_T = \lim_{n\to\infty} X_{n\wedge T} \geq 0$ de sorte que l'on obtient :

$$\mathbf{E}[X_T] = \mathbf{E}[\liminf_n X_{n\wedge T}] \leq \liminf_n \mathbf{E}[X_{n\wedge T}] \leq \mathbf{E}[X_0].$$

■

Exemple 48. Soit $X_n = Z_1 + ... + Z_n$ une marche aléatoire avec $Z_i \in L^1$ i.i.d. Soit T un temps d'arrêt pour la filtration engendrée par le processus. On suppose $\mathbf{E}(T) < \infty$, alors on a (en utilisant le cas 2. du théorème) l'équation de Wald :

$$\mathbf{E}(X_T) = \mathbf{E}(T)\mathbf{E}(X_1)$$

Exemple 49. Soit Y_n la martingale de l'exemple 43 dans le cas $r = 0$, c'est une surmartingale positive donc on peut appliquer 3. Quelle que soit le temps d'arrêt (même non borné) T, on trouve $\mathbf{E}(Y_T) \leq \mathbf{E}(Y_0)$. Il n'y a donc pas de façon d'être sûr de gagner plus en vendant à un temps aléatoire (qui ne prévoit pas le futur) cet actif financier.

Exemple 50. Soit $X_n = Y_0^{-1}(1+r)^{-n} Y_n$ la martingale de l'exemple 43 dans le cas $r \geq 0$ avec $Y_0 = y_0 \geq 0$ et $a > 1$. Soit $T = \inf\{n \in \mathbb{N}, X_n = a\}$. C'est le premier temps où Y_n vaut $ay_0(1+r)^n$ et donc le premier temps où l'actif risqué a battu d'un rapport a l'actif sans risque.

Montrons par l'absurde que $E(T) = \infty$.

Autrement dit, si on attend cet objectif pour vendre l'actif, le temps d'attente nécessaire sera en moyenne infini.

On regarde $\ln(X_n) = \sum_{k=1}^{n} \ln(1 + \frac{\sigma\epsilon_n}{1+r})$ est une marche aléatoire et

$$\mathbf{E}[|\ln(X_1)|] = \frac{1}{2}\left(\ln(1 + \frac{\sigma}{1+r}) + |\ln(1 - \frac{\sigma}{1+r})|\right) = \frac{1}{2}\left(\ln(\frac{1+r+\sigma}{1+r-\sigma})\right) < \infty.$$

On note que $\mathbf{E}[(\ln(X_1)] = \frac{1}{2}\ln(1 - \frac{\sigma^2}{(1+r)^2}) < 0$ car $|\sigma| < 1+r$. Donc on pourrait appliquer l'équation de Wald et déduire

$$\mathbf{E}[T]\mathbf{E}[\ln(X_1)] = \mathbf{E}[\ln(X_T)] = \ln(a) > 0$$

C'est la contradiction cherchée.

Chapitre 5

Introduction au mouvement Brownien et au modèle de Black et Scholes

1 Définition et propriétés fondamentales du mouvement Brownien

Définition 39. Un processus gaussien centré $(B_t)_{t\in\mathbb{R}^+}, B_t : (\Omega, \mathcal{T}, P) \to \mathbb{R}$ est un mouvement brownien (standard) si il existe $A \in \mathcal{T}$ avec $P(A) = 1$ tel que pour tout $\omega \in A$, $t \mapsto B_t(\omega)$ est continue et si il est de covariance $\mathbf{E}(B_s B_t) = \min(s, t)$.

Remarque 11. On rappelle que, pour vérifier qu'un processus est gaussien, il faut vérifier que $(B_{t_1}, ..., B_{t_n})$ est un vecteur gaussien pour tout $(t_1, ..., t_n)$. Pour vérifier qu'il est centré, on doit vérifier $\mathbf{E}(B_t) = 0$. Il n'est pas évident que le mouvement brownien existe (c'est-à-dire que l'on peut construire le processus gaussien ci-dessus avec toutes ses trajectoires continues. On l'admettra.

Par ailleurs, certains appellent $x + \mu t + \sigma B_t$ aussi mouvement brownien (non standard), partant de x, avec dérive μ et variance σt. Nous n'utiliserons que des mouvements brownien standards.

Théorème 76. *Un processus $(B_t)_{t\in\mathbb{R}^+}$ est un mouvement brownien si et seulement si il vérifie*
- *(continuité p.s.) Il existe $A \in \mathcal{T}$ avec $P(A) = 1$ tel que pour tout $\omega \in A$, $t \mapsto B_t(\omega)$ est continue*
- *(accroissements indépendants) Pour tout n et tout $t_1 < ... < t_n$ $B_{t_1}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}}$ sont indépendants*
- *(Accroissements gaussiens stationnaires) Pour tout $s < t$ $B_t - B_s \sim \mathcal{N}(0, t - s)$.*

Preuve : Si B_t vérifie les hypothèses, $(B_{t_1}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}})$ est un vecteur gaussien (suite de variables indépendantes gaussiennes), donc par transformation linéaire $(B_{t_1}, B_{t_2}, ..., B_{t_n})$ est un vecteur gaussien, donc (B_t) est un processus gaussien (centré car chaque B_t est supposé centré). Comme en plus il est continue, c'est un mouvement brownien si il a la bonne covariance, et en effet, si $s < t$

$$\mathbf{E}(B_t B_s) = \mathbf{E}((B_t - B_s)B_s) + \mathbf{E}(B_s^2) = \mathbf{E}(B_t - B_s)\mathbf{E}(B_s) + s = s = \min(s, t)$$

en utilisant la variance et l'indépendance. Par échange de s et t, on a le cas général.

Réciproquement, si (B_t) est un mouvement brownien, il est supposé *p.s.* continue et chaque $(B_{t_1}, B_{t_2}, ..., B_{t_n})$ est un vecteur gaussien donc par combinaison linéaire $(B_{t_1}, B_{t_2} - B_{t_1}, ..., B_{t_n} - B_{t_{n-1}})$ est un vecteur gaussien, et il suffit de calculer les covariances nulles pour obtenir l'indépendance.

