

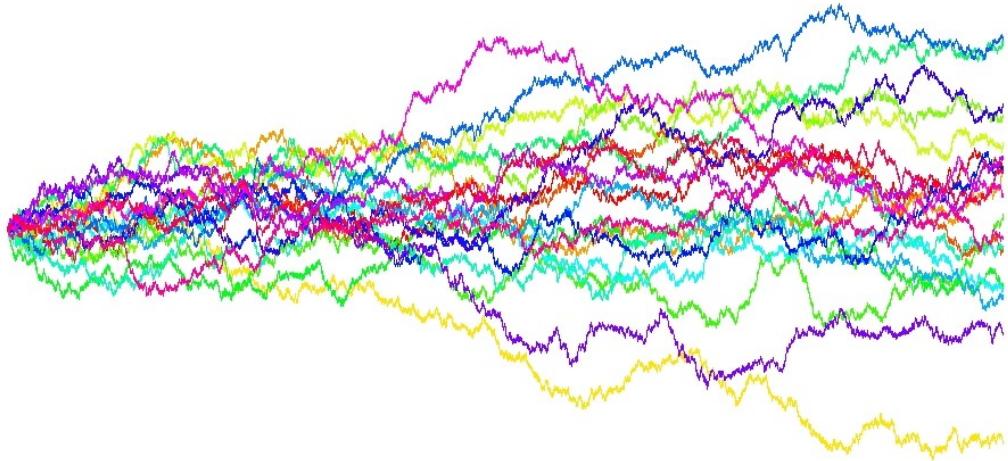
Processus stochastiques

Master 1 Maths
Université Côte d'Azur

(Version bêta du 25 octobre 2023)

Enseignants :

Roland Diel, Damien Garreau et Sylvain Rubenthaler



(Image de couverture : trajectoires de la marche aléatoire simple symétrique sur \mathbb{Z} .)

Bibliographie

Rappels à connaître pour pouvoir assister au cours dans de bonnes conditions :

- *Théorie des probabilités*, Candelopergher
- *De l'intégration aux probabilités*, Garet et Kurtzmann

Sur le contenu général du cours :

- *Probabilité*, Barbé et Ledoux
- *Probabilités (Tome 2)*, Ouvrard (très complet mais d'un abord plus difficile que le précédent)

Sur des points particuliers du cours (en anglais) :

- *Markov Chains*, Norris
- *Probability with Martingales*, Williams

Table des matières

1 Rappels : quelques théorèmes de théorie de la mesure utiles pour le cours de processus stochastiques	3
1.1 Théorème de convergence dominée	4
1.2 Théorème de convergence monotone	6
1.3 Lemme de Fatou	7
1.4 Théorèmes de Fubini	7
1.5 Théorème de dérivation sous le signe intégral	8
2 Espaces \mathbb{L}^p et fonctions caractéristiques	10
2.1 Espaces \mathbb{L}^p	10
2.2 Fonctions caractéristiques	13
3 Convergences stochastiques	17
3.1 Les différents types de convergence	17
3.2 Théorèmes limites pour les v.a.i.i.d.	25
4 Processus stochastiques (à temps discret)	29
4.1 Processus stochastiques et filtration	29
4.2 Espérance conditionnelle	33
5 Chaînes de Markov	39
5.1 Définition et premières propriétés	39
5.2 Représentation matricielle	42

1 Rappels : quelques théorèmes de théorie de la mesure utiles pour le cours de processus stochastiques

La théorie de la mesure permet de construire des objets appelés *intégrales abstraites*. Ce sont des généralisations de l'intégrale de Riemann classique vérifiant certaines de ses propriétés : linéarité, positivité, . . .

L'intégrale d'une fonction mesurable définie sur un ensemble mesuré $(\Omega, \mathcal{F}, \mu)$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ($\mathcal{B}(\mathbb{R})$ étant la tribu des boréliens) est construite en trois temps que nous rappelons très brièvement pour mettre en évidence la logique derrière la notion d'intégrale abstraite. Le lecteur intéressé par des résultats précis sur cette construction est invité à consulter les premiers ouvrages de la bibliographie.

- Dans un premier temps, il est naturel de définir l'intégrale de la fonction indicatrice d'un ensemble mesurable A comme la mesure de cet ensemble A :

$$\int_{\Omega} \mathbf{1}_A \, d\mu := \mu(A).$$

Comme on souhaite que notre intégrale soit linéaire, on peut alors étendre la définition aux fonctions étagées, i.e. ne prenant qu'un nombre fini de valeurs $\alpha_1 < \dots < \alpha_n$. Si on note A_i le sous ensemble de Ω où une fonction étagée f vaut α_i (soit $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$), on définit l'intégrale de f par rapport à μ par

$$\int_{\Omega} f \, d\mu := \sum_{i=1}^n \alpha_i \int_{\Omega} \mathbf{1}_{A_i} \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i).$$

- Ensuite, on construit l'intégrale de fonctions mesurables positives par approximation par des fonctions étagées. Notons E^+ l'ensemble des fonctions étagées positives. Pour une fonction mesurable positive, l'intégrale de $f : \Omega \rightarrow \mathbb{R}^+$ par rapport à μ est alors

$$\int_{\Omega} f \, d\mu := \sup_{h \leq f, h \in E^+} \int h \, d\mu \in [0; \infty]$$

- Finalement, on construit, quand c'est possible, l'intégrale de fonctions mesurables quelconques en considérant ses parties positive, f^+ , et négative, f^- , qui sont bien des fonctions mesurables positives. Ainsi si $f = f^+ - f^-$ est telle que $\int f^+ \, d\mu$ et $\int_{\Omega} f^- \, d\mu$ sont finies toutes les deux, on dit que f est intégrable et on pose :

$$\int_{\Omega} f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.$$

Nous ne reviendrons pas ici en détails sur les notions de mesures, de tribus ou encore de fonctions mesurables et nous renvoyons le lecteur intéressé vers un cours de théorie de la mesure et de l'intégration. Rappelons simplement que, «en pratique», toutes les fonctions rencontrées sont mesurables. En effet,

- toute fonction $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ continue par morceaux est mesurable (d, p entiers quelconques)
- par définition, toute v.a. $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (ou tout vecteur aléatoire $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) est mesurable.

Remarquons également que, par construction, l'intégrale de toute fonction positive (mesurable) est bien définie mais qu'elle peut prendre la valeur $+\infty$.

Le signification donnée à une intégrale abstraite dépend de la mesure μ considérée. Ainsi, si on choisit :

- $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, et μ la mesure de Lebesgue sur \mathbb{R}^d , on obtient l'intégrale classique :

$$\int_{\Omega} f(\omega) \mu(d\omega) = \int_{\mathbb{R}^d} f(x) dx .$$

- $(\Omega, \mathcal{F}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ et μ la mesure de comptage sur \mathbb{N} , on obtient la série :

$$\int_{\Omega} f(\omega) \mu(d\omega) = \sum_{n \in \mathbb{N}} f(n) .$$

- μ une mesure de probabilité \mathbb{P}_X sur (Ω, \mathcal{F}) , représentant la loi d'une v.a. réelle X (discrète, continue ou autre) on obtient une espérance :

$$\int_{\Omega} f(\omega) \mu(d\omega) = \int_{\Omega} f(\omega) \mathbb{P}_X(d\omega) = \mathbb{E}[f(X)] .$$

L'intérêt de la notion d'intégrale abstraite est de regrouper sous un même énoncé des résultats concernant des objets de nature a priori différente : intégrale, série, espérance. Nous utiliserons dans ce cours quelques résultats de la théorie de la mesure et de l'intégration dont nous rappelons les énoncés ci-dessous, d'abord sous une version générale théorie de la mesure, puis dans les cas particuliers de certaines mesures.

1.1 Théorème de convergence dominée

Ce théorème est le résultat le plus souvent utilisé pour passer à la limite dans une intégrale/espérance/série. Nous présentons d'abord le cas général puis le cas de mesures particulières en corollaires : ce sont les résultats utilisés en pratique dans le cours.

Théorème 1.1.1 (Version générale). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables d'un espace mesuré $(\Omega, \mathcal{F}, \mu)$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Si*

- *il existe une fonction f mesurable telle que pour μ -presque tout $\omega \in \Omega$,*

$$\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$$

- il existe une fonction g positive intégrable (i.e. telle que $\int_{\Omega} g(\omega) \mu(d\omega) < \infty$) telle que pour μ -presque tout $\omega \in \Omega$,

$$\forall n \in \mathbb{N}, |f_n(\omega)| \leq g(\omega)$$

alors f est intégrable et

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n(\omega) \mu(d\omega) = \int_{\Omega} f(\omega) \mu(d\omega) \quad \text{et} \quad \lim_{n \rightarrow \infty} \int_{\Omega} |f_n(\omega) - f(\omega)| \mu(d\omega) = 0 .$$

Corollaire 1.1.2 (Version intégrale de Riemann classique). Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions boréliennes (i.e. mesurables de $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$). Si

- il existe une fonction f mesurable telle que pour presque tout $x \in \mathbb{R}^p$,

$$\lim_{n \rightarrow \infty} f_n(x) = f(x)$$

- il existe une fonction g positive intégrable (i.e. telle que $\int_{\mathbb{R}^d} g(x) dx < \infty$) telle que pour presque tout $x \in \mathbb{R}^p$,

$$\forall n \in \mathbb{N}, |f_n(x)| \leq g(x)$$

alors f est intégrable sur \mathbb{R}^p et

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^p} f_n(x) dx = \int_{\mathbb{R}^p} f(x) dx \quad \text{et} \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^p} |f_n(x) - f(x)| dx = 0 .$$

Corollaire 1.1.3 (Version série). Soit $(u_{n,k})_{n,k \in \mathbb{N}^2}$ une suite, \tilde{A} double indice, de réels. Si

- il existe une suite $(u_k)_{k \in \mathbb{N}}$ telle que pour tout $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} u_{n,k} = u_k$
- il existe une suite $(v_k)_{k \in \mathbb{N}}$ positive sommable (i.e. telle que $\sum_{k \in \mathbb{N}} v_k < \infty$) telle que pour tout $n, k \in \mathbb{N}$, $|u_{n,k}| \leq v_k$

alors $(u_k)_{k \in \mathbb{N}}$ est sommable et

$$\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} u_{n,k} = \sum_{k \in \mathbb{N}} u_k \quad \text{et} \quad \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} |u_{n,k} - u_k| = 0 .$$

Corollaire 1.1.4 (Version probabiliste). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Si

- il existe une v.a. réelle X telle que $(X_n)_{n \in \mathbb{N}}$ converge vers X \mathbb{P} -p.s.
- il existe une v.a. Y positive intégrable (i.e. telle que $\mathbb{E}[Y] < \infty$) telle que \mathbb{P} -p.s., pour tout $n \in \mathbb{N}$, $|X_n| \leq Y$

alors X est intégrable et

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X] \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|] = 0 .$$

1.2 Théorème de convergence monotone

Moins souvent utilisé que le précédent, ce théorème a l'avantage de ne pas nécessiter la condition de domination. Elle est néanmoins remplacée (rien n'est gratuit !) par une condition de monotonie et de positivité. Remarquons que l'on n'obtient pas l'intégrabilité de la fonction limite et que les intégrales considérées peuvent donc être infinies (mais elles restent bien définies car on ne considère que des fonctions positives).

Théorème 1.2.1 (Version générale). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables positives d'un espace mesuré $(\Omega, \mathcal{F}, \mu)$ dans $([0, \infty], \mathcal{B}([0, \infty]))$. On suppose de plus la suite $(f_n)_{n \in \mathbb{N}}$ croissante : pour μ -presque tout $\omega \in \Omega$,*

$$\forall n \in \mathbb{N}, f_n(\omega) \leq f_{n+1}(\omega).$$

Alors la suite $(f_n)_{n \in \mathbb{N}}$ converge μ -presque partout vers une fonction mesurable f à valeurs dans $[0, \infty]$ et

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n(\omega) \mu(d\omega) = \int_{\Omega} f(\omega) \mu(d\omega).$$

Théorème 1.2.2 (Version intégrale de Riemann classique). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions boréliennes positives (i.e. mesurables de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dans $([0, \infty], \mathcal{B}([0, \infty]))$). On suppose de plus la suite $(f_n)_{n \in \mathbb{N}}$ croissante : pour presque tout $x \in \mathbb{R}^d$,*

$$\forall n \in \mathbb{N}, f_n(x) \leq f_{n+1}(x).$$

Alors la suite $(f_n)_{n \in \mathbb{N}}$ converge presque partout vers une fonction mesurable f à valeurs dans $[0, \infty]$ et

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f_n(x) dx = \int_{\mathbb{R}^d} f(x) dx$$

Théorème 1.2.3 (Version série). *Soit $(u_{n,k})_{n,k \in \mathbb{N}^2}$ une suite de réels positifs. On suppose que, pour tout $k \in \mathbb{N}$, la suite $(u_{n,k})_{n \in \mathbb{N}}$ est croissante :*

$$\forall n \in \mathbb{N}, u_{n,k} \leq u_{n+1,k}.$$

Alors il existe une suite $(u_k)_{k \in \mathbb{N}}$ à valeurs dans $[0, \infty]$ telle que pour tout $k \in \mathbb{N}$, $\lim_{n \rightarrow \infty} u_{n,k} = u_k$ et

$$\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} u_{n,k} = \sum_{k \in \mathbb{N}} u_k.$$

Théorème 1.2.4 (Version probabiliste). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. positives définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On suppose que la suite $(X_n)_{n \in \mathbb{N}}$ est croissante :*

$$\forall n \in \mathbb{N}, \mathbb{P}(X_n \leq X_{n+1}) = 1.$$

Alors il existe une v.a. X à valeurs dans $[0, \infty]$ telle que $(X_n)_{n \in \mathbb{N}}$ converge vers X \mathbb{P} -p.s. et

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X] \in [0; \infty].$$

1.3 Lemme de Fatou

Ce résultat permet encore d'étudier le comportement asymptotique d'une suite de fonctions/variables aléatoires positives. Il sera peu utilisé dans ce cours et peut être omis lors d'une première lecture.

Théorème 1.3.1 (Lemme de Fatou). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de $(\Omega, \mathcal{F}, \mu)$ à valeurs dans $[0, \infty]$. On a*

$$\int_{\Omega} \liminf_{n \rightarrow \infty} f_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

Théorème 1.3.2 (Lemme de Fatou, version probabiliste). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. à valeurs dans $[0, \infty]$. On a*

$$\mathbb{E} \left[\liminf_{n \rightarrow \infty} X_n \right] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n].$$

1.4 Théorèmes de Fubini

Ces théorèmes permettent d'échanger le sens d'intégration de deux intégrales abstraites. On suppose que les mesures sont choisies parmi les mesures de probabilité, de comptage sur un ensemble dénombrable ou de Lebesgue sur \mathbb{R}^d (ou un sous-ensemble de \mathbb{R}^d) pour un certain $d \in \mathbb{N}^*$.

