M1 HADARMARD

COURS DE PROBABILITÉS

Chargé de cours :

→ Édouard MAUREL-SEGALA Laboratoire de Mathématiques d'Orsay, Université Paris Sud
edouard.maurel-segala@math.u-psud.fr

Chargés de TD :

→ Nathanaël ENRIQUEZ Laboratoire de Mathématiques d'Orsay, Université Paris Sud
nenriquez@u-paris10.fr
→ Pascal MAILLLARD Laboratoire de Mathématiques d'Orsay, Université Paris Sud
pascal.maillard@u-psud.fr

Notes de cours :

→ Antoine BARRIER École Normale Supérieure de Lyon
antoine.barrier@ens-lyon.fr
https://perso.ens-lyon.fr/antoine.barrier/fr/

2017/2018(premier semestre) – Dernière mise à jour : 11 octobre 2020

Sommaire

Première partie

Généralités, outils des probabilités

CHAPITRE 1

Fondement des probabilités

Notre premier objectif est de construire un cadre pour modéliser l'aléa, le hasard. On peut par exemple penser à des jeux de hasard (dès, roulette, ...), des sondages sur des échantillons (par exemple, on observe 1000 tortues dont 453 sont bleues, cela est-il assez significatif pour conclure qu'il y a une proportion inférieure à $1/2$ de tortues bleues ?), ou encore tout un tas de situations où l'on ne maitrise pas toutes les interactions (météo, cours de la bourse, physique statistique, ...).
Ce cadre, c'est l'axiomatique de Kolmogorov ! A un problème impliquant, on va construire un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ pour modéliser l'aléa.

Commençons par rappeler les notions de tribu et de mesure :

Définition 1.1. [tribu]
Soit Ω un ensemble. On dit que $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ est une *tribu* (ou σ-algèbre) si $\mathcal{A}$ contient $\varnothing$ et Ω et est stable par union dénombrable et par passage au complémentaire.

Définition 1.2. [mesure (de probabilité)]
Soit Ω un ensemble non vide et $\mathcal{A}$ une tribu sur Ω. $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$ est une *mesure* si pour toute famille d'événements $(A_n)_{n\in\mathbb{N}}$ deux à deux disjoints, on a $\mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.
Si de plus $\mathbb{P}(\Omega) = 1$, on dit que c'est une mesure *de probabilités*.

On peut alors définir les espaces de probabilité :

Définition 1.3. [espace de probabilité]
Un *espace probabilisable* est la donnée de $(\Omega, \mathcal{A})$ tels que :

1. Ω est un ensemble non vide appelé *univers* et dont les éléments notés ω sont appelés *aléas*, ou *éléments élémentaires*,
2. $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ est une *tribu* dont les éléments sont des *événements*.

Lorsque de plus il est muni d'une mesure de probabilité $\mathbb{P} : \mathcal{A} \longrightarrow [0,1]$, on dit que $(\Omega, \mathcal{A}, \mathbb{P})$ est un *espace de probabilité*.

La donnée de ω donne toute l'information sur le modèle. Un seul ω est réalisé, mais on ne sait pas lequel : tout ce qui dépend de ω est aléatoire. Le reste est déterministe. Les événements représentent des propriétés que l'on peut observer sur l'aléa. Pour un événement A, $\mathbb{P}(A)$ représente les chances qu'a l'aléa ω d'appartenir à A, c'est-à-dire de satisfaire l'événement A.

Exemple 1.4. On jette 3 dès à 6 faces non pipés. On considère $\Omega = [\![1,6]\!]^3$ muni de $\mathcal{A} = \mathcal{P}(\Omega)$ et de $\mathcal{P}$ la probabilité uniforme : $\mathcal{P}(\{\omega\}) = \frac{1}{6^3}$ pour tout $\omega \in \Omega$.

Considérons par exemple l'événement $A =$ "les 3 dès sont pairs" $= \{2,4,6\}^3$. On peut calculer :

$$\mathbb{P}(A) = \frac{3^3}{6^3} = \frac{1}{2^3} = \frac{1}{16}$$

Remarque 1.5. On ne travaille jamais avec plusieurs espaces de probabilité et plus tard on ne précisera pas lequel, sauf dans le cas des chaînes de Markov où Ω est explicite et il y a toute une famille de probabilités.

Pour résoudre un problème, on procède donc en deux étapes : d'abord on modélise le problème par le choix d'un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ adapté, puis on travaille dans cet espace de probabilité pour le résoudre.

Exemple 1.6.

Problème On jette deux dès à six faces non pipés. Quelle est la probabilité que la somme des deux dès soit paire ?

Modèle On considère l'espace de probabilité $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ où $\Omega = [\![1,6]\!]^2$ et $\mathbb{P}$ est la probabilité uniforme, c'est-à-dire la probabilité définie par :

$$\forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{36}$$

Résolution Notons $A =$ "la somme des dès est paire" $= \{2,4,6\}^2 \cup \{1,3,5\}^2$. Alors :

$$\mathbb{P}(A) = \frac{\mathrm{card}(A)}{36} = \frac{9+9}{36} = \frac{1}{2}$$

Remarque 1.7. Lorsque Ω est fini, on prendra le plus souvent $\mathcal{A} = \mathcal{P}(\Omega)$.

Exemple 1.8. [aiguille de Buffon]

Problème On lance une aiguille de 1 cm sur un parquet de lattes d'épaisseur 1 cm. Quelle est la probabilité que l'aiguille soit à cheval sur deux lattes ?

Modèle On repère l'aiguille par (voir Figure 1.1) :

- $x \in [0,1[$ la distance du centre de l'aiguille au bord inférieur de la latte contenant ce centre,
- $\theta \in [0,\pi[$ l'angle de l'aiguille par rapport aux lattes.

et on considère alors l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ défini par :

- $\Omega = \{(x,\theta) \in [0,1[\times[0,\pi[\}$,
- $\mathcal{A} = \mathcal{B}([0,1[\times[0,\pi[)$,
- $\mathbb{P} = \frac{1}{\pi}\,\mathrm{Leb}_{|[0,1[\times[0,\pi[}$ où Leb est la mesure de Lebesgue sur $\mathbb{R}^2$.

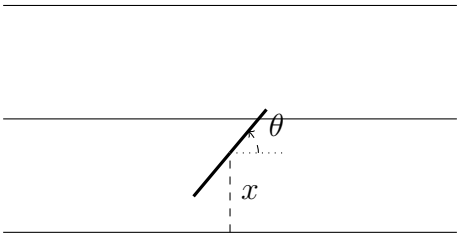


FIGURE 1.1 – Paramètrage du problème de l'aiguille de BUFFON

Résolution Soit $A =$ "l'aiguille touche deux lattes". On a

$$\begin{aligned} A^{\mathsf{c}} &= \text{"l'aiguille ne touche qu'une latte"} \\ &= \left\{(x,\theta) \in \Omega \,|\, \left(x \leq \frac{1}{2} \text{ et } x - \frac{1}{2}\sin(\theta) \geq 0\right) \text{ ou } \left(x \geq \frac{1}{2} \text{ et } x + \frac{1}{2}\sin(\theta) \leq 1\right)\right\} \end{aligned}$$

D'où :

$$\begin{aligned} \mathbb{P}(A^{\mathsf{c}}) &= \mathbb{P}\left(\left\{(x,\theta) \in \Omega \,|\, x \leq \frac{1}{2} \text{ et } x - \frac{1}{2}\sin(\theta) \geq 0\right\}\right) \\ &\qquad + \mathbb{P}\left(\left\{(x,\theta) \in \Omega \,|\, x \geq \frac{1}{2} \text{ et } x + \frac{1}{2}\sin(\theta) \leq 1\right\}\right) \\ &= 2\mathbb{P}\left(\left\{(x,\theta) \in \Omega \,|\, x \leq \frac{1}{2} \text{ et } x - \frac{1}{2}\sin(\theta) \geq 0\right\}\right) \qquad \text{par symétrie} \\ &= 2\int_{[0,\pi[}\int_{[0,1[} \frac{\mathbb{1}_{\frac{1}{2}\sin\theta \leq x < \frac{1}{2}}}{\pi}\, dx\, d\theta \\ &= \frac{1}{\pi}\int_0^\pi 1 - \sin(\theta)\, d\theta = 1 - \frac{2}{\pi} \end{aligned}$$

et finalement $\mathbb{P}(A) = \frac{2}{\pi}$.

Le choix de la tribu n'est pas toujours aussi simple. Regardons l'exemple suivant qui modélise une situation dans laquelle ce choix pose problème :

EXEMPLE 1.9.

Problème On jette un dès à six faces non biaisé une infinité de fois (chaque lancer étant réalisé indépendamment des autres). Quelle est la probabilité que le premier 6 apparaisse au bout d'un nombre pair de lancers ?

Modèle On considère $\Omega = \{1,\dots,6\}^{\mathbb{N}^*}$ et on note $\omega = (\omega_i)_{i\in\mathbb{N}^*}$ un événement élémentaire, où ω_i est le résultat du i-ème lancer.
Pour $k \in \mathbb{N}^*$ et $i_1,\dots,i_k \in \{1,\dots,6\}$, posons

$$A_{i_1,\dots,i_k} = \{\omega \in \Omega \,|\, \forall j \in [\![1,k]\!], \omega_j = i_j\}$$

On veut choisir une tribu de sorte que les $(A_{i_1,\dots,i_k})_{k\in\mathbb{N}^*,1\leq i_1,\dots,i_k\leq 6}$ soient des événements. On pose donc

$$\mathcal{A} = \sigma\left(\{A_{i_1,\dots,i_k} \,|\, k \in \mathbb{N}^* \text{ et } i_1,\dots,i_k \in [\![1,6]\!]\}\right)$$

et on considère $\mathbb{P}$ la mesure de probabilité telle que :

$$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in [\![1,6]\!], \quad \mathbb{P}(A_{i_1,\dots,i_k}) = \frac{1}{6^k} \tag{1.1}$$

On remarque déjà plusieurs difficultés : pourquoi ce choix de tribu ? Existe-t-il une telle probabilité $\mathbb{P}$? Est-elle unique ?

Résolution Soit $P =$ "le premier 6 arrive au bout d'un nombre pair de lancers" puis pour $p \in \mathbb{N}^*$, $P_p =$ "le premier 6 arrive au bout de $2p$ lancers". Alors :

$$\mathbb{P}(P) = \mathbb{P}(\sqcup_{p\in\mathbb{N}^*} P_p) = \sum_{p\in\mathbb{N}^*} \mathbb{P}(P_p) = \sum_{p\in\mathbb{N}^*} \sum_{i_1,\dots,i_{2p-1}\in[\![1,5]\!]} \mathbb{P}(A_{i_1,\dots,i_{2p-1},6}) = \sum_{p\in\mathbb{N}^*} \left(\frac{5}{6}\right)^{2p} \frac{1}{5}$$

On obtient donc $\mathbb{P}(P) = \frac{5}{11}$.

Revenons sur nos choix de probabilité et de tribu :

- Pour le choix de la tribu, il est déraisonnable de ne pas prendre les $(A_{i_1,\dots,i_k})_{k\in\mathbb{N}^*,1\leq i_1,\dots,i_k\leq 6}$, mais choisir une tribu plus grosse que la tribu qu'ils engendrent mène aux mêmes difficultés que définir la mesure de LEBESGUE sur tous les boréliens !
  La tribu $\mathcal{A}$ choisie est appelée *tribu cylindrique*. Nous la réutiliserons plus tard.
- Concernant l'existence de la probabilité $\mathbb{P}$, notons qu'il est parfois difficile de construire une infinité d'événements indépendants (on y reviendra). Cependant ici la construction est directe. Considérons, pour $x \in [0,1[$, son développement 6-adique (qui est unique) :

  $$x = \sum_{i=1}^{\infty} \frac{x_i}{6^i}$$

  où $x_i \in [\![0,5]\!]$ pour $i \in \mathbb{N}^*$ et les $(x_i)_{i\in\mathbb{N}^*}$ sont non tous égaux à 5 à partir d'un certain rang. Considérons alors l'application :

  $$\begin{array}{rccc} \Phi : & [0,1[& \longrightarrow & \Omega \\ & x = \sum_{i=1}^{\infty} \frac{x_i}{6^i} & \longmapsto & (x_i+1)_{i\in\mathbb{N}^*} \end{array}$$

  On vérifie que Φ est mesurable puisque [1] :

  $$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in [\![1,6]\!], \quad \Phi^{-1}(A_{i_1,\dots,i_k}) = \left[\sum_{j=1}^{k} \frac{i_j - 1}{6^j}, \sum_{j=1}^{k} \frac{i_j-1}{6^j} + \frac{1}{6^k}\right[$$

  Posons alors $\mathbb{P}$ la mesure image de la mesure de LEBESGUE par Φ. On a par définition, pour tout $A \in \mathcal{A}$, $\mathbb{P}(A) = \mathrm{Leb}(\Phi^{-1}(A))$ et :

  $$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in [\![1,6]\!], \quad \mathbb{P}(A_{i_1,\dots,i_k}) = \mathrm{Leb}\left(\left[\sum_{j=1}^{k} \frac{i_j - 1}{6^j}, \sum_{j=1}^{k} \frac{i_j-1}{6^j} + \frac{1}{6^k}\right[\right) = \frac{1}{6^k}$$

  Ainsi $\mathbb{P}$ existe.

1. il suffit de le vérifier sur une partie génératrice

- Concernant l'unicité, on va montrer qu'il existe une unique probabilité sur $(\Omega, \mathcal{A})$ telle que

$$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in [\![1,6]\!], \quad \mathbb{P}(A_{i_1,\dots,i_k}) = \frac{1}{6^k}$$

Pour cela on utilise les propriétés de *classes monotones*.

RAPPEL 1.10. Une classe $\mathcal{C}$ est une partie de $\mathcal{A}$. On appelle $\sigma(\mathcal{C})$ la plus petite tribu qui la contient. $\mathcal{C}$ est une tribu si elle est stable par complémentaire et union dénombrable.

DÉFINITION 1.11. [CLASSE MONOTONE]
On dit que $\mathcal{C}$ est une *classe monotone* si :

- $\Omega \in \mathcal{C}$,
- pour tout $A, B \in \mathcal{C}$ tels que $A \subseteq B$, on a $B \setminus A \in \mathcal{C}$,
- pour toute $(A_n)_{n\in\mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ suite croissante d'événements, on a $\bigcup_{n\in\mathbb{N}} \uparrow A_n \in \mathcal{C}$.

Pour une classe $\mathcal{C}$ quelconque, on note $\Lambda(\mathcal{C})$ la plus petite classe monotone contenant $\mathcal{C}$.

REMARQUE 1.12. Une tribu est une classe monotone. La notion de classe monotone possède deux grands intérêts :

- elle est en général assez simple à vérifier,
- si μ et ν sont deux mesures de probabilité alors $\{A \in \mathcal{A} \mid \mu(A) = \nu(A)\}$ est une classe monotone.