Or pour $i < n$, on a

$$\begin{aligned}\mathbf{E}((B_{t_i} - B_{t_{i-1}})(B_{t_n} - B_{t_{n-1}})) &= \mathbf{E}(B_{t_i}B_{t_n} - B_{t_{i-1}}B_{t_n} - B_{t_i}B_{t_{n-1}} + B_{t_{i-1}}B_{t_{n-1}})\\ &= \min(t_i, t_n) - \min(t_{i-1}, t_n) - \min(t_i, t_{n-1}) + \min(t_{i-1}, t_{n-1})\\ &= t_i - t_{i-1} - t_i + t_{i-1} = 0\end{aligned}$$

De même $B_t - B_s$ est une variable gaussienne et il ne reste qu'à calculer la variance pour $t > s$.

$$\begin{aligned}\mathbf{E}((B_t - B_s)^2) &= \mathbf{E}(B_t^2 - 2B_sB_t + B_s^2)\\ &= t - 2s + s = (t - s).\end{aligned}$$

On a donc obtenu toutes les propriétés. ■

Proposition 77. *(Invariance) Si B_t est un mouvement brownien, alors pour $\lambda > 0$, $(\lambda^{-1}B_{\lambda^2 t})_{t\geq 0}$ et $(B_t - B_\lambda)_{t\geq\lambda}$ sont des mouvements browniens.*

Preuve : La continuité est évidentes et par combinaison linéaire on a bien des processus gaussiens centrés, il suffit donc de vérifier la covariance pour le premier $\mathbf{E}(\lambda^{-1}B_{\lambda^2 t}\lambda^{-1}B_{\lambda^2 s}) = \lambda^{-2}\min(\lambda^2 t, \lambda^2 s) = \min(t, s)$. Pour le deuxième on utilise le théorème, comme les incréments $(B_t - B_\lambda)_{t\geq\lambda}$ sont des incréments de B_t on trouve l'indépendance et la stationnarité, comme voulu. ■

1.1 Limite de marches aléatoires

Soit $S_n = X_1 + ... + X_n$ avec X_i i.i.d, $\mathbf{E}(X_1) = 0, \mathbf{E}(X_1^2) = 1$. On s'intéresse au chemin $k \to S_k$ et on introduit pour cela une normalisation similaire au TCL pour se ramener à $[0, 1]$.

$$S_t^n = \frac{1}{\sqrt{n}}S_{\lfloor nt\rfloor}, t \in [0, 1].$$

S_1^n est la variable du TCL qui tend vers une variable gaussienne. Le processus tend vers un processus gaussien qui n'est rien d'autre que le mouvement brownien (en loi fini-dimensionnelle) :

Théorème 78. *(Donsker) Pour tout $(t_1, ..., t_m)$, $(S_{t_1}^n, ..., S_{t_m}^n)$ converge en loi vers $(B_{t_1}, ..., B_{t_m})$ avec $(B_t)_{t>0}$ un mouvement brownien.*

En fait le meilleur théorème de Donsker établit qu'une variante continue linéaire par morceau converge en un sens plus fort (en loi sur l'espace des fonctions continues).
Preuve : Il est équivalent de considérer la convergence en loi des incréments (changement de variable linéaire) soit pour $t_1 < ... < t_m$ $(S_{t_1}^n, S_{t_2}^n - S_{t_1}^n, ..., S_{t_m}^n - S_{t_{m-1}}^n)$. Or par regroupement par paquet ces variables sont indépendantes car $S_{t_m}^n - S_{t_{m-1}}^n \in \sigma(X_i, i \in [\lfloor t_{m-1}n\rfloor + 1, \lfloor t_m n\rfloor])$ qui donnent des familles disjointes d'indices donc indépendance. De plus $S_{t_m}^n - S_{t_{m-1}}^n$ a même loi que $\frac{\sqrt{\lfloor t_m n\rfloor - \lfloor t_{m-1}n\rfloor}}{\sqrt{n}}S_1^{(\lfloor t_m n\rfloor - \lfloor t_{m-1}n\rfloor)}$ qui converge donc en loi par le TCL vers une $\mathcal{N}(0, t_m - t_{m-1})$. En utilisant l'indépendance (et par exemple les fonctions caractéristiques), on obtient que $(S_{t_1}^n, S_{t_2}^n - S_{t_1}^n, ..., S_{t_m}^n - S_{t_{m-1}}^n)$ converge en loi vers des variables indépendantes ayant même loi que le mouvement brownien. ■

1.2 Quelques faits géométriques [facultatifs]

Le théorème suivant admis donne une idée de l'allure des trajectoires du mouvement brownien

Théorème 79. *Soit* $(B_t)_{t>0}$ *un mouvement brownien, alors*

1. *Pour tout* $\epsilon > 0$, *p.s.* $\sup_{s\in]0,\epsilon[} B_s > 0$ *et* $\inf_{s\in]0,\epsilon[} B_s > 0$,
2. $\sup_{s\in]0,t[} B_s$ *a même loi que* $|B_t|$,
3. *On a p.s.* $\limsup_{s\to\infty} B_s = +\infty$ *et* $\liminf_{s\to\infty} B_s = -\infty$
4. *L'ensemble des trajectoires nulle part dérivables contient un évènement de probabilité* 1.

2 Modèle de Black et Scholes

Définition 40. Un *mouvement Brownien géométrique* de taux d'intérêt μ et de volatilité σ est le processus $(S_t)_{t>0}$ donné par la formule :

$$S_t = S_0 exp\left((\mu - \frac{\sigma^2}{2})t + \sigma B_t\right),$$

avec $(B_t)_{t>0}$ un mouvement brownien.

C'est le modèle d'action proposé par Black et Scholes.