Théorème 1.4.1 (Fubini-Tonelli). *Soit $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_2, \mu_2)$ deux espaces mesurés (tels que μ_1 et μ_2 soient des mesures de probabilité, de comptage sur un ensemble dénombrable ou de Lebesgue). Soit $f : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2) \rightarrow ([0, \infty], \mathcal{B}([0, \infty]))$ une fonction mesurable positive. Alors :*

1. les fonctions

$$\omega_1 \rightarrow \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \quad \text{et} \quad \omega_2 \rightarrow \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1)$$

sont respectivement \mathcal{F}_1 et \mathcal{F}_2 mesurables

2. on a l'égalité :

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \mu_1 \otimes \mu_2(d\omega_1, d\omega_2) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \end{aligned}$$

Théorème 1.4.2 (Fubini-Lebesgue). *Soit $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_2, \mu_2)$ deux espaces mesurés. Soit f une fonction intégrable de $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \otimes \mu_2)$ à valeurs dans \mathbb{R} . Alors :*

1. les fonctions (définies presque partout)

$$\omega_1 \rightarrow \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \quad \text{et} \quad \omega_2 \rightarrow \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1)$$

sont intégrables respectivement sur $(\Omega_1, \mathcal{F}_1, \mu_1)$ et sur $(\Omega_2, \mathcal{F}_2, \mu_2)$

2. On a l'égalité :

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \mu_1 \otimes \mu_2(d\omega_1, d\omega_2) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \end{aligned}$$

Ainsi, en pratique, selon les mesures choisies, le premier théorème nous permet d'affirmer par exemple pour toute suite/fonction mesurable f positive :

- $\sum_{n \in \mathbb{N}} \sum_{k \in \mathbb{N}} f_{n,k} = \sum_{k \in \mathbb{N}} \sum_{n \in \mathbb{N}} f_{n,k}$
- $\sum_{n \in \mathbb{N}} \int_{\mathbb{R}^d} f_n(x) dx = \int_{\mathbb{R}^d} \left(\sum_{n \in \mathbb{N}} f_n(x) \right) dx$
- $\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^{d'}} f(x, y) dy \right) dx = \int_{\mathbb{R}^{d'}} \left(\int_{\mathbb{R}^d} f(x, y) dx \right) dy$
- $\sum_{n \in \mathbb{N}} \mathbb{E}[f_n(X)] = \mathbb{E} \left[\sum_{n \in \mathbb{N}} f_n(X) \right]$
- $\int_{\mathbb{R}^d} \mathbb{E}[f(x, X)] dx = \mathbb{E} \left[\int_{\mathbb{R}^d} f(x, X) dx \right]$

Pour les fonctions quelconques, on commence par appliquer le théorème de Fubini-Tonelli à $|f|$, une fois qu'on a montré que la fonction était bien intégrable par rapport à $\mu_1 \otimes \mu_2$ (i.e. que l'intégrale de $|f|$ par rapport aux deux mesures est finie), on peut appliquer le théorème de Fubini-Lebesgue à f et faire exactement le même calcul mais sans la valeur absolue.

1.5 Théorème de dérivation sous le signe intégral

Le théorème suivant est dérivé du théorème de convergence dominée (d'où l'hypothèse de domination) et permet d'étudier la régularité d'une intégrale abstraite à paramètre.

Théorème 1.5.1. Soit O un ouvert de \mathbb{R} et $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. On considère $f : O \times \Omega \rightarrow \mathbb{R}$ une fonction satisfaisant les conditions suivantes :

1. Pour tout $x \in O$, la fonction $\omega \rightarrow f(x, \omega)$ est μ -intégrable.
2. Pour μ -presque tout $\omega \in \Omega$, la fonction $x \rightarrow f(x, \omega)$ est dérivable sur O .

3. Il existe une fonction intégrable $g : \Omega \rightarrow \mathbb{R}_+$ telle que :

$$\text{Pour } \mu\text{-p.t. } \omega, \forall x \in O, \quad \left| \frac{\partial f}{\partial x}(x, \omega) \right| \leq g(\omega)$$

Alors la fonction $x \rightarrow \int_{\Omega} f(x, \omega) \mu(d\omega)$ est dérivable sur O et

$$\forall x \in O, \quad \frac{d}{dx} \int_{\Omega} f(x, \omega) \mu(d\omega) = \int_{\Omega} \frac{\partial f}{\partial x}(x, \omega) \mu(d\omega).$$

2 Espaces \mathbb{L}^p et fonctions caractéristiques

Dans tout le chapitre, nous allons introduire des outils techniques que nous utiliserons tout au long du cours. Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ désignera l'espace de probabilité sous-jacent et on identifiera les variables aléatoires et les classes de variables aléatoires \mathbb{P} -p.s. égales.

2.1 Espaces \mathbb{L}^p

Définitions

- Pour $p \in [1, \infty[$, on note $\mathbb{L}^p := \mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ l'espace vectoriel des (classes de) variables aléatoires admettant un moment d'ordre p i.e.

$$\mathbb{L}^p = \{\text{v.a. } X / \mathbb{E}[|X|^p] < \infty\}.$$

On note alors pour toute v.a. $X \in \mathbb{L}^p$,

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

Si les v.a. ne sont pas dans \mathbb{L}^p , on utilise encore parfois la notation $\|X\|_p$ avec la convention $\|X\|_p = \infty$.

Exemples :

- Les v.a de lois gaussiennes, les v.a. de lois de Poisson sont dans \mathbb{L}^p pour tout p fini mais une v.a. de loi de Cauchy $\mathcal{C}(0, 1)$ (i.e. de densité $x \rightarrow \frac{1}{\pi(1+x^2)}$) n'est dans aucun \mathbb{L}^p , quel que soit $p \geq 1$.
- Une v.a. continue X_α de densité $f_\alpha(x) = \frac{\alpha-1}{x^\alpha} \mathbf{1}_{x \geq 1}$ pour $\alpha \geq 2$ est dans $\mathbb{L}^{\alpha-1}$ mais pas dans $\mathbb{L}^{\alpha-1+\epsilon}$ quel que soit $\epsilon > 0$.
- Pour $p = \infty$, on note \mathbb{L}^∞ l'espace vectoriel des variables aléatoires \mathbb{P} -p.s. bornées i.e.

$$\mathbb{L}^\infty = \{\text{v.a. } X / \exists M > 0, \mathbb{P}(|X| \leq M) = 1\}$$

Attention, dans la définition précédente, le réel M est *non aléatoire*.

On note alors, pour toute v.a. $X \in \mathbb{L}^\infty$,

$$\|X\|_\infty = \inf \{M, |X| \leq M \text{ } \mathbb{P}\text{-p.s.}\}.$$

Exemples :

- Une v.a. X de loi de Bernoulli $\mathcal{B}(p)$, $p \in [0, 1]$, est dans \mathbb{L}^∞ et $\|X\|_\infty = 1$ pour $p > 0$.
- Une v.a. X de loi binomiale $\mathcal{B}(n, p)$, $p \in [0, 1]$, $n \in \mathbb{N}^*$, est dans \mathbb{L}^∞ et $\|X\|_\infty = n$ pour $p > 0$.
- Une v.a. X de loi uniforme $\mathcal{U}([a, b])$, $-\infty < a < b < \infty$, est dans \mathbb{L}^∞ et $\|X\|_\infty = \max(|a|, b)$.
- Les v.a. gaussiennes ne sont pas dans \mathbb{L}^∞ .

Inégalités et moments

La première inégalité importante est l'inégalité de Markov qui contrôle la queue de la distribution en fonction du moment d'ordre p .

Théorème 2.1.1 (Inégalité de Markov généralisée). *Soit $p \in [1, \infty[$. Pour tout $X \in \mathbb{L}^p$,*

$$\forall a > 0, \quad \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p]}{a^p}.$$

Démonstration. Soit $a > 0$.

$$\mathbb{P}(|X| \geq a) = \mathbb{E}[\mathbf{1}_{|X|^p \geq a^p}] \leq \mathbb{E}\left[\frac{|X|^p}{a^p} \mathbf{1}_{|X|^p \geq a^p}\right] \leq \frac{\mathbb{E}[|X|^p]}{a^p}.$$

□

On en déduit l'inégalité de Tchebitchev :

Corollaire 2.1.2 (Inégalité de Tchebitchev). *Pour tout v.a. X admettant une variance (i.e. $X \in \mathbb{L}^2$), on a l'inégalité suivante*

$$\forall a > 0, \quad \mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}[X]}{a^2}.$$

Démonstration. On applique l'inégalité de Markov à la v.a. $X - \mathbb{E}[X]$ en prenant $p = 2$. □

On poursuit avec l'inégalité de Jensen qui étend l'inégalité caractérisant les fonctions convexes réelles à des lois de probabilités quelconques.

Théorème 2.1.3 (Inégalité de Jensen). *Soit une variable aléatoire $X \in \mathbb{L}^1$ (ou positive) à valeurs dans un intervalle I et ϕ une fonction convexe $I \rightarrow \mathbb{R}$ telle que $\phi(X) \in \mathbb{L}^1$ alors :*

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

Démonstration. Admis, voir par exemple la démonstration du théorème 10.12 du livre *Probabilités 2*, de Ouvrard □

Théorème 2.1.4 (Inégalité de Hölder). *Soit $p, q \in [1, \infty]$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Pour tout $X \in \mathbb{L}^p$ et $Y \in \mathbb{L}^q$, on a que le produit $XY \in \mathbb{L}^1$ et*

$$\mathbb{E}[|XY|] = \|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Démonstration. Le cas $p = \infty$ et $q = 1$ est évident. Pour $p \in]1, \infty[$, l'inégalité peut être déduite de l'inégalité précédente. Soit $\mathbb{E}[|Y|^q] = 0$, dans ce cas la v.a. Y est nulle \mathbb{P} -p.s. et le résultat est évident. Soit $\mathbb{E}[|Y|^q] > 0$ et on utilise l'inégalité de Jensen avec $x \rightarrow x^p$ et la mesure de probabilité $\mathbb{Q}(d\omega) = \frac{|Y|^q(\omega)}{\mathbb{E}[|Y|^q]}\mathbb{P}(d\omega)$:

$$\begin{aligned} \mathbb{E}[|XY|] &= \mathbb{E}[|Y|^q] \int_{\Omega} |X|(\omega) |Y|^{1-q}(\omega) \mathbb{Q}(d\omega) \\ &\leq \mathbb{E}[|Y|^q] \left(\int_{\Omega} |X|^p(\omega) |Y|^{-q}(\omega) \mathbb{Q}(d\omega) \right)^{1/p} \\ &= \mathbb{E}[|Y|^q]^{1-1/p} \|X\|_p \end{aligned}$$

□

Remarque : Cas particulier : $p = q = 2$. On retombe alors sur l'inégalité de Cauchy-Schwarz,

$$\forall X, Y \in \mathbb{L}^2, \quad \mathbb{E}[|XY|] \leq \|X\|_2 \|Y\|_2.$$

Propriétés des espaces \mathbb{L}^p

Théorème 2.1.5. Pour tout $p \in [1, \infty]$, l'espace $\mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ est un sous espace vectoriel de l'espace des (classes de) v.a. sur Ω .

Démonstration. Soit $p \in [1, \infty[$ (le cas $p = \infty$ est laissé en exercice). Comme $\mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ est un sous ensemble de l'espace vectoriel des (classes de) v.a. sur Ω , il suffit de vérifier qu'il est stable par combinaison linéaire :

- il est clair que pour $\lambda \in \mathbb{R}$ et $X \in \mathbb{L}^p$, $\lambda X \in \mathbb{L}^p$.
- Par convexité de la fonction $x \rightarrow x^p$ sur $[0, \infty[$, on a

$$\forall x, y \geq 0, \quad \left(\frac{x}{2} + \frac{y}{2} \right)^p \leq \frac{1}{2}x^p + \frac{1}{2}y^p$$

soit

$$\forall x, y \geq 0, \quad (x + y)^p \leq 2^{p-1} (x^p + y^p).$$

Ainsi, si $X, Y \in \mathbb{L}^p$, d'après l'inégalité précédente et la linéarité de l'espérance,

$$\mathbb{E}[|X + Y|^p] \leq \mathbb{E}[(|X| + |Y|)^p] \leq 2^{p-1} (\mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p]) < \infty$$

et $X + Y \in \mathbb{L}^p$ également.

□

Théorème 2.1.6. Pour tout $p \in [1, \infty]$,

1. $\|\cdot\|_p$ est une norme sur l'espace \mathbb{L}^p i.e.
 - (a) $\|X\|_p \geq 0$ et ($\|X\|_p = 0$ ssi $X = 0$ \mathbb{P} -p.s.)
 - (b) $\forall \lambda \in \mathbb{R}, \|\lambda X\|_p = |\lambda| \|X\|_p$
 - (c) pour tous $X, Y \in \mathbb{L}^p$,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \text{ (inégalité de Minkowski).}$$

2. L'espace $(\mathbb{L}^p, \|\cdot\|_p)$ est un espace de Banach (ie les suites de Cauchy sont des suites convergentes).

Démonstration.

1. Le seul point non évident est l'inégalité de Minkowski. Le cas $p = \infty$ est simple. Pour le cas $p < \infty$, on utilise l'inégalité de Hölder avec p et $q = p/(p-1) \in]1, \infty]$. En effet, on a $|X + Y|^{p-1} \in \mathbb{L}^q$ et ainsi

$$\begin{aligned} \|X + Y\|_p^p &= \mathbb{E}[|X + Y|^p] \\ &\leq \mathbb{E}[|X| \cdot |X + Y|^{p-1}] + \mathbb{E}[|Y| \cdot |X + Y|^{p-1}] \\ &\leq (\|X\|_p + \|Y\|_p) \|(X + Y)^{p-1}\|_q \\ &= (\|X\|_p + \|Y\|_p) \|X + Y\|_p^{(p-1)} \end{aligned}$$

D'où le résultat.

2. Admis, voir par exemple la démonstration du théorème 10.12 du livre *Probabilités 2*, de Ouvrard

□

Théorème 2.1.7. Si $1 \leq p \leq q \leq \infty$, alors $\mathbb{L}^q \subset \mathbb{L}^p$ et $\|\cdot\|_p \leq \|\cdot\|_q$.

Démonstration. On utilise l'inégalité de Jensen avec la fonction $x \rightarrow x^{q/p}$. Cette fonction est bien convexe car $\frac{q}{p} \geq 1$. Soit $X \in \mathbb{L}^q$, on a donc :

$$(\mathbb{E}[|X|^p])^{q/p} \leq \mathbb{E}[|X|^{p \cdot q/p}] = \mathbb{E}[|X|^q].$$

□

Ainsi, par exemple, une variable aléatoire admettant un moment d'ordre 2 admet également un moment d'ordre 1 et $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$.

Remarque : le théorème 2.1.7 n'est valable qu'avec les espaces \mathbb{L}^p associés à des mesures finies et même de masse totale 1 pour avoir l'inégalité des normes.

2.2 Fonctions caractéristiques

La fonction caractéristique d'une variable aléatoire réelle X détermine de façon unique sa loi de probabilité. Il s'agit donc comme son nom l'indique d'une caractérisation des lois comme la fonction de répartition, la densité dans le cas de v.a. continues, ou le germe de probabilité dans le cas de v.a. discrètes. La fonction caractéristique est, à un signe près, la transformée de Fourier de la mesure de probabilité \mathbb{P}_X . Ainsi, la régularité de cette fonction est liée à l'existence des moments de la variable (cf. points 6 et 7 de la propriété 2.2.2 ci-dessous).

Définition 2.2.1. La fonction caractéristique de la v.a. réelle X est la fonction à valeurs complexes définie par :

$$\Phi_X : \begin{cases} \mathbb{R} & \rightarrow \mathbb{C} \\ t & \mapsto \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)] \end{cases}$$

Remarque : $\mathbb{E}[|e^{itX}|] = \mathbb{E}[1] = 1$ donc la fonction caractéristique est bien définie sur \mathbb{R} pour toute variable X (ce qui n'est pas le cas avec la transformée de Laplace : $t \rightarrow \mathbb{E}[e^{-tX}]$).