Ainsi deux mesures satisfaisant (1.1) coïncident sur $\Lambda\left(\{A_{i_1,\dots,i_k} \mid k \in \mathbb{N}^* \text{ et } i_1, \dots, i_k \in [\![1,6]\!]\}\right)$. En particulier, elles sont égales si

$$\Lambda\left(\{A_{i_1,\dots,i_k} \mid k \in \mathbb{N}^* \text{ et } i_1, \dots, i_k \in [\![1,6]\!]\}\right) = \sigma\left(\{A_{i_1,\dots,i_k} \mid k \in \mathbb{N}^* \text{ et } i_1, \dots, i_k \in [\![1,6]\!]\}\right) = \mathcal{A}$$

C'est en fait bien la cas. Pour le montrer on utilise le résultat suivant :

LEMME 1.13. [LEMME DES CLASSES MONOTONES (THÉORÈME DE DYNKIN)]
Si $\mathcal{C}$ est une classe stable par intersection finie alors $\Lambda(\mathcal{C}) = \sigma(\mathcal{C})$.

En considérant $\mathcal{C} = \{A_{i_1,\dots,i_k} \mid k \in \mathbb{N}^* \text{ et } i_1, \dots, i_k \in [\![1,6]\!]\} \cup \{\varnothing\}$, on vérifie que $\mathcal{C}$ est une classe stable par intersection finie, et donc que le Lemme s'applique, et on en déduit 'unicité de la probabilité.

CHAPITRE 2

Variables aléatoires

Travailler avec un espace de probabilité peut vite devenir fastidieux. On va donc tout faire pour ne plus avoir regarder l'espace de probabilité en détails. ***Pour cela, nous allons utiliser les variables aléatoires****, qui sont un objet essentiel pour s'en affranchir notamment grâce à leur loi.*

Dans la suite, $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace de probabilité et $(E, \mathcal{E})$ désigne un espace probabilisable quelconques.

DÉFINITION 2.1. [VARIABLE ALÉATOIRE]
Une *variable aléatoire* X est une fonction mesurable $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$.
Pour $\omega \in \Omega$, $X(\omega)$ est alors un *élément aléatoire* de E.
Si on a $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on dit que X est une *variable aléatoire réelle*.

Ainsi, la variable aléatoire X est l'observation faite sur l'aléa. Notons qu'une variable aléatoire n'a rien d'aléatoire : c'est une fonction ! Ce qui est aléatoire, c'est toujours le ω qui sera réalisé, et donc c'est en cela que la valeur prise par X est aléatoire !

EXEMPLE 2.2. On jette 2 dès et on considère X la somme des deux dès. X est alors une variable aléatoire :

$$\begin{array}{rccc} X : & \Omega = [\![1,6]\!]^2 & \longrightarrow & \mathbb{R} \\ & (\omega_1, \omega_2) & \longmapsto & \omega_1 + \omega_2 \end{array}$$

EXEMPLE 2.3. Dans une population de 1000 personnes, N veulent voter "Oui" à un référendum. On interroge une personne au hasard et on lui demande son opinion. On voudrait connaître la probabilité qu'il réponde "Oui".

Regardons d'abord des modélisation sans variables aléatoires :

- Tout d'abord, on peut considérer $\Omega = [\![1, 1000]\!]$ tel que les N premières personnes votent "Oui" et les autres "Non" [1]. On prend $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ la probabilité uniforme. Alors :

$$\mathbb{P}(\text{"l'individu tiré répond "Oui""}) = \frac{\text{card}([\![1, N]\!])}{\text{card}([\![1, 1000]\!])} = \frac{N}{1000}$$

1. quitte à réordonner les personnes

- On peut aussi considérer $\Omega = \{\text{"Oui"}, \text{"Non"}\}$ avec $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ telle que $\mathbb{P}(\{\text{"Oui"}\}) = \frac{N}{1000}$ et $\mathbb{P}(\{\text{"Non"}\}) = 1 - \frac{N}{1000}$. Alors :

$$\mathbb{P}(\text{"l'individu tiré répond "Oui""}) = \frac{N}{1000}$$

REMARQUE 2.4. Le choix de $(\Omega, \mathcal{A}, \mathbb{P})$ n'est alors pas unique. De plus, si on interroge une seconde personne, il faut changer Ω, et prendre par exemple $[\![1, 1000]\!]^2$.

Regardons ce qu'il se passe lorsque l'on modélise la situation par une variable aléatoire :

- Considérons $X : \Omega \longrightarrow \{\text{"Oui"}, \text{"Non"}\}$ telle que

$$\mathbb{P}(X = \text{"Oui"}) = \mathbb{P}(\text{"l'individu tiré répond "Oui""}) = \frac{N}{1000}$$

  On raisonne alors de manière à ce que cela marche sur tout $(\Omega, \mathcal{A}, \mathbb{P})$ tel que ce X existe. Désormais, si l'on s'intéresse à une seconde personne, on définit une seconde variable aléatoire $Y : \Omega \longrightarrow \{\text{"Oui"}, \text{"Non"}\}$ telle que $\mathbb{P}(X = \text{"Oui" et } Y = \text{"Oui"}) = \left(\frac{N}{1000}\right)^2$.

Passer par une variable aléatoire permet donc d'oublier l'espace de probabilité sous-jacent.

I. Evénements décrits par des variables aléatoires

EXEMPLE 2.5. Soient $X, Y, (X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires sur Ω. On note :

- $(X \leq 4)$ pour $\{\omega \in \Omega \,|\, X(\omega) \leq 4\}$,
- $(\limsup_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n X_i = +\infty)$ pour $\left\{\omega \in \Omega \,|\, \limsup_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = +\infty\right\}$,
- $(\sum_{i=1}^n X_i < +\infty)$ pour $\{\omega \in \Omega \,|\, \sum_{i=1}^n X_i(\omega) < +\infty\}$.

EXEMPLE 2.6. On jette un dès à six faces non biaisé indépendamment une infinité de fois et on note X le numéro du jet où apparaît pour la première fois "6".
Alors X est une variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$:

$$\begin{array}{rccc} X: & \Omega & \longrightarrow & \mathbb{N}^* \cup \{+\infty\} \\ & \omega & \longmapsto & \inf\{i \in \mathbb{N}^* \,|\, \omega_i = 6\} \end{array}$$

Considérons $\Omega = [\![1, 6]\!]^{\mathbb{N}^*}$ muni de $\mathcal{A}$ la tribu cylindrique et regardons l'événement $(X < +\infty)$. On a pour $n \in \mathbb{N}^*$:

$$\mathbb{P}(X = +\infty) \leq \mathbb{P}(\forall j \leq n, \omega_j \neq 6) = \sum_{i_1,\dots,i_k \in [\![1,5]\!]} \mathbb{P}(\forall j \leq n, \omega_j = i_j) \leq \left(\frac{5}{6}\right)^n \underset{n\to+\infty}{\longrightarrow} 0$$

Donc $\mathbb{P}(X = +\infty) = 0$ et $\mathbb{P}(X < +\infty) = 1$.

REMARQUE 2.7. Si un événement A vérifie $\mathbb{P}(A) = 1$, on note A p.s. (presque sûrement) ou A $\mathbb{P}$- p.p. (presque partout).
On prendra garde de bien différencier $X < +\infty$ (qui est un événement) de $X < +\infty$ p.s. (qui dit que $P(X < +\infty) = 1$, c'est-à-dire qui décrit un fait mathématique) !
On remarquera aussi que $X < +\infty$ p.s. ne signifie pas que X ne prend que des valeurs finies !

PROPOSITION 2.8. [ÉVÉNEMENTS PRESQUE SÛRS]

1. *Soient $(A_n)_{n\in\mathbb{N}}$ une suite d'événements tels que pour tout $n \in \mathbb{N}$, on a A_n p.s..*
   Alors on a $\bigcap_{n\in\mathbb{N}} A_n$ p.s..
2. *Soit A un événement presque sûr. Alors :*

$$\forall B \in \mathcal{A}, \mathbb{P}(A \cap B) = \mathbb{P}(B)$$

PREUVE

1. Comme $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n) = 0$ pour tout $n \in \mathbb{N}$, il vient
$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = 1 - \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n^c\Big) \geq 1 - \sum_{n\in\mathbb{N}} \mathbb{P}(A_n^c) = 1$$
2. Pour tout événement B on a : $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$. Or $\mathbb{P}(A^c \cap B) = 0$ puisque $\mathbb{P}(A^c \cap B) \leq \mathbb{P}(A^c) = 0$.

□

REMARQUE 2.9. Le résultat 1. ne subsiste pas dans le cas d'une intersection non dénombrable !
Soit en effet l'espace de probabilité $([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$. Considérons $A_x = \{x\}$ pour $x \in [0,1]$.
On a $\mathbb{P}(A_x) = \mathrm{Leb}(\{x\}) = 0$ donc $[0,1] \setminus \{x\}$ p.s. pour tout $x \in [0,1]$, mais

$$\mathbb{P}\Big(\bigcap_{x\in[0,1]} ([0,1] \setminus \{x\})\Big) = \mathbb{P}(\varnothing) = 0$$

II. Tribu à partir d'une variable aléatoire

On va voir que la construction de tribus définies par des variables aléatoires est cruciale !

Soit $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire mesurable. On se demande quelle est la plus petite sous-tribu de $\mathcal{A}$ qui rend X mesurable ? Notons-là $\mathcal{B}$.

Pour tout $C \in \mathcal{E}$, $\mathcal{B}$ doit contenir $X^{-1}(C)$. Donc $X^{-1}(\mathcal{E}) = \{X^{-1}(C) \mid C \in \mathcal{E}\} \subseteq \mathcal{B}$.
Par ailleurs, on vérifie facilement que $X^{-1}(\mathcal{E})$ est une tribu, et donc $\mathcal{B} = X^{-1}(\mathcal{E})$.

DÉFINITION 2.10. [TRIBU ENGENDRÉE PAR UNE VARIABLE ALÉATOIRE]
Soit $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire. On note $\sigma(X)$ la plus petite tribu de $\mathcal{A}$ qui rend X mesurable, c'est-à-dire $\sigma(X) = X^{-1}(\mathcal{E})$, et on l'appelle *tribu engendrée* par X.

REMARQUE 2.11. C'est la bonne manière de penser les tribus : une tribu est l'ensemble des propriétés qu'on peut évaluer sur l'aléa au travers de l'observation de X.

PROPOSITION 2.12. *Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. On a :*

$$X : (\Omega, \mathcal{B}) \longrightarrow (E, \mathcal{E}) \text{ est mesurable} \quad \Longleftrightarrow \quad \sigma(X) \subseteq \mathcal{B}$$

REMARQUE 2.13.

- Toute sous-tribu $\mathcal{B}$ de $\mathcal{A}$ peut apparaître comme la tribu engendrée par une variable aléatoire, puisque $\mathcal{B} = \sigma(\mathbb{1}_{\mathcal{B}})$.
- La tribu engendrée est utilisée dans des situations très diverses. On est contraint dans le cadre d'un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ de définir des tribus pas trop grosses (pour ne pas avoir d'informations inutiles) et $\sigma(.)$ est toujours une sous-tribu de $\mathcal{A}$ de "bonne taille" (c'est-à-dire contenant l'information nécessaire de la variable aléatoire, mais pas plus).
- Si A est un événement, on a $\sigma(A) = \{\varnothing, A, A^c, \Omega\}$.

III. Indépendance

*On quitte la théorie de la mesure pour faire des probabilités avec la notion d'*indépendance*, dont on va rappeler brièvement les définitions.*

DÉFINITION 2.14. [INDÉPENDANCE DE TRIBUS]
Soient $(\mathcal{B}_i)_{i\in I}$ des sous-tribus de $\mathcal{A}$. On dit que les $(\mathcal{B}_i)_{i\in I}$ sont *indépendantes* si :

$$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in I \text{ tous distincts}, \forall B_1 \in \mathcal{B}_{i_1}, \dots, B_k \in \mathcal{B}_{i_k}, \quad \mathbb{P}\Big(\bigcap_{j=1}^{k} B_j\Big) = \prod_{j=1}^{k} \mathbb{P}(B_j)$$

DÉFINITION 2.15. [INDÉPENDANCE D'ÉVÉNEMENTS]
Soient $(B_i)_{i\in I}$ des événements de $\mathcal{A}$. On dit que les $(B_i)_{i\in I}$ sont *indépendants* si :

$$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in I \text{ tous distincts}, \quad \mathbb{P}\Big(\bigcap_{j=1}^{k} B_{i_j}\Big) = \prod_{j=1}^{k} \mathbb{P}(B_{i_j})$$

DÉFINITION 2.16. [INDÉPENDANCE DE VARIABLES ALÉATOIRES]
Soient $(X_i)_{i\in I}$ des variables aléatoires. On dit que les $(X_i)_{i\in I}$ sont *indépendantes* si les tribus $(\sigma(X_i))_{i\in I}$ le sont.

REMARQUE 2.17.

1) $(X_i : (\Omega, \mathcal{A}) \longrightarrow (E_i, \mathcal{E}_i))_{i\in I}$ sont indépendantes si et seulement si $\mathbb{P}(X \in \bigcap_{j=1}^{k} C_j) = \prod_{j=1}^{k} \mathbb{P}(X \in C_j)$ pour tout $k \in \mathbb{N}^*$, $i_1, \dots, i_k \in I$ tous distincts et $C_1 \in \mathcal{E}_{i_1}, \dots, C_k \in \mathcal{E}_{i_k}$.
2) Les $(A_i)_{i\in I}$ sont indépendants si et seulement si les $(\sigma(A_i))_{i\in I}$ sont indépendantes si et seulement si les $(\mathbb{1}_{A_i})_{i\in I}$ sont indépendantes.
3) L'indépendance est une propriété sur un nombre fini d'objets. Si I est infini, les $(B_i)_{i\in I}$ sont indépendants si et seulement si toute sous-famille finie l'est.

Exemple 2.18. On lance deux dès à six faces non biaisés et on pose :

$$\begin{cases} A = \text{"le premier chiffre tiré est pair"} \\ B = \text{"le second chiffre tiré est pair"} \\ C = \text{"les deux chiffres tirés sont égaux"} \end{cases}$$

On considère lespace de probabilité $\Omega = [\![1,6]\!]^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ la probabilité uniforme. On a :

$$\begin{cases} \mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2} \\ \mathbb{P}(C) = \frac{1}{6} \\ \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{12} = \mathbb{P}(A)\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C) \\ \mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B) \\ \mathbb{P}(A \cap B \cap C) = \frac{1}{12} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \end{cases}$$

Ainsi $\{A, B\}$, $\{A, C\}$ et $\{B, C\}$ sont des systèmes d'événements indépendants, mais pas $\{A, B, C\}$!

Remarque 2.19.

- L'indépendance a lieu lorsque l'on ne peut rien déduire sur l'un des objets à partir des autres. Ici, A est indépendant de C car savoir que le premier chiffre tiré est pair n'influe pas sur les chances que les deux dès soient égaux. Par contre savoir A et C implique B, donc on n'a pas indépendance des trois événements.
- Pour deux événements A et B tels que $\mathbb{P}(B) \neq 0$, on a que A et B sont indépendants si et seulement si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$: le fait de savoir que B est réalisé n'influe pas sur A. **[à définir]**

IV. Loi d'une variable aléatoire

La dernière étape avant de se débarrasser de notre espace de probabilité est de trouver un substitut à $\mathbb{P}$. ***C'est le rôle de la loi d'une variable aléatoire.***

Définition 2.20. [Loi d'une variable aléatoire]
Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire. Sa *loi* μ_X est la mesure image de $\mathbb{P}$ par X, c'est-à-dire la mesure de probabilité sur $(E, \mathcal{E})$ définie par :

$$\forall A \in \mathcal{E}, \quad \mu_X(A) = \mathbb{P}(X \in A)$$

Remarque 2.21. Si μ est une probabilité sur $(E, \mathcal{E})$ alors il existe une variable aléatoire de loi μ. Considérer en effet $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$ et X l'identité sur Ω.