Remarque 12. Le modèle de Black et Scholes n'est qu'un modèle financier parmi d'autres, particulièrement simple. On pourrait facilement intégrer un taux d'intérêt et une volatilité dépendant du temps pour améliorer le modèle. Plus fondamentalement, on lui reproche souvent que les trajectoires du mouvement brownien géométrique sont continues alors que les marchés financiers ont des sauts (nécessité de processus de Poisson/Lévy). On reproche aussi dans la même ligne que les trajectoires du mouvement brownien sont trop régulières pour représenter qualitativement les marchés financiers.

La caractérisation du mouvement brownien permet de trouver une propriété caractéristique plus parlante probabilistiquement :

Proposition 80. *Un processus positif* $(S_t)_{t\in I\!R^+}$ *est un mouvement brownien géométrique si et seulement si il vérifie*

- *(continuité p.s.) Il existe* $A \in \mathcal{T}$ *avec* $P(A) = 1$ *tel que pour tout* $\omega \in A$, $t \mapsto S_t(\omega)$ *est continue*
- *(accroissements relatifs indépendants) Pour tout* $u < t$, S_t/S_u *est indépendant de la tribu* $\sigma(S_v; v \leq u)$.
- *(Accroissements stationnaires log-normaux) Pour tout* $u < t$, S_t/S_u *est de loi log normale* $\mathcal{LN}(\mu - \frac{\sigma^2}{2}, \sigma)$, *c'est-à-dire de loi :*

$$P_{S_t/S_u}(ds) = \frac{1}{s\sigma\sqrt{2\pi(t-u)}} exp\left(-\frac{(\ln(s) - \ln((\mu - \frac{\sigma^2}{2})(t-u))^2}{2\sigma^2(t-u)}\right) 1_{]0,\infty[}(s)ds.$$

PREUVE : Il est facile de voir que les hypothèses se traduisent pour $\frac{1}{\sigma}\left(\ln(S_t/S_0) - (\mu - \frac{\sigma^2}{2})t\right)$ sur les hypothèses du théorème 76 et donc qu'il s'agit d'un mouvement brownien (standard) B_t. ■

Le théorème suivant va être utilisé pour calculer le prix des options associés à ce modèle. Le modèle de Black et Scholes suppose donné un actif sans risque de prix à l'instant t de taux d'intérêt (instantané) r :

$$S_t^0 = exp(rt).$$

On a le cas particulier suivant du théorème de Girsanov (de changement de mesure des processus stochastiques) :

Théorème 81 (Girsanov avec dérive constante)**.** *Soit $(B_t)_{t\in[0,T]}$ un mouvement brownien (standard) dans $(\Omega, \mathcal{T}, P)$. Soit $\mathcal{F}_T = \sigma(B_t, t \leq T)$. Alors la formule suivante définit une mesure de probabilité sur $(\Omega, \mathcal{F}_T)$*

$$\widetilde{P}(A) = \mathbf{E}(1_A exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_T - \frac{T}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right)).$$

Dans $(\Omega, \mathcal{F}_T, \widetilde{P})$, $W_t = B_t + \left(\frac{\mu - r}{\sigma}\right) t$ est un mouvement brownien (standard) sur $[0, T]$.

Preuve : On commence par montrer que l'on a bien une probabilité. Comme

$$L_T := exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_T - \frac{T}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right)$$

est positive, on a bien la positivité, la sigma-additivité (par convergence monotone), il reste à noter que $\widetilde{P}(\Omega) = 1$ ce qui revient à

$$\mathbf{E}(exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_T\right)) = exp\left(\frac{T}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right)$$

C'est le calcul fait dans la preuve du lemme 12 pour la transformée de Laplace d'une variable gaussienne.

Pour montrer que W_t est un mouvement brownien, on va vérifier les hypothèses du théorème 76. On va utiliser souvent la propriété suivante (intégration par partie gaussienne) d'une variable Y gaussienne $\mathcal{N}(0, \sigma^2)$, pour tout f C^1 telle que $Yf(Y), f'(Y) \in L^1(\Omega)$, on a (facile à vérifier après transfert par IPP) :

$$\mathbf{E}(Yf(Y)) = \sigma^2 \mathbf{E}(f'(Y)). \tag{5.1}$$

On remarque que la continuité p.s. est évidente avec le même A.

On remarque d'abord pour f mesurable bornée ou $f(x) = x^k$ $k = 1, 2$ (d'abord $k = 2$ pour obtenir intégrabilité), par indépendance des accroissements du mouvement brownien :

$$\begin{aligned}\mathbf{E}(f(B_t) exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_T\right)) &= \mathbf{E}(f(B_t) exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right) exp\left(-\left(\frac{\mu - r}{\sigma}\right)(B_T - B_t)\right)) \\ &= \mathbf{E}(f(B_t) exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right))\mathbf{E}(exp\left(-\left(\frac{\mu - r}{\sigma}\right)(B_T - B_t)\right))\end{aligned}$$

donc pour $t \leq T$, $\mathbf{E}(f(B_t)L_T) = \mathbf{E}(f(B_t)L_t)$. (En fait on peut voir que L_t est une martingale à temps continu, du type des martingales exponentielles de l'exemple 40). Il est alors facile de voir que $W_t \in L^2(\Omega, \widetilde{P})$, car $B_t^2 exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right) \in L^1(\Omega)$ par une borne par une fonction $exp(cB_t) + exp(-cB_t) \in L^1(\Omega)$.

-**Calcul de l'espérance**.