Exemples :

- binomiale $\mathcal{B}(p) : t \rightarrow ((1-p) + pe^{it})^n$
- Poisson $\mathcal{P}(\lambda) : t \rightarrow e^{\lambda(e^{it}-1)}$
- Géométrique $\mathcal{G}(p) : t \rightarrow \frac{pe^{it}}{1-(1-p)e^{it}}$
- Uniforme $\mathcal{U}[a, b] : t \rightarrow \frac{e^{itb}-e^{ita}}{it(b-a)}$
- Exponentielle $\mathcal{E}(\lambda) : t \rightarrow \frac{\lambda}{\lambda-it}$
- Normale $\mathcal{N}(\mu, \sigma^2) : t \rightarrow e^{it\mu - \frac{\sigma^2}{2}t^2}$
- Cauchy $\mathcal{C}(\mu, a) : t \rightarrow e^{it\mu - a|t|}$
(Rappel : densités des lois de Cauchy : $x \rightarrow \frac{1}{\pi a \left[1 + \left(\frac{x-\mu}{a} \right)^2 \right]}$)

Voici quelques propriétés plus ou moins élémentaires des fonctions caractéristiques.

Proposition 2.2.2. *Soit X, Y deux v.a. et Φ_X et Φ_Y leur fonction caractéristique.*

1. Φ_X est une fonction continue sur \mathbb{R} bornée par 1 : $\forall t \in \mathbb{R}, |\Phi_X(t)| \leq 1$.
2. $\Phi_X(0) = 1$.
3. Pour tout $t \in \mathbb{R}$, $\Phi_X(-t) = \overline{\Phi_X(t)}$.
4. Si X et Y sont indépendantes, alors $\Phi_{X+Y} = \Phi_X \cdot \Phi_Y$.
5. Pour tout $a, b \in \mathbb{R}$, $\forall t \in \mathbb{R}, \Phi_{aX+b}(t) = e^{ibt}\Phi_X(at)$
6. Si X admet un moment d'ordre k , alors Φ_X est \mathcal{C}^k sur \mathbb{R} et pour tout $\ell \leq k$,

$$\Phi_X^{(\ell)}(0) = i^\ell \mathbb{E}[X^\ell]$$

7. Si Φ_X est \mathcal{C}^k sur \mathbb{R} pour un k pair, alors X admet un moment d'ordre k et pour tout $\ell \leq k$,

$$\Phi_X^{(\ell)}(0) = i^\ell \mathbb{E}[X^\ell]$$

Démonstration. Les points 1 à 5 peuvent être démontrés à titre d'exercice. Le point 6 est une conséquence du théorème de dérivation sous le signe intégral. On renvoie à la proposition 12.14 de *Probabilités 2*, Ouvrard) \square

Comme indiqué en préambule, les fonctions caractéristiques caractérisent la loi d'une v.a.

Théorème 2.2.3. *Deux v.a. X et Y ont même loissi $\Phi_X = \Phi_Y$.*

Démonstration. Admis (cf Théorème 12.6 de *Probabilités 2*, Ouvrard) \square

Un des intérêts des fonctions caractéristiques est de pouvoir calculer facilement la loi de somme de variables aléatoires indépendantes en utilisant le point 4 de la propriété 2.2.2.

Exemple : Soit X et Y deux variables gaussiennes indépendantes de loi respective $\mathcal{N}(\mu_X, \sigma_X^2)$ et $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Calculons la fonction caractéristique de la somme $Z = X + Y$. Par indépendance de X et Y ,

$$\forall t \in \mathbb{R}, \Phi_Z(t) = \Phi_X(t)\Phi_Y(t) = e^{it\mu_X - \frac{\sigma_X^2}{2}t^2} e^{it\mu_Y - \frac{\sigma_Y^2}{2}t^2} = e^{it(\mu_X + \mu_Y) - \frac{\sigma_X^2 + \sigma_Y^2}{2}t^2}.$$

Il s'agit de la fonction caractéristique de la loi $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$, la variable Z suit donc cette loi.

Remarque : La réciproque du point 4 de la propriété 2.2.2 est fausse. En effet, si X est une v.a. de loi de Cauchy $\mathcal{C}(0, 1)$ alors, $\Phi_{2X}(t) = \Phi_X(2t) = e^{-2|t|} = \Phi_X(t)\Phi_X(t)$ mais la variable X n'est pas indépendante d'elle-même.

Théorème 2.2.4 (Formule d'inversion).

1. Soit X une v.a. réelle telle que sa fonction caractéristique Φ_X est intégrable sur \mathbb{R} , $\int_{\mathbb{R}} |\Phi_X(t)| dt < \infty$, alors X est une v.a. continue dont la densité est donnée par :

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi_X(t) dt$$

2. Réciproquement, soit $\Phi : \mathbb{R} \rightarrow \mathbb{C}$ une fonction intégrable telle que la fonction f définie par

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi(t) dt$$

est la densité de probabilité d'une v.a. X alors Φ est la fonction caractéristique de X .

Démonstration. Admis (cf Ouvrard, Probabilités, Tome 2) □

Le second point du théorème précédent permet de calculer simplement la fonction caractéristique de la loi de Cauchy donnée dans les exemples. Considérons la fonction $\Phi : t \rightarrow e^{-|t|}$ qui est bien intégrable sur \mathbb{R} . Alors, la fonction f du théorème est bien définie et

$$\begin{aligned} \forall x \in \mathbb{R}, f(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} e^{-|t|} dt \\ &= \frac{1}{2\pi} \left(\int_0^\infty e^{-itx-t} dt + \int_{-\infty}^0 e^{-itx+t} dt \right) \\ &= \frac{1}{2\pi} \left(\left[\frac{e^{-(1+ix)t}}{-(1+ix)} \right]_0^\infty + \left[\frac{e^{(1-ix)t}}{(1-ix)} \right]_{-\infty}^0 \right) \\ &= \frac{1}{2\pi} \left(\frac{1}{1+ix} + \frac{1}{1-ix} \right) = \frac{1}{\pi(1+x^2)}. \end{aligned}$$

La fonction f est bien la densité de la loi de Cauchy $\mathcal{C}(0, 1)$ et d'après le point 2 du théorème précédent, $t \rightarrow e^{-|t|}$ est donc la fonction caractéristique de la loi de Cauchy $\mathcal{C}(0, 1)$.

Finalement, tous ces résultats s'étendent au cas de vecteurs aléatoires, si on remplace le produit classique par le produit scalaire. La *fonction caractéristique du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$* à valeurs dans \mathbb{R}^d est définie par :

$$\Phi_{\mathbf{X}} : \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R} \\ \mathbf{t} = (t_1, \dots, t_d) & \mapsto \mathbb{E}[e^{i\langle \mathbf{t} | \mathbf{X} \rangle}] = \mathbb{E}\left[e^{i\sum_{k=1}^d t_k X_k}\right] \end{cases}$$

Là encore :

Théorème 2.2.5. *La fonction caractéristique d'un vecteur aléatoire caractérise la loi du vecteur : deux vecteurs \mathbf{X} et \mathbf{Y} ont même loissi $\Phi_{\mathbf{X}} = \Phi_{\mathbf{Y}}$.*

Démonstration. Admis (cf Ouvrard, Probabilités, Tome 2) □

L'indépendance de variables aléatoires se lit sur la fonction caractéristique du vecteur aléatoire associé :

Théorème 2.2.6. Soit $\mathbf{X} = (X_1, \dots, X_d)$ un vecteur aléatoire à valeurs dans \mathbb{R}^d .

$$X_1, \dots, X_d \text{ sont indépendantesssi } \forall \mathbf{t} = (t_1, \dots, t_d) \in \mathbb{R}^d, \Phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^d \Phi_{X_k}(t_k).$$

Démonstration. On suppose dans un premier temps que les variables X_1, \dots, X_d sont indépendantes. Alors

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\left[e^{i \sum_{k=1}^d t_k X_k}\right] = \mathbb{E}\left[\prod_{k=1}^d e^{it_k X_k}\right] = \prod_{k=1}^d \mathbb{E}\left[e^{it_k X_k}\right] = \prod_{k=1}^d \Phi_{X_k}(t_k).$$

Supposons maintenant que

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^d \Phi_{X_k}(t_k).$$

Alors, le vecteur $\mathbf{X} = (X_1, \dots, X_d)$ a même fonction caractéristique que le vecteur $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_d)$, où les $\tilde{X}_1, \dots, \tilde{X}_d$ sont des variables aléatoires indépendantes telles que pour tout $k \in \{1, \dots, d\}$, $X_k \sim \tilde{X}_k$. Comme la fonction caractéristique caractérise la loi, \mathbf{X} et $\tilde{\mathbf{X}}$ ont même loi et donc (X_1, \dots, X_d) sont indépendantes. \square

3 Convergences stochastiques

3.1 Les différents types de convergence

Définitions et premières propriétés

Nous allons étudier des processus stochastiques, c'est-à-dire des suites de v.a. Comme pour les suites réelles, un point important est l'étude de leur comportement asymptotique et notamment de leur limite potentielle. Mais contrairement aux suites numériques, la notion de convergence n'est pas unique. Dans cette partie, nous présentons les différents modes de convergence utilisés et les liens qui existent entre eux.

Définition 3.1.1.

1. Une suite de v.a. réelles $(X_n)_{n \in \mathbb{N}}$ converge \mathbb{P} -presque sûrement vers une v.a. X si

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1 .$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$.

2. Une suite de v.a. réelles $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers une v.a. X si

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0 .$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$.

3. Soit $p \in [1, \infty]$. Une suite de v.a. réelles $(X_n)_{n \in \mathbb{N}}$ d'éléments de \mathbb{L}^p , converge en norme \mathbb{L}^p vers une v.a. $X \in \mathbb{L}^p$ si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0 .$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X$.

4. Une suite de v.a. réelles $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une v.a. X si pour toute fonction continue bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] .$$

Notation : $X_n \xrightarrow[n \rightarrow \infty]{loi} X$.

Ces définitions s'étendent naturellement au cas de vecteurs aléatoires. De plus, si pour les trois premiers types de convergence, les v.a. doivent toutes être définies sur un même espace de probabilité. Cela n'est pas nécessaire pour la convergence en loi qui est, comme son nom l'indique, une convergence des mesures de probabilité associées : \mathbb{P}_{X_n} vers \mathbb{P}_X .

Proposition 3.1.2. 1. La limite presque sûre (resp. en probabilité, resp. dans un \mathbb{L}^p) est unique \mathbb{P} -p.s. : si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ et $X_n \xrightarrow[n \rightarrow \infty]{p.s.} Y$ (resp. en probabilité ou dans \mathbb{L}^p), alors $\mathbb{P}(X = Y) = 1$.

2. si $X_n \xrightarrow[n \rightarrow \infty]{loi} X$ et $X_n \xrightarrow[n \rightarrow \infty]{loi} Y$, alors $\mathbb{P}_X = \mathbb{P}_Y$.

Démonstration.

1. Le résultat dans le cas presque sûr est une conséquence directe de l'unicité de la limite pour les suites réelles (ou vectorielles) non aléatoires. Pour la limite \mathbb{L}^p , le résultat est impliqué par le fait que \mathbb{L}^p soit un espace vectoriel normé, et donc métrique. Pour le cas de la convergence en probabilité, remarquons que

$$\mathbb{P}(X \neq Y) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} \left\{ |X - Y| \geq \frac{1}{n} \right\}\right) \leq \sum_{n \in \mathbb{N}^*} \mathbb{P}\left(\left\{ |X - Y| \geq \frac{1}{n} \right\}\right).$$

Or, pour tout $n \in \mathbb{N}^*$, $k \in \mathbb{N}^*$,

$$\mathbb{P}(|X - Y| \geq \frac{1}{n}) \leq \mathbb{P}(|X - X_k| + |X_k - Y| \geq \frac{1}{n}) \leq \mathbb{P}(|X - X_k| \geq \frac{1}{2n}) + \mathbb{P}(|X_k - Y| \geq \frac{1}{2n})$$

Par définition de la convergence en probabilité, on a alors : pour tout $n > 0$,

$$\lim_{k \rightarrow \infty} \mathbb{P}(|X - X_k| \geq \frac{1}{2n}) = \lim_{k \rightarrow \infty} \mathbb{P}(|X_k - Y| \geq \frac{1}{2n}) = 0.$$

Ainsi, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(|X - Y| \geq \frac{1}{n}) = 0$ et $\mathbb{P}(X \neq Y) = 0$.

2. Si X et Y sont deux limites loi d'une même suite, alors : pour toute fonction continue bornée $f : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$. Cela implique bien que X et Y ont même loi.

□

Remarque : Dans le cas vectoriel, pour la définition des convergences en probabilité, presque sûre et \mathbb{L}^p de v.a. à valeurs dans \mathbb{R}^d , le choix de la norme n'a pas d'influence sur la limite. En effet, toutes les normes sur \mathbb{R}^d sont équivalentes.

Proposition 3.1.3. On considère une v.a. X , une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$, et une fonction continue f de \mathbb{R} dans \mathbb{R} .

Si $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ (resp. $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$ ou $X_n \xrightarrow[n \rightarrow \infty]{loi} X$), alors $f(X_n) \xrightarrow[n \rightarrow \infty]{p.s.} f(X)$ (resp. $f(X_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} f(X)$ ou $f(X_n) \xrightarrow[n \rightarrow \infty]{loi} f(X)$).

Démonstration.

- *Convergence en probabilité* : On fixe un $\epsilon > 0$ et on va montrer que $\mathbb{P}(|f(X_n) - f(X)| \geq \epsilon)$ tend vers 0 lorsque n tend vers l'infini.

La fonction f étant continue, elle est uniformément continue sur tout intervalle de longueur finie (théorème de Heine). Ainsi, fixons $M > 0$, l'uniforme continuité de f sur $[-2M; 2M]$ nous dit qu'il existe $\delta > 0$,

$$\forall (x, y) \in [-2M; 2M], |x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

soit

$$\forall (x, y) \in [-2M; 2M], |f(x) - f(y)| \geq \epsilon \Rightarrow |x - y| > \delta.$$

On peut alors décomposer la probabilité d'intérêt : pour tout $n \in \mathbb{N}$,

$$\begin{aligned} & \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \\ &= \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon, |X| > M) + \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon, |X| \leq M, |X_n| > 2M) \\ &\quad + \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon, |X| \leq M, |X_n| \leq 2M) \\ &\leq \mathbb{P}(|X| > M) + \mathbb{P}(|X - X_n| > M) + \mathbb{P}(|X_n - X| > \delta) \end{aligned}$$

Ainsi, en faisant tendre n vers ∞ ,

$$\forall M > 0, \limsup_{n \rightarrow \infty} \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) \leq \mathbb{P}(|X| > M).$$

La variable X est p.s. finie donc $\lim_{M \rightarrow \infty} \mathbb{P}(|X| > M) = \mathbb{P}(|X| = \infty) = 0$ et on a bien

$$\lim_{n \rightarrow \infty} \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) = 0 \quad \text{soit} \quad f(X_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} f(X).$$

- *Convergence p.s.* : On considère l'événement $A = \{\lim_n X_n = X\}$. Par hypothèse, on a $\mathbb{P}(A) = 1$. Pour tout $\omega \in A$, $\lim_n X_n(\omega) = X(\omega)$ et par caractérisation séquentielle de la continuité, on a également $\lim_n f(X_n(\omega)) = f(X(\omega))$. Ainsi, $A \subset \{\lim_n f(X_n) = f(X)\}$. D'où le résultat.
- *Convergence en loi* : On suppose que $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$. Montrons que $f(X_n) \xrightarrow[n \rightarrow \infty]{\text{loi}} f(X)$. Soit g une fonction continue bornée. On remarque que $g(f(X_n)) = g \circ f(X_n)$. Les fonctions f et g étant continues, la fonction $g \circ f$ est également continue. De plus, g étant bornée, $g \circ f$ l'est également. Ainsi, par convergence en loi de X_n vers X , on a bien :

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(f(X_n))] = \mathbb{E}[g(f(X))]$$

$$\text{et } f(X_n) \xrightarrow[n \rightarrow \infty]{\text{loi}} f(X).$$

□

Remarques :

- La fonction f n'a en fait besoin d'être définie que sur un sous ensemble (borélien) E de \mathbb{R} tel que $\mathbb{P}(X \in E)$ et $\forall n \in \mathbb{N}, X_n \in E$ = 1, par exemple $E = \mathbb{R}_+$ si les v.a. sont toutes p.s. positives.
- Là encore, le résultat se généralise au cas de vecteurs aléatoires.
- Quel que soit $p \in [1, \infty[$, on ne conserve pas la convergence \mathbb{L}^p en composant par une fonction continue. L'appartenance à l'espace \mathbb{L}^p n'est même pas conservée a priori. Ainsi, avec la fonction $x \rightarrow x^2$, si $X \in \mathbb{L}^1$ et $X \notin \mathbb{L}^2$ alors $X^2 \notin \mathbb{L}^1$.