Lorsque X a pour loi la mesure de probabilité μ, on note $X \sim \mu$.

DÉFINITION 2.22. [ESPÉRANCE]

- Soit $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ une variable aléatoire. On définit l'*espérance* de X par
$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) \in [0, +\infty]$$
- Soit $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire telle que $\mathbb{E}[|X|] < +\infty$. On définit de même l'*espérance* de X par $\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) \in \mathbb{R}$.
- Soit $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une variable aléatoire telle que $\mathbb{E}[\|X\|] < +\infty$. On définit l'*espérance* de X par $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ où $X = (X_1, \dots, X_d)$.

  et on définit l'*espérance* de X lorsque $\mathbb{E}[\|X\|] < +\infty$ (c'est-à-dire lorsque $\forall 1 \leq i \leq d, \mathbb{E}[|X_i|] < +\infty$) par $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$.

REMARQUE 2.23. Pour le dernier point, la norme n'est pas spécifiée puisque toutes les normes sont équivalentes. Notons que la condition revient à ce que $\mathbb{E}[|X_i|] < +\infty$ pour tout $i \in [\![1, d]\!]$.

EXEMPLE 2.24. [QUELQUES LOIS DE VARIABLES ALÉATOIRES]
On distingue au moins 2 catégories de lois :

1) le cas *discret*, lorsqu'il existe A dénombrable (ou fini) tel que $X \in A$ p.s.,
2) le cas *continu*, lorsque X est à valeurs dans $\mathbb{R}^d$ et que sa loi admet une densité par rapport à la mesure de LEBESGUE.

Donnons d'abord quelques exemples de ses deux catégories de lois avant de se demander s'il peut exister d'autres catégories de lois :

1) **<u>Lois discrètes</u> : $X \in A$ p.s. avec A fini ou dénombrable.**
   a) <u>Variable aléatoire constante</u> : lorsqu'il existe $c \in \mathcal{E}$ tel que $X = c$ p.s.. On a :
   $$\mathbb{P}(X \in A) = \mathbb{P}(X = c)\mathbb{1}_{c \in A} + \underbrace{\mathbb{P}(X \neq c \cap X \in A)}_{\leq \mathbb{P}(X \neq c) = 0} = \delta_c(A)$$
   où $\delta_c : \begin{array}{ccc} \mathcal{E} & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbb{1}_{c \in A} \end{array}$. Ainsi $\boxed{X = c \text{ p.s.} \iff X \sim \delta_c}$.
   b) Plus généralement, si $A = \{a_i \mid i \in I\}$ avec I fini ou dénombrable, notons $p_i = \mathbb{P}(X = a_i)$.
   Alors automatiquement on a $\sum_{i \in I} p_i = \sum_{i \in I} \mathbb{P}(X = a_i) \underset{\sqcup}{=} \mathbb{P}(A) = 1$. Et alors, pour $B \in \mathcal{E}$:
   $$\begin{aligned} \mu_X(B) = \mathbb{P}(X \in B) &= P(X \in B, X \in A) && \text{car } X \in A \text{ p.s.} \\ &= \mathbb{P}(\sqcup_{a_i \in B} X = a_i) = \sum_{i \in I \mid a_i \in B} \mathbb{P}(X = a_i) \\ &= \sum_{i \in I} p_i \delta_{a_i}(B) = \Big(\sum_{i \in I} p_i \delta_{a_i}\Big)(B) \end{aligned}$$
   Ainsi $\boxed{\mu_X = \sum_{i \in I} p_i \delta_{a_i}}$.

REMARQUE 2.25.
- La loi d'une variable aléatoire discrète est donc donnée par l'ensemble $\{a_i \mid i \in I\}$ où cette variable aléatoire vit presque sûrement et par la suite des $p_i = \mathbb{P}(X = a_i)$.
- Pour une mesure μ, on dit que x est un atome si $\mu(\{x\}) > 0$. Ainsi une mesure discrète a pour atomes tous les $(a_i)_{i\in I}$ tels que $p_i > 0$ et est donc *portée* par ses atomes. On appelle ces mesures des *mesures purement atomiques*.
- Il y a des mesures sans atomes (comme la mesure de LEBESGUE).

Lorsque $A \subseteq \mathbb{R}^d$: soit X une variable aléatoire discrète admettant une espérance (c'est-à-dire $X \geq 0$ p.s. ou $|X|$ intégrable). Alors :

$$\mathbb{E}[X] = \int_{X^{-1}(A)} X(\omega)\, d\mathbb{P}(\omega) + \underbrace{\int_{X^{-1}(A^c)} X(\omega)\, d\mathbb{P}(\omega)}_{=0} = \sum_{i\in I} \int_{X^{-1}(a_i)} \underbrace{X(\omega)}_{a_i}\, d\mathbb{P}(\omega) = \sum_{i\in I} a_i p_i$$

c) Loi de BERNOULLI : lorsque $X \in \{0, 1\}$ p.s., notons $p = \mathbb{P}(X = 1)$.
Alors $\boxed{\mu_X = p\delta_1 + (1-p)\delta_0}$ et on appelle cette loi la *loi de* BERNOULLI *de paramètre* p, que l'on note $\mathcal{B}(p)$. Elle modélise un événement qui a une probabilité p de succès. On a $\mathbb{E}[X] = p$.

d) Loi uniforme : lorsque $A = \{a_i \mid i \in [\![1, n]\!]\}$ et $\forall i \in [\![1, n]\!], \mathbb{P}(a_i) = \frac{1}{n}$.
On a $\boxed{\mu_X = \frac{1}{n}\sum_{i=1}^{n} \delta_{a_i}}$ et on dit que X suit la *loi uniforme* sur A. On a $\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} a_i$.

e) Loi binomiale : soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(p)$. On pose $X = \sum_{i=1}^{n} X_i$, qui modélise le nombre de succès dans une répétition de n expériences indépendantes ayant une probabilité p de réussir. On a $X \in [\![0, n]\!]$ p.s.. Sa loi est ainsi déterminée par les $(\mathbb{P}(X = k))_{0\leq k\leq n}$:

$$\begin{aligned}
\mathbb{P}(X = k) &= \mathbb{P}(\text{"exactement } k\ X_i \text{ valent 1"}) \\
&= \mathbb{P}\Big(\bigsqcup_{i_1<\cdots<i_k} (\forall j \in [\![1, k]\!], X_{i_j} = 1 \text{ et } \forall i \notin \{i_1, \ldots, i_k\}, X_i = 0) \Big) \\
&= \sum_{i_1<\cdots<i_k} \mathbb{P}(\forall j \in [\![1, k]\!], X_{i_j} = 1 \text{ et } \forall i \notin \{i_1, \ldots, i_k\}, X_i = 0) \\
&= \sum_{i_1<\cdots<i_k} \Big(\prod_{j=1}^{k} \mathbb{P}(X_{i_j} = 1) \prod_{j\notin\{i_1,\ldots,i_k\}} P(X_j = 0) \Big) \qquad \text{par indépendance} \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

On dit que X suit la *loi binomiale de paramètre* p, que l'on note $\mathcal{B}(n, p)$.
On a $\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$.

f) Loi multinomiale :
On se donne $k \geq 2$ puis $p_1, \ldots, p_k \in [0, 1]$ tels que $\sum_{i=1}^{k} p_i = 1$. Soient alors $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées de loi caractérisée par $\forall j \in [\![1, k]\!], \mathbb{P}(X_i = j) = p_j$. On note, pour $j \in [\![1, k]\!]$, $N_j = \mathrm{card}(\{1 \leq i \leq n \mid X_i = j\})$.

Si on lance indépendamment des objets dans k boites, et que chaque objet à une probabilité p_j de tomber dans la boite j, alors les $(N_j)_{1\leq j\leq k}$ correspondent au nombre d'objets tombés dans chacune des k boites.
On note $N = (N_1, \dots, N_k)^\top$. On a $N \in \left\{(x_1, \dots, x_k)^\top \in \mathbb{N}^k \mid \sum_{i=1}^k x_i = n\right\}$.
N suit une loi qui généralise la loi binomiale puisque l'on sait que $N_j \sim \mathcal{B}(n, p_j)$. Quelle est la loi de N ?
Fixons $(x_1, \dots, x_k) \in \mathbb{N}^k \mid \sum_{i=1}^k x_i = n$. On a l'égalité des deux événements :

$$\left(N = (x_1, \dots, x_k)^\top\right) = \left(\bigsqcup_{\substack{(I_j)_{1\leq j\leq k} \text{ partition de } [\![1,n]\!] \\ |I_j|=x_j}} (\forall j \in [\![1,k]\!], \forall i \in I_j, X_i = j) \right)$$

D'où :

$$\begin{aligned}
\mathbb{P}(N = (x_1, \dots, x_k)^\top) &= \sum_{\substack{(I_j)_j \text{ part} \\ |I_j|=x_j}} \mathbb{P}(\forall j \in [\![1,k]\!], \forall i \in I_j, X_i = j) \\
&= \sum_{\substack{(I_j)_j \text{ part.} \\ |I_j|=x_j}} \prod_{j=1}^k \prod_{i\in I_j} p_j = \sum_{\substack{(I_j)_j \text{ part.} \\ |I_j|=x_j}} \prod_{j=1}^k p_j^{x_j} && \text{par indépendance} \\
&= \binom{n}{x_1}\binom{n-x_1}{x_2}\dots\binom{n-x_1-\dots-x_{k-1}}{x_k} \times \prod_{j=1}^k p_j^{x_j} \\
&= \frac{n!}{x_1!\dots x_k!}\prod_{j=1}^k p_j^{x_j}
\end{aligned}$$

On note $\binom{n}{x_1,\dots,x_k} = \frac{n!}{x_1!\dots x_k!}$, appelé *coefficient multinomial* et on dit que N suit la *loi multinomiale de paramètre* $(n, (p_1, \dots, p_k))$.

g) Loi de POISSON : Soient $\lambda > 0$ et $X \in \mathbb{N}$ p.s. telle que :

$$\forall k \in \mathbb{N}, \mathbb{P}(X = k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}$$

Alors on dit que X suit une *loi de* POISSON *de paramètre* λ, notée $\mathcal{P}(\lambda)$.

h) Loi géométrique : soient $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(p)$ (on montrera l'existence d'une telle suite plus tard). On pose $X = \inf\{i \in \mathbb{N}^* \mid X_i = 1\}$, avec la convention $\inf \varnothing = +\infty$. On modélise ainsi le nombre d'expériences à réalise pour obtenir un premier succès dans une suite d'expériences indépendantes ayant une probabilité $p \in]0,1[$ de réussir. On a $X \in \mathbb{N}^* \cup \{+\infty\}$ p.s. et :

$$\begin{aligned}
\forall k \in \mathbb{N}^*, \mathbb{P}(X = k) &= \mathbb{P}(\forall i \in [\![1, k-1]\!], X_i = 0 \text{ et } X_k = 1) \\
&= \prod_{i=1}^{k-1} \mathbb{P}(X_i = 0)\mathbb{P}(X_k = 1) && \text{par indépendance} \\
&= (1-p)^{k-1}p
\end{aligned}$$

Puis, pour $n \in \mathbb{N}$:

$$\mathbb{P}(X = +\infty) \leq \mathbb{P}(\forall i \in [\![1, n]\!], X_i = 0) = (1-p)^n \underset{n\to+\infty}{\longrightarrow} 0$$

Donc $\mathbb{P}(X = +\infty) = 0$.
Ainsi $X \in \mathbb{N}^*$ p.s.. On dit que X suit la *loi géométrique de paramètre* p, notée $\mathcal{G}(p)$.

2) **Lois continues** : on dit que X a une loi de densité $f : \mathbb{R}^d \longrightarrow \mathbb{R}^+$ mesurable par rapport à la mesure de LEBESGUE si pour tout A mesurable de $\mathbb{R}^d$, on a $\mathbb{P}(X \in A) = \int_A f(x)\, dx$ (intégration par rapport à la mesure de LEBESGUE sur $\mathbb{R}^d$). Forcément, on a $\int_{\mathbb{R}^d} f(x)\, dx = 1$ et f est définie Leb-p.p..
Dans ce cas, la loi de X est entièrement déterminée par f.

   a) Loi exponentielle : soit X à valeurs dans $\mathbb{R}$ ayant pour densité $f : x \longmapsto \lambda\, \mathrm{e}^{-\lambda x}\, \mathbb{1}_{x>0}$ où $\lambda > 0$ est fixé. Alors on dit que X suit la *loi exponentielle de paramètre* λ, notée $\mathcal{E}(\lambda)$.

   b) Loi uniforme : soit X à valeurs dans $\mathbb{R}^d$ ayant pour densité $\frac{\mathbb{1}_{.\in D}}{\mathrm{Leb}(D)}$ pour un mesurable D tel que $0 < \mathrm{Leb}(D) < +\infty$. Alors on dit que X suit la *loi uniforme sur* D, notée $\mathcal{U}(D)$.

   c) Loi gaussienne : soit X à valeurs dans $\mathbb{R}$ ayant pour densité $f : x \longmapsto \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$. . Alors on dit que X suit la *loi gaussienne centrée réduite*, notée $\mathcal{N}(0,1)$.
   Dans ce cas, soit $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}^*$. Posons $Y = m + \sigma X$. *Quelle est la loi de* Y *?*
   On a $Y \in \mathbb{R}$ et pour A mesurable de $\mathbb{R}$:

$$\begin{aligned}\mathbb{P}(Y \in A) = \mathbb{P}(m + \sigma X \in A) &= \mathbb{P}\left(X \in \frac{A-m}{\sigma}\right) \\ &= \int_{\frac{A-m}{\sigma}} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, dx = \int_A \frac{\mathrm{e}^{-\frac{(y-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\, dy \qquad \text{CDV } x = \frac{y-m}{\sigma}\end{aligned}$$

   Ainsi Y suit la loi de densité $g : x \longmapsto \frac{1}{\sqrt{2\pi}\sigma}\, \mathrm{e}^{-\frac{(y-m)^2}{2\sigma^2}}$, appelée *loi gaussienne de moyenne* m *et de variance* σ^2, et notée $\mathcal{N}(m, \sigma^2)$.

3) **Y a-t-il d'autres types de lois ?**
Sur $\mathbb{R}$, considérons $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{E}(\lambda)$. μ est une mesure de probabilité ni discrète, ni continue.
On a pour A mesurable de $\mathbb{R}$:

$$\mu(A) = \frac{1}{2}\mathbb{1}_{0\in A} + \frac{1}{2}\int_A \lambda\, \mathrm{e}^{-\lambda x}\, \mathbb{1}_{x>0}\, dx$$

On se demande s'il existe d'autres types de mesures de probabilité.