On utilise (5.1) pour obtenir :

$$\mathbf{E}(B_t exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)) = -\left(\frac{\mu - r}{\sigma}\right) t\mathbf{E}(exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right))$$

On déduit alors du calcul utilisant l'indépendance ci-dessus que

$$\widetilde{\mathbf{E}}(W_t) = \mathbf{E}((B_t + \left(\frac{\mu - r}{\sigma}\right) t)L_T) = \mathbf{E}((B_t + \left(\frac{\mu - r}{\sigma}\right) t)L_t) = \mathbf{E}(B_t L_t) + \left(\frac{\mu - r}{\sigma}\right) t = 0.$$

-Calcul de la variance.
On utilise (5.1) (puis le calcul précédent) pour obtenir :

$$\begin{aligned}\mathbf{E}(B_t^2 exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)) &= t\mathbf{E}(exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)) - \left(\frac{\mu - r}{\sigma}\right) t\mathbf{E}(B_t exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)) \\ &= [t + \left(\frac{\mu - r}{\sigma}\right)^2 t^2]\mathbf{E}(exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)).\end{aligned}$$

On déduit alors du calcul utilisant l'indépendance ci-dessus et du calcul de l'espérance que

$$\widetilde{\mathbf{E}}(W_t^2) = \mathbf{E}(B_t^2 L_t) + 2\mathbf{E}(B_t L_t)\left(\frac{\mu - r}{\sigma}\right) t + \left(\frac{\mu - r}{\sigma}\right)^2 t^2 = \mathbf{E}(B_t^2 L_t) - \left(\frac{\mu - r}{\sigma}\right)^2 t^2 = t.$$

-Calcul de transformées de Fourier pour obtenir les incréments gaussiens. En fait, cette étape suffit car on va recalculer la variance et la moyenne avec la transformée de Fourier. On utilise juste la preuve du lemme 12 :

$$\begin{aligned}\mathbf{E}(exp(isB_t)exp\left(-\left(\frac{\mu - r}{\sigma}\right) B_t\right)) &= exp\left(\frac{t}{2}\left(is - \frac{\mu - r}{\sigma}\right)^2\right) \\ &= exp\left(-\frac{ts^2}{2} + \frac{t}{2}\left(\frac{\mu - r}{\sigma}\right)^2 - ist\left(\frac{\mu - r}{\sigma}\right)\right)\end{aligned}$$

d'où la valeur attendue :

$$\widetilde{\mathbf{E}}(exp(isW_t)) = \mathbf{E}(exp(isB_t + ist\left(\frac{\mu - r}{\sigma}\right))L_t) = exp\left(-\frac{ts^2}{2}\right).$$

-Indépendance et stationnarité des incréments
Pour tout n et tout $t_1 < ... < t_n$, montrons que $W_{t_1}, W_{t_2} - W_{t_1}, ..., W_{t_n} - W_{t_{n-1}}$ sont indépendants. Prenons donc aussi $f_1, ...f_n$ mesurables bornées (par exemple des exponentielles donnant des fonctions caractéristiques). On note $\theta = \frac{\mu - r}{\sigma}$ et on obtient en utilisant l'indépendance des incréments de B_t sous P :

$$\begin{aligned}&\widetilde{\mathbf{E}}\left(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})...f_n(W_{t_n} - W_{t_{n-1}})\right) = \mathbf{E}\left(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})...f_n(W_{t_n} - W_{t_{n-1}})L_{t_n}\right) \\ &= \mathbf{E}\left(f_1(W_{t_1})f_2(W_{t_2} - W_{t_1})...f_n(W_{t_n} - W_{t_{n-1}})L_{t_1}\prod_{k=2}^{n} exp\left(-\theta(B_{t_k} - B_{t_{k-1}}) - \frac{t_k - t_{k-1}}{2}\theta^2\right)\right) \\ &= \mathbf{E}\left(f_1(W_{t_1})L_{t_1}\right)\prod_{k=2}^{n}\mathbf{E}\left(f_k(W_{t_k - t_{k-1}})exp\left(-\theta B_{t_k - t_{k-1}} - \frac{t_k - t_{k-1}}{2}\theta^2\right)\right) \\ &= \widetilde{\mathbf{E}}\left(f_1(W_{t_1})\right)\widetilde{\mathbf{E}}\left(f_2(W_{t_2} - W_{t_1})\right)...\widetilde{\mathbf{E}}\left(f_n(W_{t_n} - W_{t_{n-1}})\right)\end{aligned}$$

la dernière ligne étant obtenu en prenant tous les f_i égaux à 1 sauf un pour identifier les termes. La stationnarité des incréments se déduit du même calcul en utilisant celle de B_t. ■

En finance, un produit dérivé (ou option) est un contrat sur l'actif financier S_t qui produit un paiement positif H, $\mathcal{F}_T$ mesurable au temps d'échéance T. Nous ne considérerons que les options dites européennes de la forme $H = \Phi(S_T)$ avec Φ mesurable positive. La définition suivante (motivée dans la section suivante) permet de donner une valeur à l'option à l'instant t (même si l'option n'est pas coté, et permettant de détecter des "anomalies" de marché (par exemple possibilités d'arbitrages) si la cotation ne correspond pas à la valeur théorique).

Définition 41. La valeur au temps $t \leq T$ de l'option, définie par la variable aléatoire $\Phi(S_T)$ d'échéance T, est donnée par la formule utilisant la probabilité $\widetilde{P}$ du théorème 81 par :

$$V_t = \widetilde{\mathbf{E}}(e^{-r(T-t)}\Phi(S_T)|\mathcal{F}_t).$$

Le prix à l'instant $t = 0$ est la valeur de vente conseillée (par la théorie) dans ce modèle pour l'option. On va obtenir le calcul explicite dans le cas des options de type call $\Phi(x) = \max(x-K, 0) =: (x-K)_+$ et put $\Phi(x) = (K-x)_+$.

On note la fonction de répartition d'une gaussienne standard

$$N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{d} e^{-\frac{x^2}{2}} dx$$

et on pose

$$\rho_\pm(t,x) = \frac{\ln(\frac{x}{K}) + (r \pm \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}.$$

On remarquera que $\rho_+(t,x) = \rho_-(t,x) + \sigma\sqrt{T-t}$.