Proposition 3.1.4. *Soit une suite $(X_n)_{n \in \mathbb{N}}$ convergeant p.s. (resp. en probabilité, resp. dans \mathbb{L}^p) vers une variable (ou vecteur de \mathbb{R}^d) X et une suite $(Y_n)_{n \in \mathbb{N}}$ convergeant p.s. (resp. en probabilité, resp. dans \mathbb{L}^p) vers une variable (ou vecteur de \mathbb{R}^q) Y . Alors la suite (X_n, Y_n) converge p.s. (resp. en probabilité, resp. dans \mathbb{L}^p) vers le vecteur de $\mathbb{R}^d \times \mathbb{R}^p$ (X, Y) .*

Démonstration. Exercice.

□

Le résultat précédent n'est pas vrai pour la convergence en loi car la convergence des marginales ne donne pas d'information pour la loi jointe.

Le théorème suivant décrit les liens entre les différents modes de convergence.

Théorème 3.1.5.

1. Si $1 \leq p \leq q \leq \infty$, la convergence dans \mathbb{L}^q implique la convergence dans \mathbb{L}^p .
2. Pour tout $p \in [1, \infty]$, la convergence \mathbb{L}^p implique la convergence en probabilité.
3. La convergence presque sûre implique la convergence en probabilité.
4. La convergence en probabilité implique la convergence en loi.
5. Si $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X alors il existe une sous-suite de $(X_n)_{n \in \mathbb{N}}$ qui converge \mathbb{P} -p.s. vers X .
6. Si $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une constante $c \in \mathbb{R}$ (et que les v.a. $(X_n)_{n \in \mathbb{N}}$ vivent sur le même espace de probabilité), alors $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers c .

À chaque fois qu'une convergence en implique une autre, les limites sont égales.

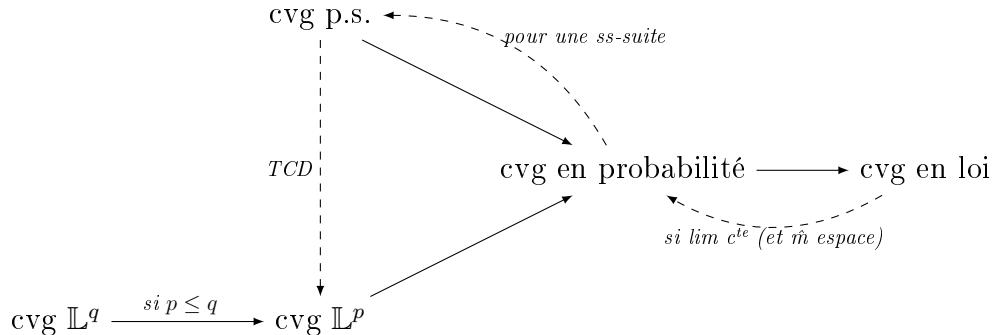


FIGURE 1 – Liens entre les différentes convergences

Démonstration.

1. C'est une conséquence directe du Théorème 2.1.7.
2. Il suffit d'utiliser l'inégalité de Markov.
3. Soit $\epsilon > 0$.

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{E}[\mathbf{1}_{|X_n - X| \geq \epsilon}]$$

La v.a. $\mathbf{1}_{|X_n - X| \geq \epsilon}$ converge p.s. vers 0 car $(X_n)_{n \in \mathbb{N}}$ converge vers X \mathbb{P} -p.s.. De plus , elle est majorée par la v.a. constante égale à 1 qui est dans \mathbb{L}^1 . D'après le théorème de convergence dominée, on a donc,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{E}[0] = 0.$$

4. Soit f une fonction continue bornée sur \mathbb{R} . Comme il y a convergence en probabilité, les variables aléatoires sont nécessairement définies sur le même espace de probabilité et pour tout $\epsilon > 0$,

$$\begin{aligned} \forall n \in \mathbb{N}, \quad & |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \mathbb{E}[|f(X_n) - f(X)|] \\ & \leq \mathbb{E}[|f(X_n) - f(X)| \mathbf{1}_{|f(X_n) - f(X)| \geq \epsilon}] + \mathbb{E}[|f(X_n) - f(X)| \mathbf{1}_{|f(X_n) - f(X)| < \epsilon}] \\ & \leq 2\|f\|_{\infty} \mathbb{P}(|f(X_n) - f(X)| \geq \epsilon) + \epsilon. \end{aligned}$$

Comme $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , d'après la propriété précédente, $f(X_n)$ converge également en probabilité vers $f(X)$. En faisant tendre n vers l'infini, on obtient

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \epsilon$$

pour tout $\epsilon > 0$. Cette limite est donc nulle et on a bien convergence en loi de $(X_n)_{n \in \mathbb{N}}$ vers X .

5. et 6. Cf Feuille de TD 2.

□

La convergence \mathbb{P} -p.s. est souvent la plus difficile à montrer, une condition suffisante utile est donnée dans la proposition suivante :

Proposition 3.1.6. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. et X une autre v.a. Si, pour tout $\epsilon > 0$, la série $\sum_{n=0}^{\infty} \mathbb{P}(|X_n - X| \geq \epsilon)$ converge, alors la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X \mathbb{P} -p.s.*

Démonstration. Cf feuille de TD 2.

□

Remarque : Cette condition n'est pas nécessaire. Considérons

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), dx)$$

et la suite de v.a. définie sur Ω : $\forall n \geq 1$, $X_n = \mathbf{1}_{[0, 1/n]}$. Pour tout $\omega \in [0, 1]$, $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Comme $\mathbb{P}(\{0\}) = 0$, $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$. Or,

$$\forall \epsilon > 0, \quad \mathbb{P}(|X_n - 0| \geq \epsilon) = \mathbb{P}([0, 1/n]) = 1/n$$

La série des $1/n$ n'étant pas sommable, la suite $(X_n)_{n \in \mathbb{N}}$ converge p.s. vers 0 mais ne vérifie pas le critère de la proposition 3.1.6.

Convergence en loi

La convergence en loi, comme son nom l'indique, est en fait une convergence des lois de probabilité et non des v.a. en elles-mêmes. Les différentes v.a. n'ont d'ailleurs pas à être définies sur le même espace de probabilité. Si la définition donnée plus haut peut paraître obscure de prime abord, les caractérisations présentées ci-dessous peuvent rendre cette notion plus intuitive.

Théorème 3.1.7 (Caractérisation de la convergence en loi). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. et X une autre variable aléatoire. On a les équivalences suivantes :*

1. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X .
2. La suite des fonctions de répartition F_{X_n} converge en tout point de continuité vers la fonction de répartition F_X .
3. Pour tout $t \in \mathbb{R}$, la suite $\Phi_{X_n}(t)$ converge vers $\Phi_X(t)$. (Théorème de Lévy)

Démonstration. Admis □

Étudions maintenant le cas particuliers des v.a. discrètes et continues :

Proposition 3.1.8 (V.a. à valeurs dans \mathbb{Z} (ou dans un sous-ensemble discret de \mathbb{R})). Si $(X_n)_{n \in \mathbb{N}}$ est une suite de v.a. à valeurs dans \mathbb{Z} et X une autre v.a. à valeurs dans \mathbb{Z} , alors

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \text{ssi } \forall k \in \mathbb{Z}, \lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k) .$$

Démonstration.

\Rightarrow) Supposons que $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$. Soit $k \in \mathbb{Z}$. Considérons la fonction réelle f_k définie par

$$f_k(x) = \begin{cases} 2(x - k) + 1 & \text{si } x \in [k - 1/2; k[\\ 2(k - x) + 1 & \text{si } x \in [k; k + 1/2[\\ 0 & \text{sinon} \end{cases}$$

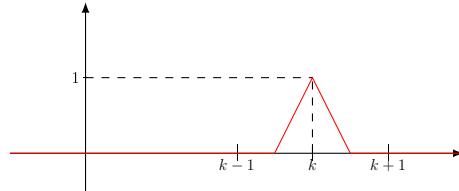


FIGURE 2 – Fonction f_k

La fonction f_k étant continue et bornée, $\mathbb{E}[f_k(X_n)] \rightarrow \mathbb{E}[f_k(X)]$. Or, $\mathbb{E}[f_k(X_n)] = \mathbb{P}(X_n = k)$ et $\mathbb{E}[f_k(X)] = \mathbb{P}(X = k)$ d'où le résultat recherché.

\Leftarrow) On fixe une fonction f continue bornée et on se donne un entier $M > 0$. Par hypothèse,

$$\mathbb{E}[f(X_n) \mathbf{1}_{|X_n| \leq M}] = \sum_{k=-M}^M f(k) \mathbb{P}(X_n = k) \xrightarrow{n \rightarrow \infty} \mathbb{E}[f(X) \mathbf{1}_{|X| \leq M}] .$$

Le problème survient lorsqu'on considère une infinité de probabilités élémentaires en même temps i.e. sur l'événement $\{|X_n| > M\}$. L'idée est de repasser à l'événement complémentaire :

$$\begin{aligned} |\mathbb{E}[f(X_n) \mathbf{1}_{|X_n| > M}] - \mathbb{E}[f(X) \mathbf{1}_{|X| > M}]| &\leq \|f\|_\infty (\mathbb{P}(|X_n| > M) + \mathbb{P}(|X| > M)) \\ &= \|f\|_\infty (1 - \mathbb{P}(|X_n| \leq M) + \mathbb{P}(|X| > M)) \end{aligned}$$

Comme $\mathbb{P}(|X_n| \leq M) = \sum_{k=-M}^M \mathbb{P}(X_n = k)$ dépend d'un nombre fini de probabilités élémentaires, on a $\lim_n \mathbb{P}(|X_n| \leq M) = \mathbb{P}(|X| \leq M)$ et donc

$$\limsup_n |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\|f\|_\infty \mathbb{P}(|X| > M) .$$

Cette inégalité étant valable pour tout $M > 0$, la limsup est nulle et on a bien la convergence en loi.

□

Proposition 3.1.9 (V.a. continues). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. de densité respective f_n et X une autre v.a. de densité f . Si $f_n(x)$ converge vers $f(x)$ pour tout réel x , alors $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$.

Démonstration. En exercice. (On pourra remarquer que $|f_n - f| = f_n + f - 2 \min(f_n, f)$.)
□

Remarque : Attention, la réciproque est fausse : la convergence en loi de v.a. continues n'implique pas la convergence des densités même si la limite a une loi continue. Voir le contre-exemple suivant : soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. telle que pour tout $n \in \mathbb{N}^*$, X_n admette pour densité :

$$x \rightarrow f_n(x) = (1 - \cos(2\pi nx))\mathbf{1}_{[0,1]}(x)$$

et X une v.a. de loi $\mathcal{U}([0, 1])$ i.e. de densité $f(x) = \mathbf{1}_{x \in [0,1]}$. On peut montrer que $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$, en étudiant les fonctions de répartition par exemple, mais il n'y a pas convergence de f_n vers f .

Pour finir sur la convergence en loi, remarquons que contrairement aux autres modes de convergence stochastique, la convergence en loi de deux suites (définies sur le même espace) n'implique pas la convergence en loi de la somme :

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \text{ et } Y_n \xrightarrow[n \rightarrow \infty]{\text{loi}} Y \not\Rightarrow X_n + Y_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X + Y$$

ou plus généralement la convergence du couple (X_n, Y_n) . Cela se voit immédiatement sur un exemple simple : considérons une variable $X \sim \mathcal{N}(0, 1)$ et les deux suites $(Y_n)_{n \in \mathbb{N}}$ et $(Z_n)_{n \in \mathbb{N}}$ définies respectivement par $Y_n = X$ et $Z_n = -X$, alors comme $-X \sim X \sim \mathcal{N}(0, 1)$, on a $Y_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$, $Z_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$ mais $Y_n + Z_n = 0$ ne converge pas en loi vers $2X$.

Comme le fait de connaître les lois marginales ne permet pas de connaître la loi du couple, il n'est pas étonnant que la convergence de chacune des lois marginales ne dise rien sur la convergence du couple. Néanmoins, lorsqu'une des v.a. limites est déterministe, la loi du couple limite est alors entièrement déterminée par la loi de l'autre v.a. et, dans ce cas, on obtient également la convergence en loi pour la suite des couples, comme l'indique le lemme suivant.

Lemme 3.1.10 (Slutsky). Soit deux suites de v.a. $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ définies sur un même espace de probabilité. Si $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une v.a. X et si $(Y_n)_{n \in \mathbb{N}}$ converge en loi vers une constante c , alors le couple $(X_n, Y_n)_{n \in \mathbb{N}}$ converge en loi vers le couple (X, c) . En particulier, $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X + c$ et $X_n \cdot Y_n \xrightarrow[n \rightarrow \infty]{\text{loi}} cX$.

Démonstration. On va utiliser les fonctions caractéristiques et commencer par séparer les variables : pour tout $(t, s) \in \mathbb{R}^2$,

$$\begin{aligned} |\Phi_{(X_n, Y_n)}(t, s) - \Phi_{(X, c)}(t, s)| &= \left| \mathbb{E} \left[e^{i(tX_n + sY_n)} \right] - \mathbb{E} \left[e^{i(tX + sc)} \right] \right| \\ &\leq \left| \mathbb{E} \left[e^{i(tX_n + sY_n)} \right] - \mathbb{E} \left[e^{i(tX_n + sc)} \right] \right| + \left| \mathbb{E} \left[e^{i(tX_n + sc)} \right] - \mathbb{E} \left[e^{i(tX + sc)} \right] \right| \\ &\leq \left| \mathbb{E} \left[e^{itX_n} (e^{isY_n} - e^{isc}) \right] \right| + |e^{isc}| \left| \mathbb{E} \left[e^{itX_n} \right] - \mathbb{E} \left[e^{itX} \right] \right| \end{aligned}$$

Comme $|e^{iu}| = 1$ pour tout réel u , on obtient pour tout $(t, s) \in \mathbb{R}^2$,

$$|\Phi_{(X_n, Y_n)}(t, s) - \Phi_{(X, c)}(t, s)| \leq \mathbb{E}[|e^{isY_n} - e^{isc}|] + |\Phi_{X_n}(t) - \Phi_X(t)|.$$

La suite $(X_n)_{n \in \mathbb{N}}$ convergeant en loi vers X , on a pour tout $t \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} |\Phi_{X_n}(t) - \Phi_X(t)| = 0.$$

Il reste à contrôler $\mathbb{E}[|e^{isY_n} - e^{isc}|]$. Or, la limite en loi de $(Y_n)_{n \in \mathbb{N}}$ étant constante, $(Y_n)_{n \in \mathbb{N}}$ converge également en probabilité vers c . De plus, la fonction $u \mapsto e^{isu}$ étant continue sur \mathbb{R} , e^{isY_n} converge en probabilité vers e^{isc} . (Le résultat a été vu pour des fonctions à valeurs réelles, pour des fonctions à valeurs complexes, il suffit de travailler coordonnée par coordonnée.)