THÉORÈME 2.26. [THÉORÈME DE DÉCOMPOSITION DE LEBESGUE]
Soit μ une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors μ se décompose sous la forme :

$$\mu = \alpha\mu_d + \beta\mu_c + \gamma\mu_s$$

avec $\alpha + \beta + \gamma = 1$, et où :

- *μ_d est une mesure discrète,*
- *μ_c est une mesure continue,*
- *μ_s est une mesure singulière, c'est-à-dire qui n'a pas d'atomes et telle que*

$$\exists B \in \mathcal{B}(\mathbb{R}) \,|\, \mu_s(B) = 1 \text{ et } \mathrm{Leb}(B) = 0$$

Cette décomposition est essentiellement unique.

REMARQUE 2.27. Pour une mesure μ à densité f, si $\mathrm{Leb}(B) = 0$, alors $\mu(B) = 0$. Si de plus $\mu(B^c) = 0$ alors $f = 0$ Leb-p.p. Cette condition s'oppose au fait d'avoir une densité !

EXEMPLE 2.28. [MESURE SINGULIÈRE]
On va construire une mesure singulière en utilisant l'ensemble triadique de CANTOR.

RAPPEL 2.29. **[DÉVELOPPEMENT b-ADIQUE]**
Soit $b \geq 2$ et $x \in [0, 1[$. x admet une unique décomposition sous la forme :

$$x = \sum_{i \geq 1} \frac{x_i}{b^i}$$

où les $x_i \in [\![0, b-1]\!]$ sont non tous égaux à $b-1$ à partir d'un certain rang.

On regarde le cas $b = 3$. Posons $K_0 = [0,1] = \left\{\sum_{i>0} \frac{x_i}{3^i} \,|\, \forall i > 0, x_i \in [\![0,2]\!] \text{ non tous ...}\right\}$, puis :

$$K_1 = \frac{1}{3}K_0 \bigcup \left(\frac{2}{3} + \frac{1}{3}K_0\right) = \left\{\sum_{i>0} \frac{x_i}{3^i} \,|\, x_1 \in \{0,2\} \text{ et } \forall i > 1, x_i \in [\![0,2]\!] \text{ non tous ...}\right\}$$

On définit alors par récurrence, pour $n \in \mathbb{N}^*$:

$$K_n = \frac{1}{3}K_{n-1} \bigcup \left(\frac{2}{3} + \frac{1}{3}K_{n-1}\right) = \left\{\sum_{i>0} \frac{x_i}{3^i} \,|\, x_1, \ldots, x_n \in \{0,2\} \text{ et } \forall i > n, x_i \in [\![0,2]\!] \text{ non tous ...}\right\}$$

On obtient ainsi l'*ensemble triadique de* CANTOR en posant :

$$K = \bigcap_{n \in \mathbb{N}} K_n = \left\{\sum_{i>0} \frac{x_i}{3^i} \,|\, \forall i, x_i \in \{0,2\} \text{ non tous ...}\right\}$$

Il est facile de voir que K est compact dans $[0,1]$. Pour autant, K est de mesure nulle. En effet, pour $n \in \mathbb{N}$:

$$\mathrm{Leb}(K_n) = \mathrm{Leb}\left(\frac{1}{3}K_{n-1}\right) + \mathrm{Leb}\left(\frac{2}{3} + \frac{1}{3}K_{n-1}\right) = \frac{2}{3}\mathrm{Leb}(K_{n-1})$$

Donc par récurrence $\mathrm{Leb}(K_n) = \left(\frac{2}{3}\right)^n$. Comme $\forall n \in \mathbb{N}, K \subset K_n$, on obtient $\mathrm{Leb}(K) = 0$.

Soit alors $X \sim \mathcal{U}([0,1])$. Posons $Y = \phi(X)$ où l'on a posé :

$$\begin{array}{rccc} \phi : & [0,1[& \longrightarrow & K \\ & x = \underbrace{\sum_{i>0} \frac{x_i}{2^i}}_{\text{forme 3-adique}} & \longmapsto & \sum_{i>0} \frac{2x_i}{3^i} \end{array}$$

On vérifie que ϕ est injective, donc pour tout $x \in [0,1[$, $\phi^{-1}(\{x\})$ est un singleton ou est vide.
Alors μ_Y n'a pas d'atomes. En effet :

$$\forall x \in [0,1[, \mu_Y(\{x\}) = \mathbb{P}(Y = x) = \mathbb{P}(\phi(X) = x) = \mathbb{P}(X \in \phi^{-1}(\{x\})) = \mathrm{Leb}(\phi^{-1}(\{x\})) = 0$$

On n'a pas non plus de densité, puisque $\mu_Y(K) = 1$ et $\mathrm{Leb}(K) = 0$.
Ainsi μ_Y est bien *singulière*.

V. Espérance

On rappelle la formule de l'espérance :

$$\mathbb{E}[.] = \int_\Omega . \, d\mathbb{P}(\omega)$$

dont on a précisé à la définition 2.22 les différents cas dans lesquels cette expression à un sens.

On a alors les théorèmes classiques :

LEMME 2.30. [LEMME DE FATOU]
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires positives presque sûrement. Alors on a :

$$\boxed{\mathbb{E}\Big[\liminf_{n\to+\infty} X_n\Big] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n]}$$

THÉORÈME 2.31. [THÉORÈME DE CONVERGENCE MONOTONE]
Soit $(X_n)_{n\in\mathbb{N}}$ une suite croissante presque sûrement de variables aléatoires positives presque sûrement. Alors on a :

$$\boxed{\mathbb{E}\Big[\lim_{n\to+\infty} \uparrow X_n\Big] = \lim_{n\to+\infty} \uparrow \mathbb{E}[X_n]}$$

où $\lim_{n\to+\infty} \uparrow X_n$ est définie presque sûrement.

THÉORÈME 2.32. [THÉORÈME DE CONVERGENCE DOMINÉE]
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles et Z une variable aléatoire telles que :

$$X_n \underset{n\to+\infty}{\longrightarrow} X \text{ p.s.} \quad \forall n \in \mathbb{N}, |X_n| \leq Z \text{ p.s.} \quad \text{et} \quad \mathbb{E}[Z] < +\infty$$

Alors on a :

$$\boxed{\mathbb{E}[X_n] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X]}$$

On veut calculer des espérances sans faire de calcul dans Ω. On va utiliser le théorème :

THÉORÈME 2.33. [LEMME DE TRANSFERT]
Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire et μ une mesure de probabilité sur $(E, \mathcal{E})$. Alors on a :

$$X \sim \mu \iff \forall f : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ mesurable positive}, \mathbb{E}[f(X)] = \int_E f(x)\, d\mu(x)$$

Le premier intérêt de ce théorème est de *calculer des espérances sans passer par Ω* :

EXEMPLE 2.34. Soit $X \sim \mathcal{E}(2)$. Que vaut $\mathbb{E}[\mathrm{e}^X]$? Par le lemme de transfert :

$$\mathbb{E}[\mathrm{e}^X] = \int_{\mathbb{R}} \mathrm{e}^x \; d\mu_X(x) = \int_{\mathbb{R}} \mathrm{e}^x \, 2\,\mathrm{e}^{-2x}\, \mathbb{1}_{x>0}\, dx = 2\int_0^{+\infty} \mathrm{e}^{-x} \; dx = 2$$

Un autre intérêt crucial du théorème est qu'il permet de *trouver la loi d'une variable aléatoire* :

EXEMPLE 2.35. Soient $X \sim \mathcal{E}(\lambda)$ et $\theta > 0$. Quelle est la loi de $Y = \theta X$? Pour f mesurable positive, on a :

$$\begin{aligned}
\mathbb{E}[f(Y)] = \mathbb{E}[f(\theta X)] &= \int_{\mathbb{R}} f(\theta x)\, d\mu_X(x) && \text{par le sens } \Longrightarrow \text{ du théorème} \\
&= \int_{\mathbb{R}} f(\theta x) \lambda\, \mathrm{e}^{-\lambda}\, \mathbb{1}_{x>0}\, dx = \int_{\mathbb{R}} f(y) \frac{\lambda}{\theta}\, \mathrm{e}^{-\frac{\lambda}{\theta} y}\, \mathbb{1}_{y>0}\, dy \\
&= \int_{\mathbb{R}} f(y)\, d\mu_{\mathcal{E}\left(\frac{\lambda}{\theta}\right)}(y)
\end{aligned}$$

Ceci étant valable pour toute fonction f mesurable positive, on en déduit, par le sens $\Longleftarrow$ du théorème, que $Y \sim \mathcal{E}\left(\frac{\lambda}{\theta}\right)$.

PREUVE

$\Longleftarrow$ Pour $B \in \mathcal{E}$, posons $f = \mathbb{1}_B$. On a $\mathbb{E}[\mathbb{1}_{X\in B}] = \mathbb{E}[f(X)]$.
Or, d'une part :

$$\mathbb{E}[\mathbb{1}_{X\in B}] = \int_\Omega \mathbb{1}_{\omega \mid X(\omega)\in B}\, d\mathbb{P}(\omega) = \mathbb{P}(X \in B) = \mu_X(B)$$

Et d'autre part :

$$\mathbb{E}[f(X)] = \int_E f(x)\, d\mu(x) = \int_E \mathbb{1}_{x\in B}\, d\mu(x) = \mu(B)$$

Ceci étant vrai pour tout $B \in \mathcal{E}$, on a $\mu_X = \mu$ donc $X \sim \mu$.

$\Longrightarrow$ Si $X \sim \mu$, pour quels f a-t-on $\mathbb{E}[f(X)] = \int f(x)\, d\mu(x)$?

- C'est vrai pour $f = \mathbb{1}_B$ puisqu'alors cela se réécrit $\mu_X(B) = \mu(B)$,
- Par linéarité, c'est vrai pour f étagée,

– Si f est mesurable positive, il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonction étagées positives telle que $f = \lim\limits_{n\to+\infty} \uparrow f_n$. Alors :

$$\mathbb{E}[f(X)] \underset{\text{TCD}}{=} \lim_{n\to+\infty} \uparrow \mathbb{E}[f_n(X)] = \lim_{n\to+\infty} \uparrow \int_E f_n(x)\,d\mu(x) \underset{\text{TCM}}{=} \int_E f(x)\,d\mu(x)$$

□

REMARQUE 2.36. Dans le théorème, on peut remplacer "pour tout f mesurable positive" par "pour tout f indicatrice" ou par "pour tout f mesurable bornée".

Regardons le cas particulier de $\mathbb{R}^d$:

THÉORÈME 2.37. *Soient X une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et μ une probabilité sur $\mathbb{R}^d$. On a l'équivalence entre :*

(i) $X \sim \mu$,
(ii) $\forall f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *mesurable positive*, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$,
(iii) $\forall f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *mesurable bornée*, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$,
(iv) $\forall f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *continue bornée*, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$,
(v) $\forall f : \mathbb{R}^d \longrightarrow \mathbb{R}$ *continue à support compact*, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$,
(vi) $\forall \lambda \in \mathbb{R}^d, \mathbb{E}[e^{i\langle\lambda|X\rangle}] = \int_{\mathbb{R}^d} e^{i\langle\lambda|x\rangle}\,d\mu(x)$,

PREUVE On a vu (i) $\iff$ (ii) $\iff$ (iii) (plus généralement sur $(E, \mathcal{E})$).
Et on a trivialement (iii) $\implies$ (iv), (iv) $\implies$ (v), (iv) $\implies$ (vi). Montrons donc (v) $\implies$ (i) puis (vi) $\implies$ (v).

– Supposons (v) : montrons que

$$\forall a_1 < b_1, \dots, a_d < b_d, \mathbb{P}(X \in C = [a_1, b_1] \times \cdots \times [a_d, b_d]) = \mu(C)$$

Fixons C. Posons, pour $n \in \mathbb{N}^*$, $g_n(x) = (1 - nd(x, C))$, qui est continue et à support compact.
On a $\|g_n\|_\infty \leq 1$ intégrable pour μ_X et μ (ce sont des mesures de probabilité), et $g_n(x) \underset{n\to+\infty}{\longrightarrow} \mathbb{1}_{x\in C}$.
D'où :

$$\mu_X(C) = \mathbb{E}[\mathbb{1}_{x\in C}] \underset{\text{TCD}}{=} \lim_{n\to+\infty} \mathbb{E}[g_n(X)] \underset{\text{(v)}}{=} \lim_{n\to+\infty} \int g_n\,d\mu \underset{\text{TCD}}{=} \int \mathbb{1}_{x\in C}\,d\mu = \mu(C)$$

μ_X et μ coïncident sur l'ensemble des pavés $\mathcal{C}$, stable par intersection finie, donc elles coïncident sur la tribu engendrée (d'après le lemme 1.13 des classe monotones). Mais comme $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{C})$, on a finalement $\mu = \mu_X$.

– Supposons (vi). Soit f à support compact, on veut approcher f par une fonction périodique. Si $T > 0$ est tel que $\text{supp}(f) \subseteq]-T/2, T/2]^d$, on définit le périodisé f_T de f en posant pour tout $x \in \mathbb{R}^d$, $f_T(x) = f(\overline{x})$ où $\overline{x}$ est l'unique élément de $]-T/2, T/2]^d$ tel que pour tout $1 \leq i \leq d$, on a $x_i - \overline{x}_i \in T\mathbb{Z}$. f_T est continue, T-périodique dans les d directions, $\|f_T\|_\infty = \|f\|_\infty$ et $f_{T|[-\frac{T}{2},\frac{T}{2}]^d} = f$.

On a $\mu\left(\overline{\left[-\frac{T}{2},\frac{T}{2}\right]^d}\right) \underset{T\to+\infty}{\longrightarrow} 0$ et $\mu_X\left(\overline{\left[-\frac{T}{2},\frac{T}{2}\right]^d}\right) \underset{T\to+\infty}{\longrightarrow} 0$. Fixant $\varepsilon > 0$, on peut trouver T tel que :

$$\int \mathbb{1}_{\|x\|_\infty \geq \frac{T}{2}}\, d\mu \leq \varepsilon \quad \Longleftrightarrow \quad \mathbb{P}\left(\|x\|_\infty \geq \frac{T}{2}\right) \leq \varepsilon$$

Et alors :

$$\begin{aligned}\mathbb{E}[f(X)] - \int f\, d\mu &= \mathbb{E}\left[f_T(X)\mathbb{1}_{\|X\|_\infty \leq \frac{T}{2}}\right] - \int f_T(x)\mathbb{1}_{\|x\|_\infty \leq \frac{T}{2}}\, d\mu(x) \\ &= \mathbb{E}[f_T(X)] - \int f_T(x)\, d\mu(x) - \Bigg(\underbrace{\mathbb{E}\left[f_T(X)\mathbb{1}_{\|X\|_\infty \geq \frac{T}{2}}\right]}_{|.|\leq \|f\|_\infty \varepsilon} - \underbrace{\int f(x)\mathbb{1}_{\|x\|_\infty \geq \frac{T}{2}}\, d\mu(x)}_{|.|\leq \|f\|_\infty \varepsilon}\Bigg)\end{aligned}$$

Il existe un polynôme trigonométrique $P_{\varepsilon,T}$ (c'est-à-dire est fonction des $\prod_{i=1}^j \cos\left(\frac{2\pi k_i x_i}{T}\right)$ et $\prod_{i=1}^j \sin\left(\frac{2\pi k_i x_i}{T}\right)$ pour $k_i \in \mathbb{Z}$) tel que $\|P_{\varepsilon,T} - f_T\|_\infty \leq \varepsilon$. D'où :

$$\left|\mathbb{E}[f(X)] - \int f\, d\mu\right| \leq \left|\mathbb{E}[P_{\varepsilon,T}(X)] - \int P_{\varepsilon,T}\, d\mu\right| + \underbrace{\left|\mathbb{E}[(f_T - P_{\varepsilon,T})(X)]\right|}_{\leq \varepsilon \|X\|} + \underbrace{\left|\int (f_T - P_{\varepsilon,T})\, d\mu\right|}_{\leq \varepsilon} + 2\|f\|_\infty \varepsilon$$

La quantité de droite tend vers 0 lorsque $\varepsilon \to 0$.