Théorème 82. *(valeur des put et call de Black-Scholes) La valeur au temps $t \leq T$ de l'option de type call $(S_T - K)_+$ d'échéance T est donné par $V_t = F_c(t, S_t)$ avec :*

$$F_c(t,x) = xN(\rho_+(t,x)) - Ke^{-r(T-t)}N(\rho_-(t,x)).$$

La valeur au temps $t \leq T$ de l'option de type put $(K - S_T)_+$ d'échéance T est donné par $V_t = F_p(t, S_t)$ avec :

$$F_p(t,x) = Ke^{-r(T-t)}N(-\rho_-(t,x)) - xN(-\rho_+(t,x)).$$

On remarquera que ces valeurs sont indépendantes de μ (grâce au changement de probabilité) et ne dépendent que de la seule volatilité σ.

Preuve : On note d'abord qu'avec les notations du Théorème de Girsanov, la valeur actualisée de l'actif est par calcul élémentaire $\widetilde{S}_t := S_t e^{-rt} = S_0 exp(\sigma W_t - \sigma^2 t/2)$ et donc

$$S_T = S_t e^{r(T-t)} exp(\sigma(W_T - W_t) - \sigma^2(T-t)/2).$$

Donc $\widetilde{\mathbf{E}}(S_T - K|\mathcal{F}_t) = S_t e^{r(T-t)}\mathbf{E}(exp(\sigma(W_T - W_t) - \sigma^2(T-t)/2)) - K = S_t e^{r(T-t)} - K = F(t, S_t)$ avec $F(t,x) = xe^{r(T-t)} - K$. Puisque $(S_T - K)_+ - (K - S_T)_+ = S_T - K$ on déduit donc, après multiplication par $e^{-r(T-t)}$ la relation que nos formules vérifient :

$$F_c(t,x) - F_p(t,x) = x - Ke^{-r(T-t)}.$$

On considère donc le cas du call. Donc avec $Y = \frac{W_T - W_t}{\sqrt{T-t}}$ on obtient par les calculs précédents avec $\theta := \sigma\sqrt{T-t}$:

$$e^{-r(T-t)}(S_T - K)_+ = (S_t exp(\theta Y - \theta^2) - Ke^{-r(T-t)})_+ = (S_t exp(\theta Y - \theta^2/2) - Ke^{-r(T-t)})1_{\{Y + \rho_-(t,S_t) \geq 0\}}.$$

En effet, on a noté $S_t exp(\theta Y - \theta^2/2) - Ke^{-r(T-t)}$ est positif si et seulement si $\ln(S_t/K) \geq -r(T-t) - (\theta Y - \theta^2/2)$ soit $\ln(S_t/K) + r(T-t) - \theta^2/2 + \theta Y \geq 0$ ou encore $Y + \rho_-(t, S_t) \geq 0$.

En prenant l'espérance conditionnelle sur $\mathcal{F}_t$ en utilisant l'indépendance des accroissements de W_t sous $\widetilde{P}$ c'est à dire celle de Y avec $\mathcal{F}_t$ et la proposition 60, on obtient $V_t = F_c(t, S_t)$ avec

$$F_c(t,x) = x\widetilde{\mathbf{E}}(exp(\theta Y - \theta^2/2)1_{\{Y+\rho_-(t,x)\geq 0\}}) - Ke^{-r(T-t)}\widetilde{\mathbf{E}}(1_{\{Y+\rho_-(t,x)\geq 0\}}).$$

Comme $Y \sim \mathcal{N}(0,1)$ sous $\widetilde{P}$ on a évidemment

$$\widetilde{\mathbf{E}}(1_{\{Y+\rho_-(t,x)\geq 0\}}) = \widetilde{P}(Y \geq -\rho_-(t,x)) = \widetilde{P}(Y \leq \rho_-(t,x)) = N(\rho_-(t,x)).$$

De même, on a la relation conclusive (obtenue par 2 transferts et changement de variable $y - \theta$) :

$$\widetilde{\mathbf{E}}(exp(\theta Y - \theta^2/2)1_{\{Y+\rho_-(t,x)\geq 0\}}) = \int_{-\rho_-(t,x)}^{\infty} exp(-(y-\theta)^2/2)\frac{dy}{\sqrt{2\pi}} = \widetilde{\mathbf{E}}(1_{\{Y+\theta+\rho_-(t,x)\geq 0\}}) = N(\rho_+(t,x)).$$

■

3 Motivation de l'intégrale stochastique pour l'étude du modèle de Black et Scholes

Le but de cette section sans preuve est de motiver l'utilisation d'une notion d'intégrale stochastique pour l'interprétation du prix défini ci-dessus dans le modèle de Black et Scholes.

Il existe une notion d'intégrale stochastique qui permet d'écrire une formule intégrale pour $f(t, B_t)$ si $f : (t,x) \mapsto f(t,x)$ est C^2 (ou même seulement C^1 en t). On obtient alors la formule d'Ito :

$$f(t,B_t) = f(0,0) + \int_0^t (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2})(s, B_s)ds + \int_0^t \frac{\partial f}{\partial x}(s, B_s)dB_s$$

Le changement notable par rapport au cas dérivable en temps (on rappelle que B_t est p.s. nulle part dérivable), c'est le terme faisant intervenir la dérivée seconde de f. Par exemple, on obtient

$$S_t = S_0 + \int_0^t \mu S_s ds + \int_0^t \sigma S_s dB_s.$$

En écriture infinitésimale $dS_t = S_t(\mu dt + \sigma dB_t)$ qui est très semblable au modèle discret de l'exemple 43 $Y_n - Y_{n-1} = Y_{n-1}(r + \sigma\epsilon_n)$. On peut manipuler ces expressions en utilisant les "règles"

$$dt^2 = 0, dtdB_t = 0, dB_t^2 = dt.$$

C'est un théorème qu'une intégrale stochastique $\int_0^t f_s dB_s$ est définie pour f_s adaptée dans $L^2([0,T], L^2(\Omega, P)) = L^2([0,T] \times \Omega, dt \otimes P)$ par l'isométrie d'Ito :

$$\mathbf{E}[|\int_0^t f_s dB_s|^2] = \int_0^t \mathbf{E}[|f(s)|^2]dt.$$

C'est souvent une "martingale" sur [0,T] si f est suffisamment borné (sinon c'est une "martingale locale").