Ainsi, pour tout $s \in \mathbb{R}$,

$$\forall \epsilon > 0, \mathbb{E}[|e^{isY_n} - e^{isc}|] \leq \epsilon \mathbb{P}(|e^{isY_n} - e^{isc}| < \epsilon) + \mathbb{E}[|e^{isY_n} - e^{isc}| \mathbf{1}_{|e^{isY_n} - e^{isc}| \geq \epsilon}] .$$

Comme p.s. $|e^{isY_n} - e^{isc}| \leq |e^{isY_n}| + |e^{isc}| = 2$, on obtient

$$\forall \epsilon > 0, \mathbb{E}[|e^{isY_n} - e^{isc}|] \leq \epsilon + 2\mathbb{P}(|e^{isY_n} - e^{isc}| \geq \epsilon) .$$

En passant à la limite, on obtient donc

$$\forall \epsilon > 0, 0 \leq \limsup_{n \rightarrow \infty} |\Phi_{(X_n, Y_n)}(t, s) - \Phi_{(X, c)}(t, s)| \leq \epsilon$$

Le résultat étant vrai pour tout $\epsilon > 0$, on obtient bien la convergence de $\Phi_{(X_n, Y_n)}(t, s)$ vers $\Phi_{(X, c)}(t, s)$ pour tout $(t, s) \in \mathbb{R}^2$ et donc la convergence en loi de (X_n, Y_n) vers (X, c) . \square

Comment montrer qu'une suite de v.a. converge pour un type de convergence donné

- On utilise les différentes implications du théorème 3.1.5.
- Pour montrer la convergence en probabilité, les inégalités vues dans la partie précédente sont utiles, on peut également essayer de calculer directement la probabilité.
- Pour montrer la convergence \mathbb{P} -p.s. (cas le plus complexe !), on peut utiliser la loi forte des Grands Nombres (présentée dans la section suivante) pour les moyennes de v.a.i.i.d, la propriété 3.1.6 ou le théorème de Borel-Cantelli.
- Pour montrer la cvg dans \mathbb{L}^p , le théorème de convergence dominée peut être utile.
- Pour montrer la convergence en loi, on peut utiliser le Théorème Central Limite (présenté dans la section suivante), utiliser une des caractérisations vues dans la section sur la convergence en loi ou montrer la convergence en probabilité.

Exemples :

- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.i.i.d. telle que $X_n \sim \mathcal{B}(1/n^2)$. Alors $(X_n)_{n \in \mathbb{N}}$ converge vers 0 dans \mathbb{L}^p pour tout $p \geq 1$, en proba et en loi et \mathbb{P} -p.s.

- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.i.i.d. telle que $X_n \sim \mathcal{B}(1/n)$. Alors $(X_n)_{n \in \mathbb{N}}$ converge vers 0 dans \mathbb{L}^p pour tout $p \geq 1$, en proba et en loi mais pas \mathbb{P} -p.s. (on peut utiliser le second lemme de Borel-Cantelli pour montrer le dernier point)
- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.i.i.d. telle que $X_n/n^2 \sim \mathcal{B}(1/n^2)$. Alors $(X_n)_{n \in \mathbb{N}}$ converge vers 0 \mathbb{P} -p.s. mais pas dans \mathbb{L}^p pour tout $p \geq 1$.
- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.i.i.d. telle que $X_n/\sqrt{n} \sim \mathcal{B}(1/n)$. Alors $(X_n)_{n \in \mathbb{N}^*}$ converge vers 0 dans \mathbb{L}^p pour $p < 2$ mais pas \mathbb{P} -p.s. ni dans \mathbb{L}^p pour tout $p \geq 2$.
- Soit $X \sim \mathcal{B}(1/2)$ et $(X_n)_{n \in \mathbb{N}}$ la suite définie par $X_n = (1 + (-1)^n)/2 - (-1)^n X$ pour tout $n \in \mathbb{N}$. Alors $(X_n)_{n \in \mathbb{N}}$ converge vers X en loi mais pas en probabilité.

3.2 Théorèmes limites pour les v.a.i.i.d.

Pour une suite $(X_n)_{n \in \mathbb{N}^*}$ de v.a., on note \bar{X}_n la moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Nous rappelons dans cette section les deux célèbres théorèmes de convergence décrivant le comportement asymptotique de la suite des moyennes empiriques de v.a.i.i.d. Commençons par un premier résultat simple qui donne l'espérance et la variance de la moyenne empirique.

Proposition 3.2.1. *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. de même loi qu'une v.a. X .*

1. *Si $X \in \mathbb{L}^1$, alors $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X]$.*

2. *Si $X \in \mathbb{L}^2$ et que les v.a. sont indépendantes, alors $\mathbb{V}[\bar{X}_n] = \frac{\mathbb{V}[X]}{n}$.*

Démonstration. Cf Exercice 14 de la feuille de TD 2. □

Cette proposition implique que, sous réserve de l'existence d'un moment d'ordre 2, si les v.a. sont i.i.d., alors (\bar{X}_n) converge dans \mathbb{L}^2 vers l'espérance $\mathbb{E}[X]$.

Avec des calculs plus techniques, cette convergence peut être généralisée : dès que les v.a. admettent une espérance, la moyenne empirique converge vers cette espérance \mathbb{P} -p.s. et dans \mathbb{L}^1 . Ce théorème, appelé Loi Forte des Grands Nombres, est parfois résumé ainsi : «la moyenne empirique converge vers la moyenne théorique».

Théorème 3.2.2 (Loi forte des Grands Nombres,). *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i.i.d. dans \mathbb{L}^1 d'espérance commune m . La moyenne empirique*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

converge \mathbb{P} -p.s. et dans \mathbb{L}^1 vers m .

Démonstration. Nous allons montrer la convergence p.s. dans le cas de v.a. dans \mathbb{L}^4 . Le résultat se généralise ensuite aux v.a. \mathbb{L}^1 , les lecteurs intéressés peuvent regarder, par exemple, le livre *Probabilité* de Barbe et Ledoux pour obtenir les détails de la (ou plutôt d'une) démonstration dans ce cas.

On suppose donc que $X_1 \in \mathbb{L}^4$ et on note $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}[X_1]$. En utilisant l'inégalité de Markov, on obtient

$$\mathbb{P}(|\bar{X}_n - m| > \epsilon) = \mathbb{P}\left(\left|\sum_{i=1}^n (X_i - m)\right|^4 > n^4 \epsilon^4\right) \leq \frac{1}{n^4 \epsilon^4} \mathbb{E}\left[\left|\sum_{i=1}^n (X_i - m)\right|^4\right].$$

Par ailleurs,

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{i=1}^n (X_i - m)\right|^4\right] &= \sum_{i=1}^n \mathbb{E}[(X_i - m)^4] + 3 \sum_{i \neq j} \mathbb{E}[(X_i - m)^2(X_j - m)^2] \\ &+ \sum_{\substack{i,j,k,l \in \{1, \dots, n\} \\ \text{au moins un indice est différent des 3 autres}}} \mathbb{E}[(X_i - m)(X_j - m)(X_k - m)(X_l - m)] \\ &= n\mathbb{E}[(X_1 - m)^4] + 3n(n-1)\sigma^4 \end{aligned}$$

En effet, chaque terme de la seconde somme vaut σ^4 et le dernier terme vaut 0 par indépendance. On a donc (en utilisant que $n(n-1) \leq n^2$),

$$\mathbb{P}(|\bar{X}_n - m| > \epsilon) \leq \frac{\mathbb{E}[(X_1 - m)^4]}{\epsilon^4 n^3} + \frac{3\sigma^4}{\epsilon^4 n^2}.$$

et pour tout $\epsilon > 0$,

$$\sum_{n \geq 1} \mathbb{P}(|\bar{X}_n - m| > \epsilon) < +\infty$$

et, par le critère de la proposition 3.1.6, $(\bar{X}_n)_n$ converge presque sûrement vers m . \square

Un peu d'histoire : Les premiers énoncés, sans démonstration, de la Loi des Grands Nombres remontent au moins au 16^e siècle par le mathématicien italien Cardan. Une première démonstration pour le cas particulier des vaillants de Bernoulli a été présentée par Bernoulli en 1713. L'appellation «Loi des Grands Nombres» semble avoir été popularisée par Poisson dans un traité de 1837. La forme actuelle décrite ci-dessus a été démontrée en 1930 par Kolmogorov.

Il existe de nombreuses généralisations de la loi forte des grands nombres (ou faible, avec une convergence en probabilité) en affaiblissant les hypothèses (indépendance ou loi commune, cf références du cours) mais cette version est la plus classique et la plus utile.

Exemples :

- on considère une suite de vaillants de loi de Bernoulli $\mathcal{B}(p)$ pour $p \in [0, 1]$. Alors $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} p$. Cela signifie, par exemple, que lorsque on lance de nombreuses fois une pièce parfaitement équilibrée, la proportion de «pile» obtenus est proche de 1/2
- Exemples en ligne :
<http://experiences.math.cnrs.fr/Loi-des-grands-nombres-et-theoreme.html>

La Loi des Grands Nombres donne la convergence de la moyenne empirique mais ne dit rien sur la vitesse de convergence, ce qui est un point essentiel pour les applications. Le théorème suivant permet de préciser cette vitesse, sous réserve d'existence d'un second moment.

Théorème 3.2.3 (Théorème central limite). *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i.i.d. dans \mathbb{L}^2 d'espérance commune m et de variance commune σ^2 . Alors*

$$Z_n = \sqrt{\frac{n}{\sigma^2}} (\bar{X}_n - m) \xrightarrow[n \rightarrow \infty]{\text{loi}} N \sim \mathcal{N}(0, 1).$$

Démonstration. CF DM 1 □

Remarquons que $\bar{X}_n = m + \sqrt{\frac{\sigma^2}{n}} Z_n$. Ainsi comme Z_n est «à peu près» une v.a. gaussienne centrée réduite pour n grand, \bar{X}_n est «à peu près» une v.a. gaussienne de moyenne m et de variance σ^2/n . Ce résultat est évident si les v.a. $(X_n)_{n \in \mathbb{N}}$ sont des v.a. gaussiennes mais le théorème central limite indique que le résultat est vrai asymptotiquement *quelle que soit* la loi d'origine des variables tant qu'elle admet un second moment fini.

Un peu d'histoire : le cas particulier des variables de Bernoulli a été énoncé par De Moivre en 1733. Une première démonstration fut proposée par Laplace en 1809 et la version actuelle et la démonstration basée sur l'utilisation de la fonction caractéristique sont issus des travaux du hongrois Polyà en 1920 et des travaux du finlandais Lindeberg et du français Lévy en 1935.

Application : Ce théorème permet de construire des intervalles de confiance asymptotique pour les moyennes (cf feuille de TD 2 et cours de Statistiques). Il permet également d'obtenir des intervalles de fluctuation. Par exemple supposons que pour tout $n \in \mathbb{N}^*$, $X_n \sim \mathcal{B}(0.5)$, et qu'on note q_β le quantile d'ordre $\beta \in]0, 1[$ de la loi normale $\mathcal{N}(0, 1)$ (ie $\mathbb{P}(N \leq q_\beta) = \beta$ si $N \sim \mathcal{N}(0, 1)$). Alors pour tout $\alpha \in]0, 1[$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(0.5 - \frac{q_{1-\alpha/2}}{2\sqrt{n}} \leq \bar{X}_n \leq 0.5 + \frac{q_{1-\alpha/2}}{2\sqrt{n}}\right) = 1 - \alpha$$

Ainsi, pour $\alpha = 0.05$, $q_{1-\alpha/2} \approx 1,96$ et en prenant $n = 100$,

$$\mathbb{P}(0,4 \leq \bar{X}_{100} \leq 0,6) \approx 0,95.$$

Exemples en ligne :

<http://experiences.math.cnrs.fr/Loi-des-grands-nombres-et-theoreme.html>

Le TCL nous dit que la moyenne empirique se comporte à peu près comme une variable de loi $\mathcal{N}(m, \frac{\sigma^2}{n})$. Il est naturel de se demander alors quelle est l'erreur commise lorsqu'on effectue cette approximation ; était-il raisonnable de prendre $n = 100$ dans l'exemple précédent ? Cette erreur est contrôlée par le théorème suivant lorsqu'il y a existence d'un moment d'ordre 3 pour la loi des variables aléatoires.

Théorème 3.2.4 (Berry-Esseen, 1941-1942). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.i.i.d. de \mathbb{L}^3 , de moyenne m , de variance σ^2 . On note F_n la fonction de répartition de*

la variable Z_n comme définie dans le TCL et Φ la fonction de répartition de la loi gaussienne $\mathcal{N}(0, 1)$ alors il existe une constante $C < 0.5$,

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{C\mathbb{E}[|X_1 - m|^3]}{\sqrt{n}\sigma^3} .$$

Ce théorème est donné pour satisfaire votre curiosité mais ne sera pas dans les sujets d'examen.

Actuellement, la valeur optimale de la constante est $\sim 0,4785$ (Tyurin, 2010). Originellement, Esseen avait obtenu une borne supérieure de 7,59. La borne inférieure donnée par Esseen en 1956, $C > 0,4097$, est toujours la meilleure actuellement.

4 Processus stochastiques (à temps discret)

Les processus stochastiques (à valeurs réels ou vectorielles) servent à modéliser de nombreux phénomènes qui évoluent avec le temps : évolution du nombre d'individus dans une population au cours des années, déplacement de particules, évolution du cours d'une action, évolution du nombre de personnes dans une file d'attente, ...

4.1 Processus stochastiques et filtration

Dans toute cette section, les différentes v.a. considérées sont définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Commençons par quelques rappels sur les tribus.

Définition 4.1.1. On appelle tribu \mathcal{B} sur Ω tout sous-ensemble de $\mathcal{P}(\Omega)$ vérifiant les axiomes suivant :

- $\emptyset \in \mathcal{B}$
- si $A \in \mathcal{B}$, alors $\overline{A} \in \mathcal{B}$
- pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{B} , on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{B}$.

Si une tribu \mathcal{B} est inclus dans une autre tribu \mathcal{A} , on dit que \mathcal{B} est une sous-tribu de \mathcal{A} . On ne considérera dans ce cours que des sous-tribus de \mathcal{F} .

Proposition 4.1.2. Les tribus sont stables par intersection et union dénombrable, par passage au complémentaire.

Démonstration. En exercice. □

Définition 4.1.3. Deux tribus \mathcal{B}_1 et \mathcal{B}_2 sont indépendantes si tout événement $A_1 \in \mathcal{B}_1$ est indépendant de tout événement $A_2 \in \mathcal{B}_2$.

Définition 4.1.4. On dit qu'une variable aléatoire X à valeur dans \mathbb{R}^d est mesurable par rapport à une tribu \mathcal{B} si pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, l'événement $\{X \in A\}$ est dans \mathcal{B} .

Proposition 4.1.5 (Tribu engendrée par des v.a.).

1. On appelle tribu engendrée par la v.a. (ou le vecteur aléatoire) X à valeurs dans \mathbb{R}^d l'ensemble des événements qui ne dépendent que de X :

$$\sigma(X) = \left\{ \{X \in A\} / A \in \mathcal{B}(\mathbb{R}^d) \right\}.$$

Il s'agit de la plus petite tribu rendant $X : \Omega \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mesurable.