□

DÉFINITION 2.38. [FONCTION CARACTÉRISTIQUE]
Si X est à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on définit sa *fonction caractéristique* par :

$$\begin{array}{rccc}\phi_X : & \mathbb{R}^d & \longrightarrow & \mathbb{C} \\ & \lambda & \longmapsto & \mathbb{E}\left[\mathrm{e}^{i\langle\lambda|X\rangle}\right]\end{array}$$

REMARQUE 2.39. On a prouvé deux résultats :

- ϕ_X ne dépend que de μ_X, donc on peut définir pour μ probabilité sur $\mathbb{R}^d$:

  $$\phi_\mu(x) = \int \mathrm{e}^{i\langle\lambda|x\rangle}\, d\mu(x)$$

  Et alors $\phi_{\mu_X} = \phi_X$.
- On a la propriété : $\phi_X = \phi_Y \Longrightarrow X$ et Y ont même loi.
  Ainsi $\boxed{\phi_\mu = \phi_\nu \Longleftrightarrow \mu = \nu}$.

EXEMPLE 2.40. Si $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ a une loi de densité $f(x_1, \ldots, x_d)$ par rapport à la mesure de LEBESGUE sur $\mathbb{R}^d$, alors pour tout $i \in [\![1, d]\!]$, X_i a une loi de densité $\int_{\mathbb{R}} f(x_1, \ldots, x_d)\, dx_1 \ldots\, dx_{i-1}\, dx_{i+1} \ldots$ par rapport à la mesure de LEBESGUE sur $\mathbb{R}$.

En effet, cherchons la loi de X_i. Pour toute g mesurable positive, on a :

$$\begin{aligned}\mathbb{E}[g(X_i)] &= \int_{\mathbb{R}^d} g(x_i)\, d\mu_{X_1,\dots,X_d}(x_1,\dots,x_d) && \text{par le lemme de transfert}\\ &= \int_{\mathbb{R}^d} g(x_i) f(x_1,\dots,x_d)\, dx_1 \dots\, dx_d\\ &= \int_{\mathbb{R}} g(x_i) \underbrace{\left(\int_{\mathbb{R}^{d-1}} f(x_1,\dots,x_d)\, dx_1 \dots\, dx_{i-1}\, dx_{i+1} \dots\, dx_d\right)}_{=d\mu_{X_i}(x_i)} dx_i && \text{par le théorème de FUBINI}\end{aligned}$$

REMARQUE 2.41. Si $X = (X_1,\dots,X_d)$ est à valeurs dans un espace produit $(E_1 \times \dots \times E_d, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d)$, μ_X détermine $\mu_{X_1} \times \dots \times \mu_{X_d}$, puisque

$$\begin{aligned}\mu_{X_i}(B) &= \mathbb{P}(X_i \in B) = \mathbb{P}(X \in E_1 \times \dots \times E_{i-1} \times B \times E_{i+1} \times \dots \times E_d)\\ &= \mu_X(E_1 \times \dots \times E_{i-1} \times B \times E_{i+1} \times \dots \times E_d)\end{aligned}$$

Mais *la réciproque est fausse* ! On peut remarquer par exemple :

- si $(X,Y) \sim \mathcal{U}([0,1]^2)$, alors $\mu_X = \mu_Y = \mathcal{U}([0,1])$
- puis si $Z \sim \mathcal{U}([0,1])$, $\tilde{X} = Z$ et $\tilde{Y} = Z$, alors $\mu_{\tilde{X}} = \mu_{\tilde{Y}} = \mathcal{U}([0,1])$, mais $\mu_{(X,Y)} \neq \mu_{(\tilde{X},\tilde{Y})}$.

VI. Espaces L^p

Pour $p \geq 1$, on définit :

$$\boxed{L^p(\mathbb{R}^d) = \left\{X \text{ variable aléatoire à valeurs dans } (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \mid \mathbb{E}[\|X\|^p < +\infty]\right\} / \{X \mid X = 0 \text{ p.s.}\}}$$

REMARQUE 2.42.

- Cela ne dépend pas du choix de la norme !
- Si $p < q$ alors $\boxed{L^q \subseteq L^p}$ (spécifique aux espaces de probabilités !)
  En effet, on a $\boxed{|x|^p \leq 1 + |x|^q}$, donc $\mathbb{E}[\|X\|^p] \leq 1 + \mathbb{E}[\|X\|^q]$.

RAPPEL 2.43.

- $L^p(\mathbb{R}^d)$, muni de la norme $\|.\|_p$ définie par $X \longmapsto \mathbb{E}[\|X\|^p]^{\frac{1}{p}}$, est un espace normé complet.
- On a l'inégalité de HÖLDER : si $\frac{1}{p} + \frac{1}{q} = 1$:

  $$\mathbb{E}[\|XY\|] \leq \|X\|_p \, \|X\|_q$$

  Dans le cas particulier où $p = q = 2$, on retrouve l'inégalité de CAUCHY-SCHWARZ

  $$\mathbb{E}[\|XY\|] \leq \mathbb{E}[\|X\|^2]^{\frac{1}{2}} \mathbb{E}[\|X\|^2]^{\frac{1}{2}}$$

On va revoir trois inégalités classiques, à commencer par :

PROPOSITION 2.44. [INÉGALITÉ DE MARKOV]
Si $X \geq 0$ p.s., alors

$$\boxed{\forall t > 0, \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}}$$

PREUVE On a $t\mathbb{1}_{x \geq t} \leq X$ p.s., donc $\mathbb{E}[t\mathbb{1}_{X \geq t}] \leq \mathbb{E}[X]$, ou encore $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$. □

DÉFINITION 2.45. [VARIANCE, COVARIANCE]
Soient X et Y deux variables aléatoires de $L^2(\mathbb{R})$.

- On définit la *variance* de X par $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.
- On définit la *covariance* de X et Y par $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

PROPOSITION 2.46. [INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV]
Soit X une variable aléatoire réelle. On a

$$\boxed{\forall t > 0, \mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}}$$

PREUVE Par l'inégalité de MARKOV :

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq t^2) \leq \frac{\mathrm{Var}(X)}{t^2}$$

□

PROPOSITION 2.47. [INÉGALITÉ DE JENSEN]
Si ϕ est convexe et $X \in L^1$:

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

REMARQUE 2.48. L'inégalité de JENSEN contient :

- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$ (en prenant la fonction valeur absolue)
- $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ (sous-cas de CAUCHY-SCHWARZ avec $Y = 1$)

PREUVE Une fonction convexe est l'enveloppe supérieure des droites qui la minore.
Considérons, à x fixé : $\sup_{a,b \in \mathbb{R}} \{ax + b \,|\, \forall y, ay + b \leq \phi(y)\}$.

Si a, b sont tels que $\forall y, ay+b \leq \phi(y)$, alors $ax+b \leq \phi(x)$. Donc $\phi(.) \geq \sup_{a,b \in \mathbb{R}} \{a. + b \,|\, \forall y, ay + b \leq \phi(y)\}$.

Par ailleurs, la pente entre x et y est $\frac{\phi(x)-\phi(y)}{y-x}$. Elle est décroissante en y, minorée par les pentes à gauche de x, donc converge vers un α lorsque $y \mapsto x^+$. $D : x \longmapsto \phi(x) + \alpha(y - x)$ est alors la tangente à droite de ϕ en x. Par convexité, elle ne recroise jamais ϕ.
Cette fonction affine apparaît dans le sup et vaut $\phi(x)$ en x. Donc $\phi(.) \leq \sup_{a,b \in \mathbb{R}} \{a. + b \,|\, \forall y, ay + b \leq \phi(y)\}$.

Ainsi $\phi(.) = \sup \{a. + b \,|\, \forall y, ay + b \leq \phi(y)\}$. En particulier :

$$\phi(\mathbb{E}[X]) = \sup_{a,b \in \mathbb{R}} \{a\mathbb{E}[X] + b \,|\, \forall y, ay + b \leq \phi(y)\} = \sup_{a,b \in \mathbb{R}} \{\mathbb{E}[aX + b] \,|\, \forall y, ay + b \leq \phi(y) \leq \mathbb{E}[\phi(X)]\}$$

□

EXEMPLE 2.49. Si $X \sim \mathcal{N}(m, \sigma^2)$ alors $\mathbb{E}[X] = m$ et $\mathrm{Var}[X] = \sigma^2$.
En effet, si $Z \sim \mathcal{N}(0,1)$, on montre (exercice) que $\mathbb{E}[Z] = 0$ et $\mathrm{Var}(Z) = 1$. Puis si $X \sim \mathcal{N}(m, \sigma^2)$, X a la même loi que $m + \sigma Z$ et donc $\mathbb{E}[X] = \mathbb{E}[m + \sigma Z] = m + \sigma\mathbb{E}[Z] = m$ puis $\mathrm{Var}(X) = \mathrm{Var}(m + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2$.

LEMME 2.50.

$$\boxed{\forall \lambda \in \mathbb{R}, \phi_{\mathcal{N}(m,\sigma^2)}(\lambda) = \mathrm{e}^{i\lambda m - \frac{\lambda^2\sigma^2}{2}}}$$

PREUVE

1) On a :

$$\forall \lambda \in \mathbb{R}, \phi_{\mathcal{N}(0,1)}(\lambda) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{i\lambda x - \frac{x^2}{2}}\, dx$$

Par le théorème de dérivation sous le signe $\int$, il vient :

$$\phi'_{\mathcal{N}(0,1)}(\lambda) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} i\, \mathrm{e}^{i\lambda x}\, x\, \mathrm{e}^{-\frac{x^2}{2}}\, dx \underset{\text{IPP}}{=} \frac{1}{\sqrt{2\pi}} \left[i\, \mathrm{e}^{i\lambda x} \left(-\frac{1}{2} \mathrm{e}^{-\frac{x^2}{2}} \right) \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} i^2 \lambda\, \mathrm{e}^{i\lambda x}\, \mathrm{e}^{-\frac{x^2}{2}}\, dx$$

et donc $\forall \lambda \in \mathbb{R}, \phi_{\mathcal{N}(0,1)}(\lambda) = -\lambda \phi'_{\mathcal{N}(0,1)}(\lambda)$. Ainsi :

$$\forall \lambda \in \mathbb{R}, \phi_{\mathcal{N}(0,1)}(\lambda) = \phi_{\mathcal{N}(0,1)}(0)\, \mathrm{e}^{-\frac{\lambda^2}{2}} = \mathrm{e}^{-\frac{\lambda^2}{2}}$$

2) Pour le cas général, il suffit de remarquer que :

$$\forall \lambda \in \mathbb{R}, \phi_{aZ+b}(\lambda) = \mathrm{e}^{ib\lambda}\, \phi_Z(a\lambda)$$

On obtient le résultat en prenant $X = m + \sigma Z$ où $Z \sim \mathcal{N}(0,1)$.

□

CHAPITRE 3

Indépendance

I. Indépendance sur les tribus

On pense à une tribu $\mathcal{B} \subseteq \mathcal{A}$ comme une tribu définie par une variable aléatoire X, c'est-à-dire $\mathcal{B} = \sigma(X)$.

On rappelle la définition :

DÉFINITION 3.1. [INDÉPENDANCE DE TRIBUS]
Soient $(\mathcal{B}_i)_{i\in I}$ des sous-tribus de $\mathcal{A}$. On dit que les $(\mathcal{B}_i)_{i\in I}$ sont *indépendantes* si :

$$\forall k \in \mathbb{N}^*, \forall i_1, \dots, i_k \in I \text{ tous distincts}, \forall B_1 \in \mathcal{B}_{i_1}, \dots, B_k \in \mathcal{B}_{i_k}, \mathbb{P}\Big(\bigcap_{j=1}^{k} B_j\Big) = \prod_{j=1}^{k} \mathbb{P}(B_j)$$

Intuitivement, cela signifie que l'on ne peut pas tirer d'information sur l'un des $\mathcal{B}_i$ à partir des autres.

On souhaite une stabilité de la notion d'indépendance lorsque l'on considère des variables aléatoires fonctions de variables aléatoires indépendantes. On espère en particulier :

- si X, Y, Z sont indépendantes alors $F(X, Y)$ est indépendante de Z.
- si $(X_i)_{i\in I}$ et $(Y_j)_{j\in J}$ sont des variables aléatoires indépendantes alors $F((X_i)_{i\in I})$ et $G((Y_j)_{j\in J})$ sont indépendantes.
- si $(X_i)_{i\in I}$ et $(Y_i)_{i\in I}$ sont des variables aléatoires indépendantes alors $(F_i(X_i, Y_i))_{i\in I}$ forme une famille de variables aléatoires indépendantes.

Ce sont trois cas particulier du *lemme de regroupement* :

LEMME 3.2. [LEMME DE REGROUPEMENT]
Soit $(\mathcal{B}_i)_{i\in I}$ une famille de sous-tribus de $\mathcal{A}$ indépendantes.
Soit $(J_r)_{r\in R}$ une partition de I. Pour $r \in R$, on note $\hat{\mathcal{B}}_r = \sigma(\{\mathcal{B}_i \mid i \in J_r\})$ la plus petite sous-tribu de $\mathcal{A}$ contenant les $(\mathcal{B}_i)_{i\in J_r}$.

Alors $(\hat{\mathcal{B}}_r)_{r\in R}$ est une famille de tribus indépendantes.

REMARQUE 3.3. On retrouve par exemple le troisième point si $\mathcal{B}_i = \sigma(X_i)$ et $\mathcal{D}_i = \sigma(Y_i)$: le lemme nous dit que $(\sigma(\mathcal{B}_i, \mathcal{D}_i))_{i\in I}$ est une famille de tribus indépendantes. Et on a $\sigma(F(X_i, Y_i)) \subseteq \sigma((X_i, Y_i)) = \sigma(\mathcal{B}_i, \mathcal{D}_i)$. Ainsi les $(\sigma(F(X_i, Y_i)))_{i\in I}$ sont indépendantes, donc $(F(X_i, Y_i))_{i\in I}$ le sont aussi.

Pour démontrer le lemme de regroupement, on va utiliser les classes monotones :

LEMME 3.4. *Soit $(\mathcal{B}_i)_{i\in I}$ une famille de tribus de $\mathcal{A}$ puis pour $i \in I$, soit $\mathcal{C}_i \subset \mathcal{B}_i$ où $\mathcal{C}_i$ est une classe stable par intersection finie telle que $\sigma(\mathcal{C}_i) = \mathcal{B}_i$.*

Alors les $(\mathcal{B}_i)_{i\in I}$ sont indépendants si et seulement si :

$$\forall k \in \mathbb{N}^*, \forall i_1, i_2, \dots, i_k \in I \text{ tous distincts}, \forall C_1 \in \mathcal{C}_{i_1}, \dots, C_k \in \mathcal{C}_{i_k}, \mathbb{P}\Big(\bigcap_{j=1}^{k} C_j\Big) = \prod_{j=1}^{k} \mathbb{P}(C_j)$$

PREUVE Si les $(\mathcal{B}_i)_{i\in I}$ sont indépendants, la condition est vérifiée.