Une stratégie (ou portefeuille) pour le couple (S_t^0, S_t) est une paire de valeurs $\phi_t = (\alpha_t, \beta_t)$ représentant les valeurs achetées de l'actif sans risque S_t^0 et de l'actif risqué S_t au temps t. La valeur du portefeuille est donc :

$$V_t(\phi) = \alpha_t S_t^0 + \beta_t S_t$$

Celui-ci est auto-financé si (avec des hypothèses appropriées sur ϕ_t pour que les intégrales aient un sens)

$$V_t(\phi) = V_0(\phi) + \int_0^t \alpha_s dS_s^0 + \int_0^t \beta_s dS_s$$

qui revient à demander, dans l'intervalle $[t, t+dt]$, que la variation du portefeuille $dV_t(\phi) = \alpha_s dS_t^0 + \beta_s dS_t$ ne viennent que des variations des valeurs des actifs et pas d'apports au portefeuille ($d\alpha_s$,$d\beta_s$). On suppose en particulier $(\alpha_t), (\beta_t)$ adaptés à la filtration $\mathcal{F}_t$. Si on regarde les valeurs actualisées $\widetilde{S}_t = (S_t^0)^{-1} S_t, \widetilde{V}_t(\phi) = (S_t^0)^{-1} V_t(\phi) = \alpha_t + \beta_t \widetilde{S}_t$, on a les relations (pour un portefeuille autofinancé) $d\widetilde{S}_t = \widetilde{S}_t[(\mu - r)dt + \sigma dB_t]$, et

$$d\widetilde{V}_t(\phi) = -re^{-rt} dt V_t(\phi) + e^{-rt} dV_t(\phi) = -r(\alpha_t + \beta_t \widetilde{S}_t)dt + r\alpha_t dt + \beta_t e^{-rt} dS_t = \beta_t d\widetilde{S}_t$$

si bien que la valeur de α_t est déterminée par

$$\alpha_t = V_t(\phi) - \beta_t \widetilde{S}_t = V_0(\phi) + \int_0^t \beta_t d\widetilde{S}_t - \beta_t \widetilde{S}_t$$

On retrouve la condition $\mu = r$ pour que $\widetilde{S}_t$ et $\widetilde{V}_t(\phi)$ soient des martingales (locales) sous P (similaires à l'exemple 43, ce sont toujours des martingales locales sous $\widetilde{P}$).

On peut montrer que si $V_t(\phi) \geq 0$ (on dit admissible) alors il n'y a pas de possibilité d'arbitrage, c'est à dire de stratégies auto-financées avec $V_0 = 0, V_T \geq 0$ et $P(V_T > 0) > 0$.

Une option $\Phi(S_T)$ est dite répliquée si il existe un portefeuille autofinancée tel que $V_T(\phi) = \Phi(S_T)$. La valeur de l'option (au sens du paragraphe précédent). $V_t = V_t(\phi)$ est la valeur d'un tel portefeuille autofinancé au temps t

On peut montrer l'existence du portefeuille sous l'hypothèse $\mathbf{E}(L_T^2 \Phi(S_T)^2) < \infty$. Elle nécessite des outils de calculs stochastiques dépassant la portée de ce cours (et la condition $\mathbf{E}(L_T^2 H^2) = \widetilde{\mathbf{E}}(H^2) < \infty$ fonctionne pour des options non-européennes $H \neq \Phi(S_T)$ mais la solution est alors moins explicite).

Le calcul du portefeuille simulant est utile en finance pour la “couverture de l'option" (par le vendeur vendant au prix $V_0(\phi)$ qui utilisera cette stratégie pour produire une richesse $V_T(\phi)$ au temps T et qu'il doit donner à l'acheteur). Donc le cas du put et du call, on a les formules

$$\alpha_t = e^{-rt}(F_{p/c}(t, S_t) - \beta_t S_t), \quad \beta_t = \frac{\partial F_{p/c}}{\partial x}(t, S_t).$$

Pour le call on a $\beta_t = N(\rho_+(t, S_t))$ et pour le put $\beta_t = -N(-\rho_+(t, S_t))$.

Annexe A

Rappel sur le théorème de Radon-Nikodym et Théorème de Dunford-Pettis

Ce complément pourrait pour l'essentiel être ajouté au chapitre 4.1. Nous expliquons un théorème de théorie de la mesure qui permet de dire quand une mesure provient d'une densité dans $L^1(\Omega, \mu)$. On en déduit une application à un théorème de compacité que nous avons utilisé pour la preuve du cas uniformément continue du théorème de convergence des martingale, le théorème de Dunford-Pettis 84.

Définition 42. Si μ, ν sont des mesures de probabilités sur $(\Omega, \mathcal{T})$, on dit que μ est absolument continue par rapport à ν et on note $\mu \ll \nu$ si pour tout $A \in \mathcal{T}$, $\nu(A) = 0$ implique que $\mu(A) = 0$

Définition 43. Si μ, ν sont des mesures de probabilités sur $(\Omega, \mathcal{T})$, on dit que μ admet une densité $h \in L^1(\Omega, \nu)$ par rapport à ν et on note $h = \frac{d\mu}{d\nu}$, si $h \geq 0$ p.s. et pour tout $A \in \mathcal{T}$:

$$\int_\Omega 1_A h d\nu = \mu(A).$$

Les définitions s'étendent aux mesures σ-finies, mais on considère seulement ici le cas de probabilités.

Théorème 83 (de Radon-Nikodym)**.** *Pour toutes mesures de probabilités μ, ν sur $(\Omega, \mathcal{T})$, il y a équivalence entre $\mu \ll \nu$ et l'existence d'une densité $h = \frac{d\mu}{d\nu} \in L^1(\Omega, \nu)$ de μ par rapport à ν, et la densité est alors unique ν-p.s.*

PREUVE : Si on a deux densités h, k, $\int_\Omega 1_A(h-k)d\nu = 0$ pour tout A $\mathcal{T}$ mesurable, donc par la construction de l'intégrale aussi $\int_\Omega fhd\nu = \int_\Omega fkd\nu$ d'abord pour f mesurable positive (par TCM) puis pour f mesurable bornée donc par dualité $h - k = 0$ dans $L^1(\Omega, \nu)$ donc ν-p.s.