2. On appelle tribu engendrée par la famille de v.a. $\{X_i, i \in I\}$ la plus petite tribu rendant les v.a. $X_i : \Omega \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ mesurables. On la note $\sigma(X_i, i \in I)$.

La tribu $\sigma(X_i, i \in I)$ représente l'information contenue dans les variables $(X_i)_{i \in I}$.

Remarques : Deux v.a. X et Y sont indépendantesssi $\sigma(X)$ est indépendante de $\sigma(Y)$. On dira également que X est indépendante d'une tribu \mathcal{B} si $\sigma(X)$ est indépendante de \mathcal{B} au sens de la définition 4.1.3.

Proposition 4.1.6. Une variable aléatoire Y à valeurs dans \mathbb{R}^d est mesurable par rapport à la tribu $\sigma(X_i, i \in I)$, (où I est un ensemble fini ou dénombrable)ssi Y s'écrit comme une fonction (mesurable) des $(X_i)_{i \in I}$.

Démonstration. Admis (cf cours de théorie de la mesure). □

Exemples :

- $Y = \cos(X)$ est mesurable par rapport à $\sigma(X)$ ou même par rapport à $\sigma(X^2)$ car $Y = \cos(\sqrt{X^2})$ mais pas par rapport à $\sigma(\mathbf{1}_{X>0})$.
- L'événement $\{X_1 \geq 1\}$ appartient à $\sigma(X_1)$. L'événement $\{X_1 > X_2, X_1 X_2 > 0\}$ appartient $\sigma(X_1, X_2)$ mais pas à $\sigma(X_1)$ en général. En effet, si je connais la réalisation de la variable X_1 , je peux dire si elle est plus grande que 1 et donc si l'événement $\{X_1 \geq 1\}$ est réalisé. Par contre, si je ne dispose d'aucune information sur X_2 , je serais incapable de dire si $\{X_1 > X_2, X_1 X_2 > 0\}$ est réalisé.

Ces rappels vont nous permettre d'introduire les outils nécessaires à l'étude des processus stochastiques.

Définition 4.1.7 (Processus stochastiques).

1. On appelle processus stochastique (ou aléatoire) à temps discret *une suite de v.a. $(X_n)_{n \in \mathbb{N}}$* .
2. On appelle filtration *une suite croissante de sous-tribus de \mathcal{F} , i.e. une suite de sous-tribus $(\mathcal{F}_n)_{n \in \mathbb{N}}$ telle que pour tout $n \in \mathbb{N}$, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$* .
3. On dit qu'un processus $(X_n)_{n \in \mathbb{N}}$ est adapté à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si pour tout $n \in \mathbb{N}$, X_n est \mathcal{F}_n -mesurable.

On voit immédiatement que

Proposition 4.1.8. L'ensemble des processus stochastiques à valeurs dans \mathbb{R}^d , $d \in \mathbb{N}^*$, et définis sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ est un \mathbb{R} -espace vectoriel.

Une question naturelle dans l'étude de processus stochastiques est d'estimer la probabilité d'événements faisant intervenir le futur du processus lorsqu'on dispose déjà d'informations sur le passé et le présent du processus. Cette notion d'informations disponibles à un instant donné est donc un point essentiel de la modélisation. Dans ce cadre, elle est représentée par la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$: la sous-tribu \mathcal{F}_n représente l'information à disposition à l'instant n . Le fait que la suite soit croissante signifie qu'il n'y a pas de perte d'informations au cours du temps. Finalement, dire qu'un processus $(X_n)_{n \in \mathbb{N}}$ est adapté à une filtration signifie que l'on dispose de suffisamment d'informations à chaque instant n dans \mathcal{F}_n pour connaître la valeur exacte X_n du processus à cet instant.

Exemples :

- Exemple essentiel de filtration : la filtration naturelle $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$ associée à un processus $(X_n)_{n \in \mathbb{N}}$ est définie par $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$ pour tout $n \in \mathbb{N}$. Le processus $(X_n)_{n \in \mathbb{N}}$ est toujours adapté à $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$.

- On considère un processus $(X_n)_{n \in \mathbb{N}}$.
 - On pose pour tout $n \in \mathbb{N}$, $Y_n = |X_n|$. Alors le processus $(X_n)_{n \in \mathbb{N}}$ n'est pas a priori adapté à $(\mathcal{F}_n^Y)_{n \in \mathbb{N}}$ car on a perdu l'information du signe, alors que le processus $(Y_n)_{n \in \mathbb{N}}$ est bien adapté à $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$.
 - Le processus $(Z_n)_{n \in \mathbb{N}}$ défini par $Z_n = X_{n+1} - X_n$ pour tout $n \in \mathbb{N}$ n'est pas adapté à $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$ car Z_n dépend de la v.a. X_{n+1} qui n'est pas a priori mesurable par rapport à \mathcal{F}_n^X .
- Exemple important de processus : *marche aléatoire simple sur \mathbb{Z}* . Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i.i.d. telle que $\mathbb{P}(X_1 = 1) = p$ et $\mathbb{P}(X_1 = -1) = 1 - p$ pour un certain $p \in [0, 1]$. On appelle *marche aléatoire simple sur \mathbb{Z} issue de θ* le processus $(S_n)_{n \in \mathbb{N}}$ défini par

$$\begin{cases} S_0 = 0 \\ \forall n \in \mathbb{N}, \quad S_{n+1} = S_n + X_{n+1} \end{cases}$$

L'horloge naturelle utilisée pour représenter l'évolution du temps, déterministe, n'a pas de raison d'être adaptée à l'évolution du processus considéré et aux phénomènes étudiés. L'idée, formalisée dans la définition suivante, est alors d'introduire des temps aléatoires qui vont eux être directement liés au comportement du processus.

Définition 4.1.9 (Temps d'arrêt).

1. *On appelle temps d'arrêt pour une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une v.a. T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ telle que pour tout $n \in \mathbb{N}$, $\{T \leq n\} \in \mathcal{F}_n$.*
2. *On appelle temps d'arrêt pour un processus $(X_n)_{n \in \mathbb{N}}$ une v.a. T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ telle que pour tout $n \in \mathbb{N}$, $\{T \leq n\} \in \mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$.*

Intuitivement, cela signifie que pour tout instant n , il est possible d'affirmer ou de réfuter l'affirmation $T \leq n$ en ayant à sa disposition uniquement l'information donnée par la tribu \mathcal{F}_n (ou par la connaissance de X_0, \dots, X_n).

Exemple fondamental : temps d'atteinte d'un ensemble borélien A . Le temps

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

est un temps d'arrêt pour le processus $(X_n)_{n \in \mathbb{N}}$. En effet, pour tout entier $n \in \mathbb{N}$,

$$\{T \leq n\} = \{\exists k \leq n, X_k \in A\} = \bigcup_{k=1}^n \{X_k \in A\} \in \mathcal{F}_n^X.$$

Par exemple, le temps d'atteinte d'un point $a \in \mathbb{Z}$ pour la marche simple $(S_n)_{n \in \mathbb{N}}$: $T = \min \{n \in \mathbb{N}, S_n = a\}$.

Attention, si tous les temps d'atteinte sont des temps d'arrêt, les deux notions ne sont pas identiques et il existe des temps d'arrêt qui ne sont pas des temps d'atteinte, par exemple $\tilde{T} = \min \{n \in \mathbb{N}, X_n = 2X_{n-1}\}$. De plus, de nombreuses variables à valeurs entières bien que construites à partir de la suite $(X_n)_{n \in \mathbb{N}}$ ne sont pas des temps d'arrêt par

rapport à la filtration naturelle : par exemple, la variable $\hat{T} = \min \{n \in \mathbb{N}, X_n = 2X_{n+1}\}$ car elle dépend du «futur» du processus.

Remarque : les constantes déterministes sont toujours des temps d'arrêt quelle que soit la filtration considérée. (cf exercice feuille de TD 3)

Proposition 4.1.10. Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une filtration et T une v.a. à valeurs dans $\mathbb{N} \cup \{\infty\}$. Les trois propositions suivantes sont équivalentes :

- (1) T est un t.a. par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ (i.e. $\forall n \in \mathbb{N}, \{T \leq n\} \in \mathcal{F}_n$)
- (2) $\forall n \in \mathbb{N}, \{T = n\} \in \mathcal{F}_n$
- (3) $\forall n \in \mathbb{N}, \{T > n\} \in \mathcal{F}_n$

Démonstration. Cf Feuille de TD 3. □

Proposition 4.1.11. Soit T un temps d'arrêt par rapport à une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. On appelle tribu antérieure à T l'ensemble des événements $A \in \mathcal{F}$ tels que :

$$A \cap \{T \leq n\} \in \mathcal{F}_n .$$

On note cet ensemble \mathcal{F}_T .

Démonstration. Il suffit de vérifier les différents points de la notion de tribu.

- En prenant $A = \emptyset$, il est immédiat que $\emptyset \in \mathcal{F}_T$
- Soit $A \in \mathcal{F}_T$. Alors, pour tout $n \in \mathbb{N}$, $A \cap \{T \leq n\} \in \mathcal{F}_n$ et donc

$$\overline{A} \cap \{T \leq n\} = \overline{A \cup \{T > n\}} = \overline{A \cap \{T \leq n\} \cup \{T > n\}} \in \mathcal{F}_n$$

et donc $\overline{A} \in \mathcal{F}_T$

- Soit $(A_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{F}_T . Alors pour tout $n \in \mathbb{N}$,

$$\{T \leq n\} \cap \bigcup_{k \in \mathbb{N}} A_k = \bigcup_{k \in \mathbb{N}} \{T \leq n\} \cap A_k \in \mathcal{F}_n$$

D'où le résultat. □

Remarque :

- Le temps d'arrêt T est \mathcal{F}_T -mesurable.
- Intuitivement, la tribu \mathcal{F}_T^X représente l'ensemble de l'information disponible lorsqu'on connaît X_0, X_1, \dots, X_T .

Proposition 4.1.12. Soit S et T deux temps d'arrêt par rapport à une même filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Si $S \leq T$ alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Démonstration. Soit $A \in \mathcal{F}_S$. Alors, comme $S \leq T$,

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

□

4.2 Espérance conditionnelle

Nous venons de voir que l'information à disposition pouvait être encodée dans la notion de sous-tribus. Nous étudions maintenant comment utiliser ce fait pour estimer des probabilités ou plus généralement des moments de variables aléatoires en prenant en compte cette information. Vous connaissez déjà la notion de probabilité conditionnellement à un événement de probabilité non nulle : $\mathbb{P}_B(A) = \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ pour un événement B tel que $\mathbb{P}(B) \neq 0$. Ainsi si on prend l'exemple de la marche simple avec $p = 2/3$:

$$\begin{aligned}\mathbb{P}(S_2 = 2) &= 4/9, \quad \mathbb{P}(S_2 = 0) = 2/9 + 2/9 = 4/9, \quad \mathbb{P}(S_2 = -2) = 1/9 \\ \mathbb{P}(S_2 = 2|S_1 = -1) &= 0, \quad \mathbb{P}(S_2 = 0|S_1 = -1) = 2/3, \quad \mathbb{P}(S_2 = -2|S_1 = -1) = 1/3\end{aligned}$$

On voit bien que le conditionnement influe sur la valeur de la variable S_2 . Malheureusement, le problème avec cette notion est qu'elle est fixée à un événement B donné et non à une tribu. Ainsi, si on veut utiliser l'information liée à une variable Y , il faudra étudier les lois selon toutes les valeurs possibles de Y . C'est possible si Y prend un nombre dénombrable de valeurs, mais pas si Y est une v.a. continue. On va donc créer un nouvel outil, appelé *espérance conditionnelle* qui va à la fois englober le cas classique des probabilités conditionnelles et en même temps être applicable avec n'importe quelle information pourvue qu'elle soit représentée par une tribu.

Théorème 4.2.1 (Kolmogorov, 1933). *Soit $X \in \mathbb{L}^1$ une v.a. intégrable et \mathcal{B} une sous-tribu de \mathcal{F} . Alors il existe une v.a. aléatoire $Y \in \mathbb{L}^1(\Omega, \mathcal{B}, \mathbb{P})$ (i.e. Y est \mathcal{B} -mesurable et $\mathbb{E}[|Y|] < \infty$) qui vérifie*

$$\forall A \in \mathcal{B}, \quad \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y].$$

Cette variable est unique \mathbb{P} -p.s, on la note $\mathbb{E}[X|\mathcal{B}]$.

Démonstration. Cf DM 2

□

Attention, malgré son nom, l'espérance conditionnelle est une variable aléatoire ! Intuitivement, il s'agit de la valeur moyenne de X lorsqu'on connaît toute l'information contenue dans \mathcal{B} . Dans le cas où $\mathcal{B} = \sigma(Z_1, \dots, Z_n)$, on note aussi $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X|Z_1, \dots, Z_n]$.

Remarque : Si la variable $X \in \mathbb{L}^2$, l'espérance conditionnelle $\mathbb{E}[X|\mathcal{B}]$ est la meilleure approximation, au sens \mathbb{L}^2 de X qui est \mathcal{B} -mesurable. Ainsi $\mathbb{E}[X|Y]$ est la v.a. $h(Y)$ telle que la fonction h minimise

$$\mathbb{E}[(X - h(Y))^2].$$

Proposition 4.2.2 (Calcul pratique de l'espérance conditionnelle).

- Si Z est une v.a. discrète alors pour toute v.a. $X \in \mathbb{L}^1$,

$$\mathbb{E}[X|Z] = \sum_{z \in Z(\Omega)} \mathbb{E}[X|Z=z] \mathbf{1}_{\{z\}}(Z) = \sum_{z \in Z(\Omega)} \frac{\mathbb{E}[X \mathbf{1}_{Z=z}]}{\mathbb{P}(Z=z)} \mathbf{1}_{Z=z}.$$

Cette formule se généralise à n v.a. discrètes Z_1, Z_2, \dots, Z_n :

$$\mathbb{E}[X|Z_1, \dots, Z_n] = \sum_{z_1 \in Z_1(\Omega)} \cdots \sum_{z_n \in Z_n(\Omega)} \mathbb{E}[X|Z_1 = z_1, \dots, Z_n = z_n] \mathbf{1}_{Z_1=z_1, \dots, Z_n=z_n}$$

- Si (X, Z) admet une densité jointe $p_{X,Z}$ alors pour toute fonction (borélienne) $h : \mathbb{R} \rightarrow \mathbb{R}$ tq $h(X) \in \mathbb{L}^1$,

$$\mathbb{E}[h(X)|Z] = \int_{\mathbb{R}} h(x) \frac{p_{X,Z}(x, Z)}{p_Z(Z)} dx \quad \mathbb{P}\text{-p.s.}$$

Remarquons que $\mathbb{P}(p_Z(Z) > 0) = 1$ et qu'il n'y a donc pas de problème d'existence dans la formule précédente.