Réciproquement, il suffit de traiter le cas où $I = [\![1, n]\!]$ (la propriété est vraie si et seulement si elle est vraie pour toute sous-famille finie).
Si les $(\mathcal{A}_i)_{i\in I}$ sont des classes de $\mathcal{A}$, définissons la propriété :

$$\mathscr{P}_{\mathcal{A}_1,\dots,\mathcal{A}_n} : \quad \forall k \in [\![1, n]\!], \forall i_1 < \dots < i_k \in I, \forall A_1 \in \mathcal{A}_{i_1}, \dots, A_k \in \mathcal{A}_{i_k}, \mathbb{P}\Big(\bigcap_{j=1}^{k} A_j\Big) = \prod_{j=1}^{k} \mathbb{P}(A_j)$$

On a $\mathscr{P}_{\mathcal{C}_1,\dots,\mathcal{C}_n}$. On veut montrer $\mathscr{P}_{\mathcal{B}_1,\dots,\mathcal{B}_n}$.
Il suffit de prouver que si les $(\mathcal{A}_i)_{i\in I}$ sont stables par intersection finie, alors on a :

$$\mathscr{P}_{\mathcal{A}_1,\dots,\mathcal{A}_n} \Longrightarrow \mathscr{P}_{\sigma(\mathcal{A}_1),\dots,\mathcal{A}_n} \tag{3.1}$$

Et effet, dans ce cas, on aura :

$$\mathscr{P}_{\mathcal{C}_1,\dots,\mathcal{C}_n} \Longrightarrow \mathscr{P}_{\sigma(\mathcal{C}_1),\dots,\mathcal{C}_n} \Longrightarrow \mathscr{P}_{\sigma(\mathcal{C}_1),\sigma(\mathcal{C}_2),\dots,\mathcal{C}_n} \Longrightarrow \dots \Longrightarrow \mathscr{P}_{\sigma(\mathcal{C}_1),\dots,\sigma(\mathcal{C}_n)} = \mathscr{P}_{\mathcal{B}_1,\dots,\mathcal{B}_n}$$

Montrons (3.1). Supposons $\mathscr{P}_{\mathcal{A}_1,\dots,\mathcal{A}_n}$ et $\mathcal{A}_1$ stable par intersection finie. Posons

$$\mathcal{M} = \left\{ A \in \mathcal{A} \,|\, \forall k \geq 2, \forall i_2 < \dots < i_k \in [\![2, n]\!], \forall A_2 \in \mathcal{A}_{i_2}, \dots, A_n \in \mathcal{A}_{i_n}, \mathbb{P}\Big(A \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) = \mathbb{P}(A) \prod_{j=2}^{k} \mathbb{P}(A_j) \right\}$$

On a $\mathscr{P}_{\mathcal{M},\mathcal{A}_2,\dots,\mathcal{A}_n}$.
Il est clair que $\mathcal{A}_1 \subseteq \mathcal{M}$. Vérifions que $\mathcal{M}$ est une classe monotone :

- $\mathcal{M}$ est non vide (contient Ω),

– Si $A \subseteq B$ sont dans $\mathcal{M}$:

$$\begin{aligned}
\mathbb{P}\Big([B \setminus A] \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) &= \mathbb{P}\Big(\Big[B \bigcap [\bigcap_{j=2}^{k} A_j]\Big] \setminus \Big[A \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big]\Big) \\
&= \mathbb{P}\Big(B \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) - \mathbb{P}\Big(A \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) \quad \text{par inclusion} \\
&= \mathbb{P}(B)\mathbb{P}\Big(\bigcap_{j=2}^{k} A_j\Big) - \mathbb{P}(A)\mathbb{P}\Big(\bigcap_{j=2}^{k} A_j\Big) \qquad \text{puisque } A, B \in \mathcal{M} \\
&= \mathbb{P}(B \setminus A)\mathbb{P}\Big(\bigcap_{j=2}^{k} A_j\Big)
\end{aligned}$$

Donc $B \setminus A \in \mathcal{M}$.

– Si $(B_m)_{m\in\mathbb{N}}$ est une suite croissante dans $\mathcal{M}$:

$$\begin{aligned}
\mathbb{P}\Big(\Big[\bigcup_{m\in\mathbb{N}} \uparrow B_m\Big] \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) &= \mathbb{P}\Big(\bigcup_{m\in\mathbb{N}} \uparrow \Big[B_m \bigcap \bigcap_{j=2}^{k} A_j\Big]\Big) \\
&= \lim_{m\to+\infty} \uparrow \mathbb{P}\Big(B_m \bigcap \Big[\bigcap_{j=2}^{k} A_j\Big]\Big) \\
&= \lim_{m\to+\infty} \uparrow \mathbb{P}(B_m)\mathbb{P}\Big(\bigcap_{j=2}^{k} A_j\Big) \qquad \text{car } B_m \in \mathcal{M} \\
&= \mathbb{P}\Big(\bigcup_{m\in\mathbb{N}} \uparrow B_m\Big)\mathbb{P}\Big(\bigcap_{j=2}^{k} A_j\Big)
\end{aligned}$$

Donc $\bigcup_{m\in\mathbb{N}} \uparrow B_m \in \mathcal{M}$.

Donc $\mathcal{M}$ est bien une classe monotone. Et comme $\mathcal{A}_1$ est stable par intersection finie, le lemme 1.13 des classes monotones donne que $\sigma(\mathcal{A}_1) = \mathcal{M}(\mathcal{A}_1) \subseteq \mathcal{M}$, et donc on a $\mathscr{P}_{\sigma(\mathcal{A}_1),\dots,\mathcal{A}_n}$, ce qui conclut. □

REMARQUE 3.5. On remarque l'intérêt des classes monotones, qui sont faciles à vérifier mais très puissantes grâce au lemme !

PREUVE [LEMME DE REGROUPEMENT]

On veut l'indépendance des $(\hat{\mathcal{B}}_r)_{r\in R}$. Cherchons des classes qui les engendrent. Soit, pour $r \in R$ fixé :

$$\mathcal{C}_r = \left\{ \bigcap_{\ell=1}^{k} A_\ell \,|\, k \geq 1 \text{ et } \exists j_1, \dots, j_k \in J_r \text{ tous distincts} \,|\, A_1 \in \mathcal{B}_{j_1}, \dots, A_k \in \mathcal{B}_{j_k} \right\}$$

Il est immédiat que $\mathcal{C}_r$ est une classe stable par intersection finie.
De plus, on a, pour tout $j \in J_r$, $\mathcal{B}_j \subseteq \mathcal{C}_r$, donc $\mathcal{B}_j \subseteq \sigma(\mathcal{C}_r)$. Ainsi $\hat{\mathcal{B}}_r = \sigma(\{\mathcal{B}_j \,|\, j \in J_r\}) \subseteq \sigma(\mathcal{C}_r)$.
D'après le lemme précédent, il suffit donc de prouver que les $(\mathcal{C}_r)_{r\in R}$ vérifient la propriété d'indépendance.

Soit $p \in \mathbb{N}^*$. Pour $r \in [\![1,p]\!]$, on fixe $k_r \in \mathbb{N}^*$ et on pose $B_r = \bigcap_{i=1}^{k_r} A_i^r \in \mathcal{C}_r$ où pour $i \in [\![1,k_r]\!]$, $A_i^r \in \mathcal{B}_{j_i}$ où $j_i \in J_r$ avec les $(j_i)_{1\leq i\leq k_r}$ tous distincts. On a :

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{r=1}^{p} B_r\Big) &= \mathbb{P}\Big(\bigcap_{r=1}^{p}\bigcap_{i=1}^{k_r} A_i^r\Big) \\
&= \prod_{r=1}^{p}\prod_{i=1}^{k_r} \mathbb{P}(A_i^r) && \text{car les } (\mathcal{B}_i)_{i\in I} \text{ sont indépendants et on a une partition} \\
&= \prod_{r=1}^{p} \mathbb{P}\Big(\bigcap_{i=1}^{k_r} A_i^r\Big) && \text{car les } (\mathcal{B}_i)_{i\in J_r} \text{ sont indépendants} \\
&= \prod_{r=1}^{p} \mathbb{P}(B_r)
\end{aligned}$$

On a donc $\mathscr{P}_{\mathcal{C}_1,\dots,\mathcal{C}_n}$ ce qui implique $\mathscr{P}_{\hat{\mathcal{B}}_1,\dots,\hat{\mathcal{B}}_r}$. □

II. Indépendance de variables aléatoires

On rappelle la définition :

DÉFINITION 3.6. [INDÉPENDANCE DE VARIABLES ALÉATOIRES]
Soient $(X_i)_{i\in I}$ des variables aléatoires. On dit que les $(X_i)_{i\in I}$ sont *indépendantes* si les tribus $(\sigma(X_i))_{i\in I}$ le sont.

REMARQUE 3.7. Soient pour $i \in I$, $X_i : (\Omega,\mathcal{A}) \longrightarrow (E_i,\mathcal{E}_i)$ et $F_i : (E_i,\mathcal{E}_i) \longrightarrow (G_i,\mathcal{G}_i)$. Alors, d'après la partie précédente, si les $(X_i)_{i\in I}$ sont indépendantes, les $(F_i(X_i))_{i\in I}$ le sont aussi, puisque :

$$\sigma(F_i(X_i)) = (F_i\circ X_i)^{-1}(\mathcal{G}_i) = X_i^{-1}(F_i^{-1}(\mathcal{G}_i)) \subseteq X_i^{-1}(\mathcal{E}_i) = \sigma(X_i)$$

PROPOSITION 3.8. [CARACTÉRISATION DE L'INDÉPENDANCE]
Soient $n \in \mathbb{N}^*$ et pour $i \in [\![1,n]\!]$, $X_i : (\Omega,\mathcal{A}) \longrightarrow (E_i,\mathcal{E}_i)$ des variables aléatoires. On a l'équivalence entre les propositions :

(i) *les $(X_i)_{1\leq i\leq n}$ sont indépendantes,*

(ii) $$\forall (A_i)_{1\leq i\leq n} \in \textstyle\prod_{i=1}^n \mathcal{E}_i, \mathbb{P}(\forall i \in [\![1,n]\!], X_i \in A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$$

(iii) $$\mu_{(X_1,\dots,X_n)} = \mu_{X_1}\otimes\cdots\otimes\mu_{X_n}$$

Si de plus on a $(E_i,\mathcal{E}_i) = (\mathbb{R}^{d_i},\mathcal{B}(\mathbb{R}^{d_i}))$ pour tout $i \in [\![1,n]\!]$, alors on a aussi équivalence avec les propositions suivantes :

(iv) *pour toute famille de fonctions mesurables positives $(f_i : (E_i,\mathcal{E}_i) \longrightarrow (\mathbb{R},\mathcal{B}(\mathbb{R})))_{1\leq i\leq n}$, on a :*

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

(v) $$\forall \lambda_1 \in \mathbb{R}^{d_1},\dots,\lambda_n \in \mathbb{R}^{d_n}, \phi_{(X_1,\dots,X_n)}(\lambda_1,\dots,\lambda_n) = \textstyle\prod_{i=1}^n \phi_{X_i}(\lambda_i)$$

RAPPEL 3.9. Si μ est une mesure de probabilité sur $(E,\mathcal{E})$ et ν est une mesure de probabilité sur $(F,\mathcal{F})$, alors $\mu\otimes\nu$ est l'unique mesure de probabilité sur $(E\times F,\mathcal{E}\otimes\mathcal{F})$ telle que $\mu\otimes\nu(A\times B)=\mu(A)\nu(B)$ pour tout $A\in\mathcal{E}, B\in\mathcal{F}$. L'unicité provient du fait que les pavés forment une classe stable par intersection finie.

Enfin, soit pour $i\in[\![1,n]\!]$, μ_i une mesure de probabilités sur $(E_i,\mathcal{E}_i)$ et f_i mesurable positive. Alors on a :

$$\int f_1(x_1)\dots f_1(x_1)\,d(\mu_1\otimes\cdots\otimes\mu_n)(x_1,\dots,x_n)=\prod_{i=1}^n\int f_i(x_i)\,d\mu_i(x_i) \tag{3.2}$$

En effet, le résultat est vrai pour les fonctions indicatrices, puis par linéarité pour les fonctions étagées, et enfin par convergence monotone pour toutes les fonctions.

PREUVE

(i) $\Longrightarrow$ (ii) C'est trivial.

(ii) $\Longrightarrow$ (i) On peut enlever des indices en prenant $A_i=E_i$.

(ii) $\Longrightarrow$ (iii) On a, pour $(A_i)_{1\le i\le n}\in\prod_{i=1}^n\mathcal{E}_i$:

$$\mu_{(X_1,\dots,X_n)}\Big(\prod_{i=1}^n A_i\Big)=\mathbb{P}\Big((X_1,\dots,X_n)\in\prod_{i=1}^n A_i\Big)=\prod_{i=1}^n\mathbb{P}(X_i\in A_i)=\prod_{i=1}^n\mu_{X_i}(A_i)=\Big(\bigotimes_{i=1}^n\mu_{X_i}\Big)\Big(\prod_{i=1}^n A_i\Big)$$

Ainsi, les deux mesures coïncident sur les pavés, qui engendrent la tribu, donc sont égales.

(iii) $\Longrightarrow$ (iv) Si $\mu_{(X_1,\dots,X_n)}=\mu_{X_1}\otimes\cdots\otimes\mu_{X_n}$, alors :

$$\begin{aligned}
\mathbb{E}\Big[\prod_{i=1}^n f_i(X_i)\Big]&=\int f_1(x_1)\dots f_n(x_n)\,d\mu_{(X_1,\dots,X_n)}(x_1,\dots,x_n) && \text{par le lemme de tranfert}\\
&=\int f_1(x_1)\dots f_n(x_n)\,d(\mu_{X_1}\otimes\cdots\otimes\mu_{X_n})(x_1,\dots,x_n)\\
&=\prod_{i=1}^n\int f_i(x_i)\,d\mu_{X_i}(x_i) && \text{en utilisant (3.2)}\\
&=\prod_{i=1}^n\mathbb{E}[f_i(X_i)] && \text{par le lemme de tranfert}
\end{aligned}$$

(iv) $\Longrightarrow$ (ii) On applique en prenant pour $i\in[\![1,n]\!]$, $f_i=\mathbb{1}_{A_i}$.