De plus, si on a existence d'une densité et si $\nu(A) = 0$, par TCM, $\int_\Omega 1_A h = \lim_{n\to\infty} \int_\Omega 1_A(h \wedge n) = 0$ car $|\int_\Omega 1_A(h \wedge n)d\nu| \leq ||(h \wedge n)|||_2||1_A||_2 \leq n\nu(A)^{1/2} = 0$ par Cauchy-Schwartz. Donc $\mu(A) = 0$ c'est à dire on a montré $\mu \ll \nu$.

La partie difficile est l'existence d'une densité si $\mu \ll \nu$. On va utiliser le théorème de représentation de Riesz (ou sa variante pour la dualité de L^1, le théorème 51). Soit $\mu_\alpha = \mu + \alpha\nu$ avec $\alpha > 0$. L'idée est simple on s'attend à avoir une densité $\frac{d\mu_\alpha}{d\nu} = \alpha + h$ strictement positive et donc $\frac{d\nu}{d\mu_\alpha} = \frac{1}{\alpha+h}$ bornée par $1/\alpha$ donc dans L^2 ensuite $\alpha(1 - \frac{\alpha}{\alpha+h}) = \alpha\frac{h}{\alpha+h} \to_{\alpha\to\infty} h$ et on devrait pouvoir retrouver h ainsi.

Appliquons cette idée, si $f \in L^1(\Omega, d\mu_\alpha)$, on a

$$\int |f| d\nu = \frac{1}{\alpha} \int |f| d\alpha\nu \leq \frac{1}{\alpha} \int |f| d\mu_\alpha$$

Donc $f \in L^1(\Omega, d\nu)$ et $f \mapsto \int f d\nu$ définit une forme linéaire continue sur $L^1(\Omega, d\mu_\alpha)$, donc par le théorème 51, il existe $h_\alpha \in L^\infty(\Omega, d\mu_\alpha)$ telle que pour tout $f \in L^1(\Omega, d\mu_\alpha)$ on a

$$\int f d\nu = \int f h_\alpha d\mu_\alpha.$$

Et de plus, on a $||h_\alpha||_{L^\infty(\mu_\alpha)} \leq 1/\alpha$. Si $f = 1_{\{h_\alpha < 0\}}$, on obtient $\int \max(0, h_\alpha) d\mu_\alpha \geq 0$ donc vaut 0, donc

$$\nu(\{h_\alpha < 0\}) \leq \frac{1}{\alpha} \mu_\alpha(\{h_\alpha < 0\}) = 0$$

donc $h_\alpha \geq 0$, ν p.s.

On montre maintenant la monotonie attendue pour h_α (si on veut qu'elle soit égale à un $\frac{1}{\alpha+h}$) Si $\beta > \alpha$, on a pour f positive bornée en utilisant $\mu_\alpha(g) \leq \mu_\beta(g)$ pour g positive ν-p.s,

$$\int f h_\beta d\mu_\beta = \int f d\nu = \int f h_\alpha d\mu_\alpha \leq \int f h_\alpha d\mu_\beta$$

car $f h_\alpha$ posivite ν-p.s. par le résultat précédent, donc comme c'est valable pour tout $f \geq 0$, on a $h_\beta \leq h_\alpha \mu_\beta$-p.s. donc ν-p.s.

Finalement, on a l'identité

$$\int f d\mu = \int f d\mu_\alpha - \int f \alpha d\nu = \int f(1 - \alpha h_\alpha) d\mu_\alpha = \int f \alpha (1 - \alpha h_\alpha) d\nu + \int f(1 - \alpha h_\alpha) d\mu.$$

Par $||h_\alpha||_{L^\infty(\mu_\alpha)} \leq 1/\alpha$. on a $1 - \alpha h_\alpha \geq 0$ μ_α-p.s. donc ν-p.s. En raisonnant comme avant on obtient $(1 - \alpha h_\alpha) \geq (1 - \beta h_\beta)$ ν-p.s. Donc , par l'égalité précédente (toujours pour f positive en utilisant la croissance de $\alpha \to \alpha h_\alpha$ ν-p.s. par ce qu'on vient de voir donc μ-p.s. par l'hypothèse $\mu \ll \nu$) :

$$\int f \alpha (1 - \alpha h_\alpha) d\nu = \int f \alpha h_\alpha d\mu \leq \int f \beta h_\beta d\mu = \int f \beta (1 - \beta h_\beta) d\nu$$

soit $\alpha(1 - \alpha h_\alpha) \leq \beta(1 - \beta h_\beta)$, ν-p.s. donc converge vers un h en croissant et par convergence monotone et l'égalité avant on obtient

$$\int f h d\nu = \lim_{\alpha \to \infty} \int f \alpha (1 - \alpha h_\alpha) d\nu = \lim_{\alpha \to \infty} \int f d\mu - \int f(1 - \alpha h_\alpha) d\mu \leq \int f d\mu.$$

Donc pour $f = 1$ on trouve $h \in L^1(\Omega, d\nu)$. Or par la monotonie de la limite définissant h, on a

$$(1 - \alpha h_\alpha) = \frac{\alpha(1 - \alpha h_\alpha)}{\alpha} \leq \frac{h}{\alpha} \to_{\alpha \to \infty} 0$$

ν-p.s. puisque h est fini ν-p.s. donc en utilisant encore l'hypothèse, aussi μ-p.s. Comme on a vu la monotonie en α par convergence monotone, on déduit $\int f(1 - \alpha h_\alpha) d\mu \to 0$ et donc finalement l'égalité attendue qui conclut la preuve :

$$\int f h d\nu = \lim_{\alpha \to \infty} \int f d\mu - \int f(1 - \alpha h_\alpha) d\mu = \int f d\mu.$$