Démonstration. Il suffit de vérifier que les constructions proposées satisfont les points du théorème définissant l'espérance conditionnelle. Vérifions-le pour le second point, l'autre est laissé en exercice. Posons

$$\phi(Z) = \int_{\mathbb{R}} h(x) \frac{p_{X,Z}(x, Z)}{p_Z(Z)} dx.$$

La v.a. $\phi(Z)$ est $\sigma(Z)$ -mesurable. Montrons que $\phi(Z) \in \mathbb{L}^1$:

$$\begin{aligned} \mathbb{E}[|\phi(Z)|] &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} h(x) \frac{p_{X,Z}(x, z)}{p_Z(z)} dx \right| p_Z(z) dz \\ &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |h(x)| p_{X,Z}(x, z) dx dz \\ &= \mathbb{E}[|h(X)|] < \infty. \end{aligned}$$

Ainsi, $\phi(Z) \in \mathbb{L}^1$ et pour tout $\{Z \in A\} \in \sigma(Z)$,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{Z \in A} \phi(Z)] &= \int_{\mathbb{R}} \mathbf{1}_{z \in A} \left(\int_{\mathbb{R}} h(x) \frac{p_{X,Z}(x, z)}{p_Z(z)} dx \right) p_Z(z) dz \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x) \mathbf{1}_{z \in A} p_{X,Z}(x, z) dx dz \\ &= \mathbb{E}[\mathbf{1}_{Z \in A} h(X)]. \end{aligned}$$

La variable $\phi(Z)$ vérifie bien tous les points de la définition de $\mathbb{E}[h(X)|Z]$. Par unicité, ces deux v.a. sont donc égales p.s. \square

La formule de l'espérance conditionnelle d'une v.a. X par rapport à une tribu engendrée par une variable discrète Z montre bien que, dans ce cas, $\mathbb{E}[X|Z]$ encode la même information que l'ensemble des probabilités conditionnelles

$$\{\mathbb{P}(\cdot|Z=a), a \text{ tel que } \mathbb{P}(Z=a)>0\}.$$

En effet, l'espérance conditionnelle est alors une v.a. discrète qui est égale à l'espérance de X sous la probabilité $\mathbb{P}(\cdot|Z=a)$ lorsque l'événement $\{Z=a\}$ est réalisé.

La formule dans le cas où l'on conditionne par rapport à une tribu engendrée par une variable continue peut être retenue en remarquant qu'il s'agit de l'analogie du cas discret en remplaçant les probabilités par des densités et la somme par une intégrale.

Exemples :

- Soient X et Y deux v.a.r indépendantes de loi $\mathcal{U}(\{1, \dots, 6\})$. On pose $Z = \max(X, Y)$. Alors

$$\mathbb{E}[X | Z] = \sum_{n=1}^6 \mathbb{E}[X | Z=n] \mathbf{1}_{Z=n}$$

On vérifie facilement que, pour tout $n \in \{1, \dots, 6\}$, $\mathbb{P}(Z = n) = \frac{2n-1}{36}$ et que

$$\mathbb{P}(X = k | Z = n) = \begin{cases} \frac{1}{2n-1} & \text{si } k < n \\ \frac{n}{2n-1} & \text{si } k = n \end{cases}$$

On voit alors que

$$\mathbb{E}[X | Z] = \sum_{n=1}^6 \frac{n(3n-1)}{2(2n-1)} \mathbf{1}_{Z=n} = \frac{Z(3Z-1)}{2(2Z-1)}.$$

- Soit (X, Y) un couple de variables aléatoires de densité

$$f_{(X,Y)} : (x, y) \rightarrow \frac{1}{2\sqrt{2\pi}} e^{-(x-\sqrt{y})^2/2} \mathbf{1}_{[0;2]}(y).$$

On vérifie facilement que Y a pour densité $f_Y : y \rightarrow \mathbf{1}_{[0;2]}(y)/2$ et ainsi

$$\mathbb{E}[X | Y] = \int_{\mathbb{R}} x \frac{\frac{1}{2\sqrt{2\pi}} e^{-(x-\sqrt{Y})^2/2} \mathbf{1}_{[0;2]}(Y)}{\mathbf{1}_{[0;2]}(Y)/2} dx.$$

Comme la v.a. Y est p.s. à valeurs dans $[0; 2]$, $\mathbf{1}_{[0;2]}(Y) = 1$ p.s. et

$$\mathbb{E}[X | Y] = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} e^{-(x-\sqrt{Y})^2/2} dx = \sqrt{Y}.$$

Propriétés de l'espérance conditionnelle

On fixe ici une sous tribu \mathcal{B} et toutes les v.a. considérées sont supposées être définies sur le même espace.

L'idée générale derrière toutes les propriétés qui suivent est de penser l'espérance conditionnellement à \mathcal{B} comme une espérance où tout ce qui \mathcal{B} -mesurable est traité comme une constante. Ces différentes propriétés se démontrent (plus ou moins aisément !) en utilisant la caractérisation de l'espérance conditionnelle présentée dans le théorème 4.2.1.

- (1) *Linéarité* : $\mathbb{E}[\alpha X + \beta Y | \mathcal{B}] = \alpha \mathbb{E}[X | \mathcal{B}] + \beta \mathbb{E}[Y | \mathcal{B}]$ \mathbb{P} -p.s.
- (2) *Positivité* : si $X \geq 0$ \mathbb{P} -p.s., alors $\mathbb{E}[X | \mathcal{B}] \geq 0$ \mathbb{P} -p.s. et donc si $X \geq Y$ \mathbb{P} -p.s., alors $\mathbb{E}[X | \mathcal{B}] \geq \mathbb{E}[Y | \mathcal{B}]$ p.s.
- (3) *Théorème de convergence dominée conditionnel* : Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. Si il existe $Y \in \mathbb{L}^1$ tel que
 - $\forall n \in \mathbb{N}, |X_n| \leq Y$ \mathbb{P} -p.s.
 - $(X_n)_{n \in \mathbb{N}}$ converge \mathbb{P} -p.s. vers une v.a. X .
alors $X \in \mathbb{L}^1$ et $\mathbb{E}[X_n | \mathcal{B}]$ converge vers $\mathbb{E}[X | \mathcal{B}]$ \mathbb{P} -p.s.

- (4) *Inégalité de Jensen conditionnelle* : Soit $X \in \mathbb{L}^1$. Si ϕ est une fonction convexe telle que $\mathbb{E}[|\phi(X)|] < \infty$ alors

$$\phi(\mathbb{E}[X|\mathcal{B}]) \leq \mathbb{E}[\phi(X)|\mathcal{B}] .$$

En particulier,

$$|\mathbb{E}[X|\mathcal{B}]| \leq \mathbb{E}[|X| |\mathcal{B}] .$$

Il existe également un équivalent “conditionnel” pour le théorème de convergence monotone et le lemme de Fatou, et nous renvoyons le lecteur intéressé par un énoncé précis au livre d’Ouvrard, *Probabilités 2*.

Voici ensuite les propriétés plus spécifiques à l’espérance conditionnelle.

- (5) Si X est \mathcal{B} -mesurable, $\mathbb{E}[X|\mathcal{B}] = X$

- (6) $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$

- (7) Si $X \in \mathbb{L}^p$ pour un $p \geq 1$, alors $\mathbb{E}[X|\mathcal{B}] \in \mathbb{L}^p$ également.

- (8) Si $\mathcal{A} \subset \mathcal{B}$ deux sous-tribus,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{A}]|\mathcal{B}] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]|\mathcal{A}] = \mathbb{E}[X|\mathcal{A}] .$$

- (9) Pour toute v.a. Z \mathcal{B} -mesurable et pour toute v.a. $X \in \mathbb{L}^1$ telles que $XZ \in \mathbb{L}^1$, alors

$$\mathbb{E}[XZ|\mathcal{B}] = Z\mathbb{E}[X|\mathcal{B}] .$$

- (10) Si X est indépendante de \mathcal{B} , $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

- (11) Soit X une v.a. indépendante de \mathcal{B} et Y une v.a. \mathcal{B} -mesurable. Pour toute fonction $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ telle que $f(X, Y) \in \mathbb{L}^1$, on a

$$\mathbb{E}[f(X, Y)|\mathcal{B}] = \int_{\mathbb{R}} f(x, Y) \mathbb{P}_X(dx) .$$

Démonstration. On montre simplement quelques propriétés, on renvoie le lecteur intéressé aux références du cours pour la démonstration des autres.

- (1) Il suffit de vérifier que la variable $Z = \alpha\mathbb{E}[X|\mathcal{B}] + \beta\mathbb{E}[Y|\mathcal{B}]$ vérifie les axiomes de l’espérance conditionnelle $\mathbb{E}[\alpha X + \beta Y|\mathcal{B}]$. Par définition, Z est \mathcal{B} -mesurable et \mathbb{L}^1 . De plus, pour tout $A \in \mathcal{B}$, par linéarité de l’espérance classique,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A Z] &= \alpha\mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathcal{B}]] + \beta\mathbb{E}[\mathbf{1}_A \mathbb{E}[Y|\mathcal{B}]] \\ &= \alpha\mathbb{E}[\mathbf{1}_A X] + \beta\mathbb{E}[\mathbf{1}_A Y] = \mathbb{E}[\mathbf{1}_A(\alpha X + \beta Y)] . \end{aligned}$$

La v.a. $\mathbb{E}[\alpha X + \beta Y|\mathcal{B}]$ étant l’unique (p.s.) v.a. de $\mathbb{L}^1(\Omega, \mathcal{B}, \mathbb{P})$ vérifiant l’égalité ci-dessus, on a bien :

$$\mathbb{E}[\alpha X + \beta Y|\mathcal{B}] = \alpha\mathbb{E}[X|\mathcal{B}] + \beta\mathbb{E}[Y|\mathcal{B}] \quad \text{p.s.}$$

(2) Supposons $X \geq 0$ p.s. L'espérance

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}] \mathbf{1}_{\mathbb{E}[X|\mathcal{B}]<0}]$$

est négative en tant qu'espérance d'une v.a. négative. Or, par définition de l'espérance conditionnelle, l'événement $\{\mathbb{E}[X|\mathcal{B}] < 0\}$ est dans \mathcal{B} et donc

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}] \mathbf{1}_{\mathbb{E}[X|\mathcal{B}]<0}] = \mathbb{E}[X \mathbf{1}_{\mathbb{E}[X|\mathcal{B}]<0}] \geq 0.$$

Ainsi cet espérance est nulle. La v.a. $\mathbb{E}[X|\mathcal{B}] \mathbf{1}_{\mathbb{E}[X|\mathcal{B}]<0}$ étant négative ou nulle et d'espérance nulle, elle est donc nulle p.s. et donc p.s. $\mathbb{E}[X|\mathcal{B}] \geq 0$.

En appliquant ce résultat à la v.a. $X - Y$, on obtient la seconde partie du résultat.

(3) Nous allons montrer que

$$\limsup_n \mathbb{E}[X_n|\mathcal{B}] \leq \mathbb{E}[X|\mathcal{B}] \text{ p.s.} \quad (1)$$

Un raisonnement analogue permet d'obtenir la minoration $\liminf_n \mathbb{E}[X_n|\mathcal{B}] \geq \mathbb{E}[X|\mathcal{B}]$ p.s. qui implique le résultat voulu.

Soit $\epsilon > 0$. On considère l'événement

$$A_\epsilon = \left\{ \limsup_n \mathbb{E}[X_n|\mathcal{B}] \geq \mathbb{E}[X|\mathcal{B}] + \epsilon \right\}$$

et montre que $\mathbb{P}(A_\epsilon) = 0$. En passant à la limite, on obtient (1). Par définition de A_ϵ , la minoration suivante est vérifiée :

$$\mathbb{E}\left[\left(\limsup_n \mathbb{E}[X_n|\mathcal{B}] - \mathbb{E}[X|\mathcal{B}]\right) \mathbf{1}_{A_\epsilon}\right] \geq \epsilon \mathbb{P}(A_\epsilon). \quad (2)$$

D'autre part, considérons la suite de v.a. $Z_n = \sup_{k \geq n} \mathbb{E}[X_k|\mathcal{B}] \mathbf{1}_{A_\epsilon}$. Les v.a. X_k étant uniformément bornée par la v.a. Y , les Z_n sont également bornées par $\mathbb{E}[Y|\mathcal{B}] \in \mathbb{L}^1$ d'après le point (2) :

$$\forall n \in \mathbb{N}, |Z_n| \leq \mathbb{E}[Y|\mathcal{B}].$$

De plus, les v.a. Z_n sont p.s. décroissantes et convergent donc p.s. vers la v.a.

$$Z_\infty = \limsup_n \mathbb{E}[X_k|\mathcal{B}] \mathbf{1}_{A_\epsilon} = \limsup_n \mathbb{E}[X_n|\mathcal{B}] \mathbf{1}_{A_\epsilon}.$$

Le théorème de convergence dominée classique nous donne donc l'égalité suivante :

$$\mathbb{E}\left[\limsup_n \mathbb{E}[X_n|\mathcal{B}] \mathbf{1}_{A_\epsilon}\right] = \lim_{n \rightarrow \infty} \mathbb{E}[Z_n].$$

De plus, pour tout $\ell \leq n \in \mathbb{N}$, $X_\ell \leq \sup_{k \geq n} X_k$, donc

$$Z_n = \sup_{k \geq n} \mathbb{E}[X_k|\mathcal{B}] \mathbf{1}_{A_\epsilon} \leq \mathbb{E}\left[\sup_{k \geq n} X_k|\mathcal{B}\right] \mathbf{1}_{A_\epsilon}.$$

L'événement A_ϵ appartenant à \mathcal{B} , la définition de l'espérance conditionnelle nous montre que :

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\mathbb{E}\left[\sup_{k \geq n} X_k|\mathcal{B}\right] \mathbf{1}_{A_\epsilon}\right] = \mathbb{E}\left[\sup_{k \geq n} X_k \mathbf{1}_{A_\epsilon}\right].$$

Ainsi,

$$\begin{aligned} \mathbb{E}\left[\limsup_n \mathbb{E}[X_n|\mathcal{B}] \mathbf{1}_{A_\epsilon}\right] &= \lim_{n \rightarrow \infty} \mathbb{E}[Z_n] \leq \lim_{n \rightarrow \infty} \mathbb{E}\left[\sup_{k \geq n} X_k \mathbf{1}_{A_\epsilon}\right] \\ &= \mathbb{E}[X \mathbf{1}_{A_\epsilon}] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}] \mathbf{1}_{A_\epsilon}], \end{aligned}$$

la dernière ligne étant obtenue par une nouvelle application du TCD puis de la définition de l'espérance conditionnelle. On a donc

$$\mathbb{E} \left[\left(\limsup_n \mathbb{E}[X_n | \mathcal{B}] - \mathbb{E}[X | \mathcal{B}] \right) \mathbf{1}_{A_\epsilon} \right] \leq 0$$

ce qui, avec (2), montre bien que $\mathbb{P}(A_\epsilon) = 0$.

- (4) Admis. Cf Proposition 11.28 de *Probabilités 2*, Ouvrard
- (5) La v.a. X est bien \mathcal{B} -mesurable, \mathbb{L}^1 et vérifie naturellement l'égalité : $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A X]$ pour tout $A \in \mathcal{B}$. Donc $X = \mathbb{E}[X | \mathcal{B}]$.
- (6) Il suffit de prendre $A = \Omega$ dans l'égalité caractérisant l'espérance conditionnelle du théorème 4.2.1.
- (7) Il suffit d'appliquer l'inégalité de Jensen conditionnelle avec la fonction convexe $x \mapsto |x|^p$.
- (8) Montrons que pour $\mathcal{A} \subset \mathcal{B}$ deux sous-tribus,

$$\mathbb{E}[\mathbb{E}[X | \mathcal{B}] | \mathcal{A}] = \mathbb{E}[X | \mathcal{B}],$$

l'autre égalité est une conséquence directe de (5). La v.a. $\mathbb{E}[\mathbb{E}[X | \mathcal{B}] | \mathcal{A}]$ est \mathcal{A} -mesurable et \mathbb{L}^1 par définition. De plus, pour tout $A \in \mathcal{A}$,

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}[\mathbb{E}[X | \mathcal{B}] | \mathcal{A}]] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X | \mathcal{B}]].$$

Et comme $A \in \mathcal{A} \subset \mathcal{B}$, encore une fois par définition de l'espérance conditionnelle,

$$\mathbb{E}[\mathbf{1}_A \mathbb{E}[X | \mathcal{B}]] = \mathbb{E}[\mathbf{1}_A X],$$

ce qui montre l'égalité recherchée.