(iii) $\Longleftrightarrow$ (v) Calculons :

$$\begin{aligned}
\phi_{(X_1,\dots,X_n)}(\lambda_1,\dots,\lambda_n)&=\int \mathrm{e}^{i\langle(\lambda_i)|(x_i)\rangle}\;d(\mu_{X_1}\otimes\cdots\otimes\mu_{X_n})(x_1,\dots,x_n)\\
&=\int \mathrm{e}^{i\langle(\lambda_1)|(x_1)\rangle}\dots\mathrm{e}^{i\langle(\lambda_n)|(x_n)\rangle}\;d\mu_{X_1}(x_1)\dots\,d\mu_{X_n}(x_n)\\
&=\prod_{i=1}^n\int \mathrm{e}^{i\langle(\lambda_i)|(x_i)\rangle}\;d\mu_{X_i}(x_i) && \text{en utilisant (3.2)}\\
&=\prod_{i=1}^n\phi_{X_i}(\lambda_i)
\end{aligned}$$

Donc (v) $\Longleftrightarrow\phi_{(X_1,\dots,X_n)}=\prod_{i=1}^n\phi_{X_i}\Longleftrightarrow\mu_{(X_1,\dots,X_n)}=\mu_{X_1}\otimes\cdots\otimes\mu_{X_n}\Longleftrightarrow$ (iii).

□

PROPOSITION 3.10. [CARACTÉRISATION DANS LE CAS À DENSITÉ]
Pour $i \in [\![1,n]\!]$, on considère une variable aléatoire $X_i : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^{d_i}, \mathcal{B}(\mathbb{R}^{d_i}))$.

(i) Si pour tout $i \in [\![1,n]\!]$, μ_{X_i} admet une densité ρ_i par rapport à la mesure de LEBESGUE ($\mu_{X_i}(x) = \rho_i(x)\,dx$) et si les $(X_i)_{1\le i\le n}$ sont indépendantes, alors $\mu_{(X_1,\dots,X_n)}$ est une loi à densité sur $\mathbb{R}^{\sum_{i=1}^n d_i}$ et on a :

$$\mu_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \prod_{i=1}^n \rho_i(x_i)\,dx_i$$

(ii) Si $\mu_{(X_1,\dots,X_n)} = \tilde\rho_1(x_1)\dots\tilde\rho_n(x_n)\,dx_1\dots\,dx_n$ avec $\tilde\rho_i : \mathbb{R}^{d_i} \longrightarrow \mathbb{R}$ mesurable, alors les $(X_i)_{1\le i\le n}$ sont indépendantes et ont des lois à densité vérifiant pour tout $i \in [\![1,n]\!]$:

$$\mu_{X_i}(x) = \rho_i(x)\,dx \quad \text{où} \quad \rho_i(x) = \frac{\tilde\rho_i(x)}{\int_{\mathbb{R}^{d_i}} \tilde\rho_i(x)\,dx}$$

PREUVE

– Vérifions l'égalité sur les pavés. On a, pour $(A_i)_{1\le i\le n} \in \prod_{i=1}^n \mathcal{B}(\mathbb{R}^{d_i})$:

$$\begin{aligned}\mu_{(X_1,\dots,X_n)}(A_1\times\dots\times A_n) = \mathbb{P}(\forall i\in[\![1,n]\!], X_i\in A_i) &= \prod_{i=1}^n \mu_{X_i}(A_i) \quad \text{par indépendance}\\ &= \prod_{i=1}^n \int_{A_i} d\mu_{X_i} = \prod_{i=1}^n \int_{A_i} \rho_i\,dx_i\\ &= \int_{A_1\times\dots\times A_n} \rho_1(x_1)\dots\rho_n(x_n)\,dx_1\dots\,dx_n \quad \text{par le théorème de FUBINI}\end{aligned}$$

Ainsi, les deux mesures coïncident sur les pavés, qui engendrent la tribu, donc sont égales.

– On cherche, à $i \in [\![1,n]\!]$ fixé, la loi de X_i. On a, pour $A_i \in \mathcal{B}(\mathbb{R}^{d_i})$:

$$\begin{aligned}\mathbb{P}(X_i\in A_i) &= \mathbb{P}((X_1,\dots,X_n)\in\mathbb{R}^{d_1}\times\dots\times\mathbb{R}^{d_{i-1}}\times A_i\times\mathbb{R}^{d_{i+1}}\times\dots\times\mathbb{R}^{d_n})\\ &= \int_{x_i\in A_i} \tilde\rho_1(x_1)\dots\tilde\rho_n(x_n)\,dx_1\dots\,dx_n\\ &= \int_{A_i}\tilde\rho_i(x_i)\,dx_i \prod_{j\neq i}\int_{\mathbb{R}^{d_j}}\tilde\rho_j(x_j)\,dx_j \qquad \text{par le th. de FUBINI}\end{aligned}$$

Et $\tilde\rho_1(x_1)\dots\tilde\rho_n(x_n)\,dx_1\dots\,dx_n$ est une probabilité donc si on intègre sur $\mathbb{R}^{\sum_{i=1}^n d_i}$, on obtient 1.
En particulier, $\prod_{j\neq i}\int_{\mathbb{R}^{d_j}}\tilde\rho_j(x_j)\,dx_j = \left(\int_{\mathbb{R}^{d_i}}\tilde\rho_i(x)\,dx\right)^{-1}$, d'où :

$$\mathbb{P}(X_i\in A_i) = \int_{A_i}\frac{\tilde\rho_i(x)}{\int_{\mathbb{R}^{d_i}}\tilde\rho_i(t)\,dt}\,dx$$

Ou encore $\mu_{X_i}(x) = \dfrac{\tilde\rho_i(x)}{\int_{\mathbb{R}^{d_i}}\tilde\rho_i(t)\,dt}\,dx$. Puis :

$$\mu_{(X_1,\dots,X_n)}(x) = \prod_{i=1}^n \frac{\tilde\rho_i(x_i)}{\int_{\mathbb{R}^{d_i}}\tilde\rho_i(t)\,dt}\,dx_i = \prod_{i=1}^n \tilde\rho_i(x_i)\,dx_i = \bigotimes_{i=1}^n \mu_{X_i}$$

Donc les $(X_i)_{1\leq i\leq n}$ sont indépendantes.

□

III. Convolution

DÉFINITION 3.11. [CONVOLUTION DE MESURES DE PROBABILITÉS]
Si μ et ν sont des probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (cela se généralise sur $\mathbb{R}^d$), on définit $\mu * \nu$ comme étant la loi de $X + Y$ où $X \sim \mu$, $Y \sim \nu$, et X et Y sont indépendantes.

REMARQUE 3.12. C'est une mesure de probabilité sur $\mathbb{R}$ bien définie.
En général la loi de $f(Z)$ ne dépend que de f et de la loi de Z :

$$\mu_{f(Z)}(A) = \mathbb{P}(f(Z) \in A) = \mathbb{P}(Z \in f^{-1}(A)) = \mu_Z(f^{-1}(A))$$

Ici, $X+Y = \Delta((X,Y))$ où Δ est la fonction somme, et (X,Y) a pour loi $\mu\otimes\nu$ par indépendance :

$$(X,Y) \sim \mu \otimes \nu \Longrightarrow X + Y \sim \mu * \nu$$

PROPOSITION 3.13. *On a les propriétés suivantes :*

(i) *la convolution est commutative,*

(ii) *la convolution est associative,*

(iii) δ_0 *est un élément neutre,*

(iv) $$\phi_{\mu*\nu} = \phi_\mu \phi_\nu$$

(v) *Si* $\mu(x) = f(x)\,dx$ *et* $\nu(x) = g(x)\,dx$, *alors*

$$\mu * \nu(x) = (f * g)(x)\,dx = f * g(x) = \left(\int f(y)g(x-y)\,dy\right)\,dx$$

PREUVE

(i) C'est trivial puisque $X + Y = Y + X$,

(ii) Soient X, Y, Z indépendantes telles que $X \sim \mu$, $Y \sim \nu$ et $Z \sim \lambda$.
Alors $X + Y \sim \mu * \nu$ donc $(X + Y) + Z = X + Y + Z \sim (\mu * \nu) * \lambda$ puisque $(X + Y)$ et Z sont indépendantes par regroupement.
De même, $X + (Y + Z) = X + Y + Z \sim \mu * (\nu * \lambda)$.

(iii) C'est trivial puisque $0 \sim \delta_0$.

(iv) On a :

$$\phi_{\mu*\nu}(\lambda) = \phi_{X+Y}(\lambda) = \mathbb{E}[\mathrm{e}^{i\lambda(X+Y)}] = \mathbb{E}[\mathrm{e}^{i\lambda X}\,\mathrm{e}^{i\lambda Y}] \underset{\text{indép}}{=} \mathbb{E}[\mathrm{e}^{i\lambda X}]\mathbb{E}[\mathrm{e}^{i\lambda Y}] = \phi_\mu(\lambda)\phi_\nu(\lambda)$$

NOTE 3.14. Attention à ne pas confondre $\phi_{\mu*\nu}(\lambda) = \phi_\mu(\lambda)\phi_\nu(\lambda)$ et $\phi_{\mu\otimes\nu}(\lambda) = \phi_\mu(\lambda_1)\phi_\nu(\lambda_2)$!

(v) Calculons $\mu_{X+Y} = \mu * \nu$. Soient $\Phi : \mathbb{R} \longrightarrow \mathbb{R}$ mesurable positive, $X \sim \mu$ et $Y \sim \nu$ indépen-

dantes. On a :

$$\begin{aligned}
\mathbb{E}[\Phi(X+Y)] &= \int \Phi(x+y)\, d\mu_{(X,Y)}(x,y) && \text{par le lemme de transfert} \\
&= \int \Phi(x+y) f(x) g(y)\, dx\, dy \\
&= \int \Phi(z) f(x) g(z-x)\, dx\, dz \\
&= \int \Phi(z) \left(\int f(x) g(z-x)\, dx \right) dz && \text{par le théorème de FUBINI}
\end{aligned}$$

Donc $\mu_{X+Y}(z) = \mu * \nu(z) = (f * g)(z)\, dz$.

□

EXEMPLE 3.15.

- On a $\mathcal{B}(n,p) = \mathcal{B}(p) * \cdots * \mathcal{B}(p)$, puis $\boxed{\mathcal{B}(n+m,p) = \mathcal{B}(n,p) * \mathcal{B}(m,p)}$.
- On a vu que $\phi_{\mathcal{N}(m,\sigma^2)}(\lambda) = \mathrm{e}^{i\lambda m}\, \mathrm{e}^{-\frac{\lambda^2\sigma^2}{2}}$. Ainsi :

$$\phi_{\mathcal{N}(m_1,\sigma_1^2)*\mathcal{N}(m_2,\sigma_2^2)} = \phi_{\mathcal{N}(m_1,\sigma_1^2)} \phi_{\mathcal{N}(m_2,\sigma_2^2)}$$

  Donc $\boxed{\mathcal{N}(m_1,\sigma_1^2) * \mathcal{N}(m_2,\sigma_2^2) = \mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)}$.
- Si $X \sim \mathcal{P}(\lambda)$:

$$\phi_X(t) = \mathbb{E}[\mathrm{e}^{itX}] = \sum_{k\in\mathbb{N}^*} \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda}\, \mathrm{e}^{itk} = \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda\, \mathrm{e}^{it}} = \mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}$$

  Et donc si $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$ et X et Y sont indépendantes, alors $\boxed{X+Y \sim \mathcal{P}(\lambda+\mu)}$.

CHAPITRE 4

Vecteurs gaussiens

I. Lois gaussiennes réelles ($d = 1$)

On considère la loi $\mathcal{N}(0,1)$ de densité $x \longmapsto \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$, et plus généralement on définit les lois $\mathcal{N}(m,\sigma^2)$ comme étant les lois de $m+\sigma Z$ où $Z \sim \mathcal{N}(0,1)$. Par convention on pose $\mathcal{N}(m,0) = \delta_m$.

PROPOSITION 4.1. *On a l'équivalence :*

$$\begin{aligned} X \sim \mathcal{N}(m,\sigma^2) &\iff X \overset{\mathcal{L}}{=} m + \sigma Z \quad \text{où } Z \sim \mathcal{N}(0,1) \\ &\iff \forall \lambda \in \mathbb{R}, \phi_X(\lambda) = \mathrm{e}^{i\lambda m}\, \mathrm{e}^{-\frac{\lambda^2\sigma^2}{2}} \\ &\underset{\sigma\neq 0}{\iff} X \sim \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2}}\, dx \end{aligned}$$

A la loi $\mathcal{N}(m,\sigma^2)$ correspond une courbe centrée en m et d'étalement σ. Si $X \sim \mathcal{N}(m,\sigma^2)$, on a :

$$\boxed{\mathbb{E}[X] = m \quad \text{et} \quad \mathrm{Var}(X) = \sigma^2}$$

II. Gaussiennes en dimension $d \geq 2$

DÉFINITION 4.2. [VECTEUR GAUSSIEN]
Soit $X = (X_1, \ldots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$. On dit que c'est un *vecteur gaussien* si toute combinaison linéaire (ou affine) des $(X_i)_{1\leq i\leq d}$ a une loi gaussienne sur $\mathbb{R}$, c'est-à-dire si :

$$\forall \lambda \in \mathbb{R}^d, \exists (m_\lambda, \sigma_\lambda) \in \mathbb{R}\times\mathbb{R}^+ \mid \langle \lambda \mid X\rangle = \sum_{i=1}^{d} \lambda_i X_i \sim \mathcal{N}(m_\lambda, \sigma_\lambda^2)$$

REMARQUE 4.3. Il est équivalent de demander linéaire ou affine.

DÉFINITION 4.4. [VECTEUR GAUSSIEN STANDARD]
$\mathcal{N}_{\mathbb{R}^d}(0_{\mathbb{R}^d}, \mathrm{Id})$ est la loi de $Z = (Z_1, \dots, Z_d)$ où les $(Z_i)_{1\le i\le d}$ sont indépendantes et identiquement distribuées, de loi $\mathcal{N}(0,1)$.
On a $Z \sim \frac{\mathrm{e}^{-\frac{x_1^2}{2}}}{\sqrt{2\pi}} \dots \frac{\mathrm{e}^{-\frac{x_n^2}{2}}}{\sqrt{2\pi}}\, dx_1 \dots\, dx_n = \frac{\mathrm{e}^{-\frac{\|x\|^2}{2}}}{(2\pi)^{\frac{d}{2}}}\, dx$ où dx est la mesure de LEBESGUE sur $\mathbb{R}^d$.

PROPOSITION 4.5. *Soient $X = (X_1, \dots, X_d)$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $m \in \mathbb{R}^d$ et $A \in \mathcal{M}_d(\mathbb{R})$.*

On pose $\Sigma = A\,{}^tA$. Alors on a les équivalences :

(i) $X = m + AZ$ *avec* $Z \sim \mathcal{N}_{\mathbb{R}^d}(0, \mathrm{Id})$,

(ii) $\forall \lambda \in \mathbb{R}^d, \phi_X(\lambda) = \mathrm{e}^{i\langle\lambda|m\rangle}\, \mathrm{e}^{-\frac{{}^t\lambda\Sigma\lambda}{2}}$,

(iii) X *est un vecteur gaussien et on a :*

$$\mathbb{E}[X] = m \quad \textit{et} \quad \Gamma_X := (\mathrm{Cov}(X_i, X_j))_{1\le i,j\le d} = \Sigma$$

Si de plus Σ est inversible, on a aussi équivalence avec le point :

(iv) $X \sim \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\, {}^t(x-m)\Sigma^{-1}(x-m)\right) dx.$

DÉFINITION 4.6. [VECTEUR GAUSSIEN]
Si X vérifie l'une de ces propriétés, on dira que X est un *vecteur gaussien* de loi $\mathcal{N}(m, \Sigma)$, m étant son espérance et Σ sa matrice de covariance.