■

On peut maintenant rappeler et prouver le théorème 84 :

Théorème 84 (Dunford-Pettis)**.** *Soit une suite (X_n) dans $L^1(\Omega, \mathcal{T}, P)$ avec $\mathcal{T}$ une tribu dénombrablement engendrée. On a l'équivalence entre*

1. *(X_n) est uniformément intégrable*
2. *(X_n) admet une sous-suite (X_{n_k}) ayant pour limite faible $X \in L^1$, c'est-à-dire :*
$$\forall f \in L^\infty(\Omega), \quad \mathbf{E}((X_{n_k} - X)f) \to 0.$$
3. *(X_n) est bornée dans L^1 et pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que si $A \in \mathcal{T}$ vérifie $P(A) \leq \eta$ alors pour tout n, $\mathbf{E}(1_A|X_n|) \leq \epsilon$.*

C'est surtout l'équivalence entre 1. et 2. qui est difficile et porte le nom de théorème de Dunford-Pettis. L'hypothèse "dénombrablement engendrée" n'est pas nécessaire (cf. Delacherie-Meyer *Probabilités et Potentiel* Vol 1 p 27) mais nous la faisons pour simplifier.

Preuve : On commence par l'équivalence entre 1 et 3. Supposons 3. et fixons $\epsilon > 0$, η t.q. $P(A) \leq \eta$ implique $\mathbf{E}(1_A|X_n|) \leq \epsilon$. Par l'inégalité de Markov $P(|X_n| \geq c) \leq \frac{\sup_{n \in \mathbb{N}} \mathbf{E}(|X_n|)}{c} \leq \eta$ dès que $c \geq \frac{\sup_{n \in \mathbb{N}} \mathbf{E}(|X_n|)}{\eta}$, en appliquant alors à $A = \{|X_n| \geq c\}$, on déduit $\sup_n \mathbf{E}(1_\{|X_n| \geq c\}|X_n|) \leq \epsilon$. Et donc $\lim_{c \to \infty} \mathbf{E}(1_\{|X_n| \geq c\}|X_n|) = 0$ qui est l'uniforme intégrabilité recherchée.

Réciproquement, pour $\epsilon < 0$ fixé, on prend $c > 0$ tel que $\sup_n \mathbf{E}(1_{\{|X_n| \geq c\}}|X_n|) \leq \epsilon/2$, (en particulier
$$\mathbf{E}(|X_n|) = \mathbf{E}(1_{\{|X_n| \geq c\}}|X_n|) + \mathbf{E}(1_{\{|X_n| < c\}}|X_n|) \leq c + \epsilon/2$$
donc X_n et bornée dans L^1, de sorte que
$$\mathbf{E}(1_A|X_n|) = \mathbf{E}(1_A 1_{\{|X_n| \geq c\}}|X_n|) + \mathbf{E}(1_A 1_{\{|X_n| < c\}}|X_n|) \leq \mathbf{E}(1_{\{|X_n| \geq c\}}|X_n|) + \mathbf{E}(1_A 1_{\{|X_n| < c\}} c) \leq \epsilon/2 + P(A)c$$
qui est borné par ϵ dès que $P(A) \leq \eta = \epsilon/2c$ qui convient.

On suppose maintenant 3 et on montre 2. Si $\mathcal{T} = \sigma(A_n, n \in \mathbb{N})$, $\mathcal{A}$ l'algèbre engendré par les A_n c'est à dire les unions finis d'intersections finis de A_n, A_n^c (qui n'est en général pas une σ algèbres) qui est stable par, complémentaire union finie et intersection finie. Il est facile de voir que $\mathcal{A}$ est dénombrable.

En séparant les parties positives,négatives, on peut supposer $X_n \geq 0$ et par extraction diagonale, on trouve n_k telle que $\mathbf{E}[X_{n_k} 1_A] \to \mu(A)$ converge pour tout $A \in \mathcal{A}$.

Il est facile de voir que $\mu(\Omega) < \infty$ vu que (X_n) est bornée dans L^1 (par 3.) μ est additive sur les unions disjointes finies (par additivité de $1 \mapsto \mathbf{E}[X_{n_k} 1_A]$ qui est une mesure et passage à la limite). De plus, par 3., soit ϵ positive, on a un η tel que $P(A) \leq \eta$ implique $\mathbf{E}[X_{n_k} 1_A] \leq \epsilon$ donc $\mu(A) \leq \epsilon$.

En particulier si $P(A) = 0$, on a $\mu(A) = 0$.

Un résultat classique de théorie de la mesure dit que μ s'étend de façon unique sur $\sigma(\mathcal{A})$ en une mesure μ^* (cf. par exemple Barbe-Ledoux Th 1.49). Il est facile de voir que l'on a encore si $P(A) = 0$, on a $\mu^*(A) = 0$. Donc, $\mu^* \ll P$ et par le théorème de Radom-Nikodym, il existe $X \in L^1$ telle que $\mathbf{E}(X 1_A) = \mu(A) = \lim_{n \to_i nfty} \mathbf{E}[X_{n_k} 1_A]$. Il en est donc de même pour toute fonction étagée f_m (resp. g_m) d'une suite décroissante (resp. croissante) convergeant vers f mesurable positive bornée

D'où on a les deux inégalités donnant l'égalité
$$\limsup_{n \to \infty} \mathbf{E}[X_{n_k} f] \leq \lim_{n \to \infty} \mathbf{E}[X_{n_k} f_m] = \mathbf{E}(X f_m) \to \mathbf{E}(Xf)$$
$$\liminf_{n \to \infty} \mathbf{E}[X_{n_k} f] \geq \lim_{n \to \infty} \mathbf{E}[X_{n_k} g_m] = \mathbf{E}(X g_m) \to \mathbf{E}(Xf).$$
On a donc obtenu 2.

On laisse en exercice l'implication de 3. vers 1. que l'on n'a pas utilisé dans le cours.

■