- (9) On suit le schéma classique en théorie de l'intégration : on montre d'abord le résultat pour Z une indicatrice, puis pour une fonction étagée puis pour une v.a. positive, en utilisant le fait que toute v.a. positive est limite simple croissante de fonctions étagées, puis pour une v.a. quelconque.
- (10) La v.a. (qui est en fait constante !) $\mathbb{E}[X]$ est bien \mathbb{L}^1 et \mathcal{B} -mesurable. De plus, pour tout $A \in \mathcal{B}$ par indépendance,

$$\mathbb{E}[\mathbb{E}[X] \mathbf{1}_A] = \mathbb{E}[X] \mathbb{E}[\mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A],$$

ce qui montre le résultat recherché.

La propriété (11) est proposée à la démonstration en exercice de la feuille de TD 3.

□

5 Chaînes de Markov

Intuitivement, il s'agit de processus aléatoires dont le comportement futur ne dépend que de l'état présent et non du chemin employé par le processus pour y arriver. L'ensemble des valeurs possibles, noté ici E , est appelé *l'espace des états* de la chaîne de Markov. On ne considérera, dans tout ce chapitre, que des chaînes de Markov prenant un nombre fini ou infini dénombrable de valeurs (i.e. l'ensemble E est fini ou infini dénombrable).

5.1 Définition et premières propriétés

On introduit ici les chaînes de Markov par une définition “élémentaire” qui n'est valable que dans les cas fini ou dénombrable. On verra par la suite qu'il existe une définition équivalente à partir de la notion d'espérance conditionnelle. Cette seconde définition fait intervenir des outils probabilistes plus avancés mais a, entre autre intérêt, l'avantage de pouvoir naturellement s'étendre au cas continu en temps et en espace.

Définition 5.1.1. *Une matrice de transition (ou matrice stochastique) est une application $P : E \times E \rightarrow [0, 1]$ telle que*

$$\forall x \in E, \sum_{y \in E} P(x, y) = 1 .$$

Remarques :

- Pour tout $x \in E$, $P(x, \cdot)$ est un germe de probabilité sur E .
- Attention à la terminologie qui peut être trompeuse : si E est infini, la somme précédente est une série et P n'est pas une «vraie» matrice !

Exemple :

$$E = \{1, 2, 3\} \quad \text{et} \quad P = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.5 & 0 & 0.5 \\ 0.3 & 0.7 & 0 \end{pmatrix} .$$

Définition 5.1.2. *Un processus stochastique $(X_n)_{n \in \mathbb{N}}$ à valeurs dans E est une chaîne de Markov (homogène) de matrice de transition P si pour tout $n \in \mathbb{N}$ et pour tout $(x_0, \dots, x_{n+1}) \in E^{n+2}$ tels que $\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) > 0$:*

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) &= \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) \\ &= P(x_n, x_{n+1}) . \end{aligned}$$

Remarques :

- On peut considérer des chaînes de Markov *non homogènes* :

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

i.e. la matrice de transition change à chaque instant n . Sans précision contraire, toutes les chaînes de Markov considérées dans ce cours et les TDs associés seront des chaînes de Markov homogènes.

- On appelle X_0 l'*état initial* de la chaîne de Markov. Attention, il s'agit a priori d'une variable aléatoire.

Notation : Pour une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$,

- on note \mathbb{P}_x et \mathbb{E}_x pour indiquer que la v.a. X_0 est presque sûrement constante égale à la valeur x .
- Si μ est une loi de probabilité sur E , on note \mathbb{P}_μ et \mathbb{E}_μ pour indiquer que la loi de X_0 est μ . La loi μ est alors appelée *loi initiale* de la chaîne de Markov.

Théorème 5.1.3. *La matrice de transition et la loi initiale caractérisent la loi de la chaîne de Markov :*

1. Équations de Chapman-Kolmogorov : *Pour une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de transition P et de loi initiale μ , on a, pour tout $n \in \mathbb{N}$, pour tout $(x_0, \dots, x_n) \in E^{n+1}$,*

$$\mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu(\{x_0\})P(x_0, x_1) \dots P(x_{n-1}, x_n).$$

Inversement, cette égalité caractérise la loi de la chaîne de Markov.

2. Soit P une matrice de transition et μ une loi sur E . Alors il existe un processus $(X_n)_{n \in \mathbb{N}}$ à valeurs dans E qui est une chaîne de Markov de loi initiale μ et de transition P .

Démonstration. La démonstration de ce théorème utilise le théorème d'extension de Kolmogorov qui n'est pas un attendu de ce cours, mais dont nous présentons ici une version pour le lecteur intéressé (une démonstration du théorème d'extension peut être trouvée dans *Probability and Measure* de Billingsley).

Théorème 5.1.4 (Théorème d'extension de Kolmogorov). *Soit une suite $(P_n)_{n \in \mathbb{N}}$ de mesure de probabilités telle que, pour tout $n \in \mathbb{N}$,*

1. P_n est une mesure de probabilités sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$
2. pour tout borélien $A_n \in \mathcal{B}(\mathbb{R}^n)$, on ait $P_{n+1}(A_n \times \mathbb{R}) = P_n(A_n)$.

Alors il existe une unique mesure de probabilités P sur $(\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R})^\mathbb{N})$ telle que pour tout $n \in \mathbb{N}$,

$$\forall A_n \in \mathcal{B}(\mathbb{R}^n), P(A_n \times \mathbb{R}^\mathbb{N}) = P_n(A_n).$$

On peut alors utiliser ce théorème pour démontrer le résultat d'existence et d'unicité des chaînes de Markov.

- Par récurrence sur $n \in \mathbb{N}$, à partir de la définition. Le cas $n = 0$ est évident par la définition de \mathbb{P}_μ . Supposons alors les équations de Chapman-Kolmogorov vérifiées pour un $n \in \mathbb{N}$ quels que soient $(x_0, \dots, x_n) \in E^{n+1}$. Alors, pour tout $(x_0, \dots, x_{n+1}) \in E^{n+2}$, soit $\mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_{n+1} = x_n) = 0$ et le résultat est évident, soit cette probabilité est strictement positive et par hypothèse de récurrence et définition d'une chaîne de Markov,

$$\begin{aligned} & \mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_{n+1} = x_{n+1}) \\ &= \mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \mathbb{P}_\mu(X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= (\mu(\{x_0\}) P(x_0, x_1) \dots P(x_{n-1}, x_n)) P(x_n, x_{n+1}) \end{aligned}$$

Ce qui montre que les équations de Chapman-Kolmogorov sont toujours vérifiées pour $n + 1$. Le résultat est donc vrai pour tout $n \in \mathbb{N}$ d'après le principe de récurrence. Ces équations caractérisent la loi de la chaîne de Markov car elles caractérisent la loi des n premiers termes pour $n \in \mathbb{N}$ quelconque et il y a unicité d'après le théorème d'extension de Kolmogorov. En effet, soit P et \tilde{P} les lois respectives, sur $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\mathbb{N}})$, de deux chaînes de Markov $(X_n)_{n \in \mathbb{N}}$ et $(\tilde{X}_n)_{n \in \mathbb{N}}$ satisfaisant la même équation de Chapman-Kolmogorov, i.e. ayant même probabilités de transition et même loi initiale. Cela signifie que les lois des n premières marginales de $(X_n)_{n \in \mathbb{N}}$ et de $(\tilde{X}_n)_{n \in \mathbb{N}}$ sont les mêmes et donc d'après le théorème d'extension de Kolmogorov, les lois des processus également.

- Posons pour tout $n \in \mathbb{N}$,

$$\forall (x_0, \dots, x_n), P_n(x_0, \dots, x_n) = \mu(\{x_0\}) P(x_0, x_1) \dots P(x_{n-1}, x_n).$$

Chaque P_n induit une mesure de probabilités sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ vérifiant, pour tout borélien $A_n \in \mathcal{B}(\mathbb{R}^n)$, $P_{n+1}(A_n \times \mathbb{R}) = P_n(A_n)$. Grâce au théorème d'extension de Kolmogorov, on sait qu'il existe une loi P sur les suites à valeurs réelles dont les marginales sont données par les P_n . Cela montre bien l'existence de la chaîne de Markov en considérant le processus canonique sur l'espace $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\mathbb{N}}, P)$.

□

Ainsi, si deux chaînes de Markov $(X_n)_{n \in \mathbb{N}}$ et $(\tilde{X}_n)_{n \in \mathbb{N}}$ ont même loi initiale μ et même matrice de transition P , alors pour tout $n \in \mathbb{N}$,

$$\forall (x_0, \dots, x_n) \in E^{n+1}, \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(\tilde{X}_0 = x_0, \dots, \tilde{X}_n = x_n).$$

Exemples :

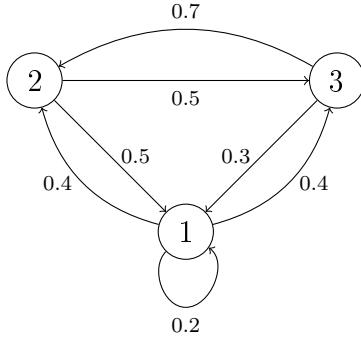
- Représentation d'une chaîne de Markov sous forme de graphe : Considérons la chaîne de Markov définie par la matrice

$$P = \begin{pmatrix} 0.2 & 0.4 & 0.4 \\ 0.5 & 0 & 0.5 \\ 0.3 & 0.7 & 0 \end{pmatrix}.$$

On peut la représenter par le graphe suivant où les noeuds représentent les différents états de la chaîne et les flèches les transitions possibles.

- On appelle *marche aléatoire simple sur \mathbb{Z} issue de $x \in \mathbb{Z}$* , une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ ayant pour état initial $X_0 = x$ et pour transitions, pour un $p \in]0, 1[$ fixé,

$$P(i, j) = \begin{cases} p & \text{si } j = i + 1 \\ 1 - p & \text{si } j = i - 1 \\ 0 & \text{si } |i - j| \neq 1 \end{cases}.$$



Comment montrer qu'un processus est une chaîne de Markov ?

Soit on revient à la définition, soit on utilise le théorème ci-dessous qui est souvent plus pratique.

Théorème 5.1.5. Soit $(X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans E et $(U_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i.i.d. à valeurs dans un espace F (non nécessaire dénombrable). On suppose de plus que la suite $(U_n)_{n \in \mathbb{N}}$ est indépendante de X_0 et qu'il existe une fonction $f : E \times F \rightarrow E$ (mesurable) telle que

$$\forall n \in \mathbb{N}, \quad X_{n+1} = f(X_n, U_{n+1}).$$

Alors le processus $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov.

Démonstration. On vérifie que le processus construit satisfait bien la définition de Chaîne de Markov. \square

Exemple : Soit $(U_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i.i.d. telles que, pour un certain $p \in]0, 1[$, pour tout $n \in \mathbb{N}$, $\mathbb{P}(U_n = 1) = p$ et $\mathbb{P}(U_n = -1) = 1 - p$. On considère alors le processus $(X_n)_{n \in \mathbb{N}}$ défini par

$$\begin{aligned} X_0 &= x \text{ pour un } x \in \mathbb{Z} \text{ fixé,} \\ \forall n \in \mathbb{N}, \quad X_{n+1} &= X_n + U_{n+1}. \end{aligned}$$

Le théorème précédent nous dit que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov. On voit de plus qu'il s'agit d'une marche aléatoire simple sur \mathbb{Z} issue de x .

5.2 Représentation matricielle

Comme l'ensemble E est fini ou dénombrable, on a vu que les probabilités de transition pouvaient être représentées par des "matrices" (ayant potentiellement une infinité de lignes et de colonnes). On peut également identifier une loi de probabilité μ sur E à un "vecteur" ligne $(\mu_x)_{x \in E}$ des poids de probabilité (potentiellement infini) et une fonction $f : E \rightarrow \mathbb{R}$ à un vecteur colonne $(f_x)_{x \in E}$. Ainsi, si une v.a. X suit la loi μ sur E , on peut écrire matriciellement l'espérance de X .

$$\mathbb{E}[f(X)] = \int_E f(x) \mu(dx) = \sum_{x \in E} f(x) \mu(x) = \sum_{x \in E} \mu_x f_x = \mu \cdot f.$$

On peut de plus généraliser le produit matriciel classique à ces matrices stochastiques infinies :

Définition 5.2.1. La matrice produit de deux matrices stochastiques est l'application $PQ : E \times E \rightarrow [0, \infty]$ définie par :

$$\forall x, y \in E, \quad PQ(x, y) := \sum_{z \in E} P(x, z)Q(z, y) .$$

Proposition 5.2.2. Le produit de deux matrices stochastiques est encore une matrice stochastique.

Démonstration. On a bien :

$$\forall x, y \in E, \quad 0 \leq PQ(x, y) \leq \sum_{z \in E} P(x, z) = 1 < \infty$$

et, de plus,

$$\forall x \in E, \quad \sum_{y \in E} PQ(x, y) = \sum_{y \in E} \sum_{z \in E} P(x, z)Q(z, y) = \sum_{z \in E} P(x, z) \sum_{y \in E} Q(z, y) = \sum_{z \in E} P(x, z) = 1 .$$

□

On peut alors définir récursivement P^n :

$$P^0 = I_d \text{ et } \forall n \in \mathbb{N}, \quad P^{n+1} = PP^n .$$

Ce produit a un sens probabiliste :

$$\begin{aligned} \mathbb{P}(X_2 = y | X_0 = x) &= \sum_{z \in E} \mathbb{P}(X_2 = y | X_0 = x, X_1 = z) \mathbb{P}(X_1 = z | X_0 = x) \\ &= \sum_{z \in E} \mathbb{P}(X_2 = y | X_1 = z) \mathbb{P}(X_1 = z | X_0 = x) \\ &= \sum_{z \in E} P(z, y)P(x, z) = P^2(x, y) . \end{aligned}$$

Plus généralement,

Proposition 5.2.3. Pour tout $n \in \mathbb{N}$, $\mathbb{P}(X_n = y | X_0 = x) = P^n(x, y)$.

Démonstration. On itère la formule précédente par récurrence. □

On peut ainsi ramener les calculs de probabilités sur la chaîne de Markov à des problèmes d'algèbre linéaire et utiliser les résultats classiques, dans le cas où l'espace d'état E est fini :

Proposition 5.2.4. Soit $f : E \rightarrow \mathbb{R}$ bornée, $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de matrice de transition P et μ une loi de probabilité sur E . Alors pour tout $n \in \mathbb{N}$, $x \in E$;

1. $\mathbb{P}_\mu(X_n = x) = (\mu P^n)_x$
2. $\mathbb{E}_x[f(X_n)] = (P^n f)_x$
3. $\mathbb{E}_\mu[f(X_n)] = \mu P^n f$

Démonstration. Exercice □

Exemple : On considère la matrice de transition $P = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 1 \end{pmatrix}$ sur l'espace d'état $E = \{A, B\}$. On montre par récurrence que pour tout $n \in \mathbb{N}$, $P^n = \begin{pmatrix} 2^{-n} & 1 - 2^{-n} \\ 0 & 1 \end{pmatrix}$. Ainsi, pour tout $n \in \mathbb{N}$, $\mathbb{P}_A(X_n = A) = P_{11}^n = 2^{-n}$.