NOTE 4.7. Σ joue le rôle de σ^2.

PREUVE

(i) $\iff$ (ii) En général, on a $(\star)$ $\quad \forall \lambda \in \mathbb{R}^d, \phi_{m+AY}(\lambda) = \mathrm{e}^{i\langle\lambda|m\rangle}\, \phi_Y({}^tA\lambda)$.
En effet :

$$\phi_{m+AY}(\lambda) = \mathbb{E}[\mathrm{e}^{i\langle\lambda|m+AY\rangle}] = \mathrm{e}^{i\langle\lambda|m\rangle}\, \mathbb{E}[\mathrm{e}^{i\,{}^t\lambda AY}] = \mathrm{e}^{i\langle\lambda|m\rangle}\, \mathbb{E}[\mathrm{e}^{i\,{}^t({}^tA\lambda)Y}] = \mathrm{e}^{i\langle\lambda|m\rangle}\, \phi_Y({}^tA\lambda)$$

Si $Z = (Z_1, \dots, Z_d) \sim \mathcal{N}_{\mathbb{R}^d}(0, \mathrm{Id})$:

$$\phi_Z(\lambda) = \prod_{i=1}^d \phi_{Z_i}(\lambda_i) = \prod_{i=1}^d \mathrm{e}^{-\frac{\lambda_i^2}{2}} = \mathrm{e}^{-\frac{{}^t\lambda\lambda}{2}}$$

Donc, en utilisant $(\star)$:

$$\phi_{m+AZ}(\lambda) = \mathrm{e}^{i\langle\lambda|m\rangle}\, \mathrm{e}^{-\frac{1}{2}\,{}^t({}^tA\lambda)\,{}^tA\lambda} = \mathrm{e}^{i\langle\lambda|m\rangle}\, \mathrm{e}^{-\frac{1}{2}\,{}^t\lambda\Sigma\lambda}$$

L'équivalence (i) $\iff$ (ii) découle alors de $\phi_X = \phi_{m+AZ} \iff X \overset{\mathcal{L}}{=} m + AZ$.

(i) $\iff$ (iv) Il suffit de vérifier que $m + AZ$ a une loi de densité qui suit cette formule si A est inversible.
Soit f mesurable positive. On a :

$$\mathbb{E}[f(m+AZ)] \underset{\text{transfert}}{=} \int_{\mathbb{R}^d} f\left(m + A\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}\right) \underbrace{d\mu_Z(x_1, \dots, x_d)}_{=\frac{1}{(2\pi)^{\frac{d}{2}}}\mathrm{e}^{-\frac{\|x\|^2}{2}}\, dx}$$

On effectue alors le CDV $\begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = m + A \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$ (affine donc C^1, inversible puisque A l'est), et on obtient :

$$\mathbb{E}[f(m+AZ)] = \underbrace{\det(A)^{-1}}_{\substack{=\det(A)^{-1} \\ =\det(\Sigma)^{-1/2}}} \int_{\mathbb{R}^d} f(y) \frac{1}{(2\pi^{\frac{d}{2}})} \exp\Big(-\frac{1}{2}\, {}^t(y-m)\Sigma^{-1}(y-m) \Big)\, dx$$

en remarquant que $x = A^{-1}(y-m)$ implique $\|x\|_2^2 = {}^t xx = {}^t(y-m)\Sigma^{-1}(y-m)$.

(i) $\Longrightarrow$ (iii) On utilise deux résultats :

$$\mathbb{E}[m+AY] = m + A\mathbb{E}[Y] \quad \text{et} \quad \Gamma_{m+AY} = A\Gamma_Y({}^tA)$$

En effet, on a :

$$(\mathbb{E}[m+AY])_i = \mathbb{E}\Big[m_i + \sum_{j=1}^d A_{ij}Y_j\Big] = m_i + \sum_{j=1}^d A_{ij}\mathbb{E}[Y_j] = (m + A\,\mathbb{E}[Y])_i$$

Puis :

$$\begin{aligned}
(\Gamma_{m+AY})_{ij} &= \mathrm{Cov}((m+AY)_i, (m+AY)_j) \\
&= \mathrm{Cov}\Big(\sum_{k=1}^d A_{ik}Y_k, \sum_{\ell=1}^d A_{j\ell}Y_\ell \Big) \\
&= \sum_{k=1}^d \sum_{\ell=1}^d A_{ik}A_{j\ell}\,\mathrm{Cov}(Y_k, Y_\ell) \qquad \text{par bilinéarité} \\
&= \sum_{k=1}^d \sum_{\ell=1}^d A_{ik}(\Gamma_Y)_{k\ell}({}^tA)_{\ell j} \\
&= (A\Gamma_Y\,{}^tA)_{ij}
\end{aligned}$$

Comme $Z \sim \mathcal{N}_{\mathbb{R}^d}(0, \mathrm{Id})$, on a $\mathbb{E}[Z] = 0$ et $\Gamma_Z = \mathrm{Id}$.
Or, puisque $X \overset{\mathcal{L}}{=} m + AZ$:

$$\forall \lambda \in \mathbb{R}^d, \langle \lambda \mid X \rangle \overset{\mathcal{L}}{=} \langle \lambda \mid m + AZ \rangle$$

Ce dernier produit scalaire étant une combinaison affine des $(Z_i)_{1\le i\le d}$, c'est une gaussienne puisque c'est une famille stable par application affine et convolution.
Donc X est un vecteur gaussien et on a $\mathbb{E}[X] = m + A\mathbb{E}[Z] = m$, puis $\Gamma_X = A\Gamma_Z\,{}^tA = \Sigma$.

(iii) $\Longrightarrow$ (ii) On cherche la fonction caractéristique de X. Soit $\lambda \in \mathbb{R}^d$.
On sait que $\langle \lambda \mid X \rangle \sim \mathcal{N}(m_\lambda, \sigma_\lambda^2)$, avec

$$\begin{cases}
m_\lambda = \mathbb{E}[\langle \lambda \mid X \rangle] = \mathbb{E}\Big[\sum_{i=1}^d \lambda_i X_i \Big] = \sum_{i=1}^d \lambda_i m_i = \langle \lambda \mid m \rangle \\
\sigma_\lambda^2 = \mathrm{Var}\Big(\sum_{i=1}^d \lambda_i X_i \Big) = \sum_{i=1}^d \sum_{j=1}^d \lambda_i\lambda_j\,\mathrm{Cov}(X_i, X_j) = {}^t\lambda\Sigma\lambda
\end{cases}$$

Donc on a $\langle \lambda \mid X \rangle \sim \mathcal{N}(\langle \lambda \mid m \rangle, {}^t\lambda\Sigma\lambda)$. Ainsi :

$$\phi_X(\lambda) = \mathbb{E}[\mathrm{e}^{i\langle \lambda | X \rangle}] = \phi_{\langle \lambda | X \rangle}(1) = \mathrm{e}^{im_\lambda}\,\mathrm{e}^{-\frac{\sigma_\lambda^2}{2}} = \mathrm{e}^{i\langle \lambda | m \rangle}\,\mathrm{e}^{-\frac{1}{2}\,{}^t\lambda\Sigma\lambda}$$

□

On en déduit le résultat :

COROLLAIRE 4.8.

Si $X = (X_1, X_2) \sim \mathcal{N}\left(m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$ *avec* $X_1 \in \mathbb{R}^p$ *et* $X_2 \in \mathbb{R}^q$ *puis* $m_1 \in \mathbb{R}^p$, $m_2 \in \mathbb{R}^q$, $\Sigma_{11} \in \mathcal{M}_p(\mathbb{R})$ *et* $\Sigma_{22} \in \mathcal{M}_q(\mathbb{R})$, *alors on a :*

- $X_1 \sim \mathcal{N}(m_1, \Sigma_{11})$,
- X_1 *et* X_2 *sont indépendants si et seulement si* $\Sigma_{12} = 0$.

PREUVE

- Le premier point est trivial,
- Pour le second point, montrons les deux implications :
  - $\Longrightarrow$ si on a indépendance alors les covariances entre coordonnées de X_1 et de X_2 sont nulles.
  - $\Longleftarrow$ on calcule la fonction caractéristique et on obtient en utilisant la linéarité du produit scalaire :

$$\phi_{(X_1,X_2)}(\lambda_1, \lambda_2) = \phi_{X_1}(\lambda_1)\phi_{X_2}(\lambda_2)$$

□

EXEMPLE 4.9. Soit $X \sim \mathcal{N}(0,1)$ indépendant de $Z \sim \mathcal{U}(\{1,-1\})$.

Posons $Y = ZX$ et considèrons le vecteur $(X,Y)^\perp$.
Alors X et Y ne sont pas indépendantes, puisque $|X| = |Y|$ p.s. mais $\mathbb{P}(|X| < 1 \cap |Y| > 1) = 0$ et $\mathbb{P}(|X| < 1)\mathbb{P}(|Y| > 1) > 0$.
De plus :

$$\begin{aligned}
\mathbb{P}(Y \in A) &= \mathbb{P}(Y \in A \cap Z = 1) + \mathbb{P}(Y \in A \cap Z = -1) \\
&= \mathbb{P}(Y \in A \,|\, Z = 1)\mathbb{P}(Z = 1) + \mathbb{P}(Y \in A \,|\, Z = -1)\mathbb{P}(Z = -1) \\
&= \frac{1}{2}\Big(\mathbb{P}(X \in A) + \mathbb{P}(X \in -A)\Big) \\
&= \mathbb{P}(X \in A) \qquad \text{par parité de } \mathcal{N}(0,1)
\end{aligned}$$

Donc $Y \sim \mathcal{N}(0,1)$ et on a :

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[ZX^2] = \mathbb{E}[Z]\mathbb{E}[X^2] = 0$$

On en conclut :

- que $\mathrm{Cov}(X,Y) = 0$ n'implique pas que X et Y indépendants,
- que les coordonnées sont gaussiennes n'implique pas que c'est un vecteur gaussien.

CHAPITRE 5

Construction de suites de variables aléatoires

On va s'intéresser dans ce chapitre à la construction de variables aléatoires : si on se donne $(\mu_i)_{i\in\mathbb{N}}$ une suite de mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, peut-on trouver sur $(\Omega, \mathcal{A}, \mathbb{P})$ une suite de variables aléatoires $(X_i)_{i\in\mathbb{N}}$ indépendantes telles que pour tout $i \in \mathbb{N}$, $X_i \sim \mu_i$? Moralement on aura alors $X = (X_i)_{i\in\mathbb{N}} \sim \otimes_{i\in\mathbb{N}}\mu_i$.

I. Premier cas : construction de X_1 de loi μ_1

On peut considérer la réalisation canonique :

$$\begin{array}{rccc} X_1 : & (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_1) & \longrightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ & x & \longmapsto & x \end{array}$$

Mais on veut quelque chose de plus *constructif* avec comme espace de départ $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$

DÉFINITION 5.1. [FONCTION DE RÉPARTITION]
Si X est une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on définit sa *fonction de répartition* par :

$$\begin{array}{rccc} F_X : & \mathbb{R} & \longrightarrow & [0,1] \\ & t & \longmapsto & \mathbb{P}(X \leq t) \end{array}$$

REMARQUE 5.2. F_X ne dépend que de la loi de X, on a $\forall t \in \mathbb{R}, F_X(t) = \mu_X(]-\infty, t])$.

PROPOSITION 5.3. *On a :* $\boxed{F_\mu = F_\nu \Longrightarrow \mu = \nu}$

PREUVE F_μ caractérise μ sur $\mathcal{C} = \{]-\infty, t] \mid t \in \mathbb{R}\}$, qui est une classe stable par intersection finie et qui engendre la tribu. □

PROPOSITION 5.4. *F_X est croissante, continue à droite et vérifie*

$$\lim_{t\to-\infty} F_X(t) = 0 \quad \text{et} \quad \lim_{t\to+\infty} F_X(t) = 1$$

PREUVE La preuve est laissée en exercice. □

On a une "réciproque" :

PROPOSITION 5.5. *Soit $F : \mathbb{R} \longrightarrow [0,1]$ croissante, continue à droite, vérifiant :*

$$\lim_{t\to-\infty} F(t) = 0 \quad \text{et} \quad \lim_{t\to+\infty} F(t) = 1$$

et considérons, pour $u \in [0,1]$, $F^{<-1>}(u) := \inf\{t \in \mathbb{R} \mid F(t) \geq u\}$.

Alors la variable aléatoire $F^{<-1>}(U)$ où $U \sim \mathcal{U}([0,1])$ a pour fonction de répartition F.

REMARQUE 5.6. On a $F_\mu^{<-1>} \sim \mu$, donc pour toute loi, $F_\mu^{<-1>} : ([0,1], \mathcal{B}([0,1]), \mathrm{Leb}) \longrightarrow \mathbb{R}$ est une variable aléatoire de loi μ. **Cela résout notre problème.**

EXEMPLE 5.7. Si $\mu = \mathcal{E}(\lambda)$, on vérifie que $\forall t, F_\mu(t) = (1 - \mathrm{e}^{-\lambda t})\mathbb{1}_{t\geq 0}$. On en déduit :

$$\forall 0 \leq u \leq 1, F_\mu^{<-1>}(u) = -\frac{\ln(1-u)}{\lambda}$$

Ainsi, si $U \sim \mathcal{U}([0,1])$, on a $-\frac{\ln(1-U)}{\lambda} \sim \mathcal{E}(\lambda)$.

PREUVE Pour $0 < u < 1$, on pose $A_u = \{t \in \mathbb{R} \mid F(t) \geq u\}$. On a

- $\lim_{t\to+\infty} F(t) = 1 > u$ donc $A_u \neq \varnothing$ et $\lim_{t\to-\infty} F(t) = 0$ donc $A_u \neq \mathbb{R}$,
- F est croissante donc si $t \in A_u$, on a $[t, +\infty[\subset A_u$. Ainsi A_u est de la forme $A_u = [a, +\infty[$ ou $A_u =]a, +\infty[$ pour un $a \in \mathbb{R}$ fixé.
  Par continuité à droite, et puisque $\forall n \in \mathbb{N}^*, a + \frac{1}{n} \in A_u$ (c'est-à-dire $F(a + \frac{1}{n}) \geq u$), on a que $F(a) \geq u$, et donc$a \in A_u$.

Donc finalement : $A_u = [a, +\infty[= [F^{<-1>}(u), +\infty[$. On en déduit :

$$F_{F^{<-1>}(U)}(t) = \mathbb{P}\Big(F^{<-1>}(U) \leq t \bigcap 0 < U < 1\Big) = \mathbb{P}(t \in A_U) = \mathbb{P}(F(t) \geq U) = \int_0^1 \mathbb{1}_{u\leq F(t)}\, du = F(t)$$

□

REMARQUE 5.8. Ceci caractérise les fonctions de répartitions.

II. Deuxième cas : construction de $(X_i)_{i\in\mathbb{N}^*}$ variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(1/2)$

PROPOSITION 5.9. *Soient Z et $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires telles que $Z \in [0,1]$ p.s., $X_i \in \{0,1\}$ p.s. pour tout $i \in \mathbb{N}$ et $Z = \sum_{i=1}^{+\infty} \frac{X_i}{2^i}$ p.s..*

Alors $Z \sim \mathcal{U}([0,1])$ si et seulement si $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(1/2)$.

REMARQUE 5.10. **Cela résout le problème** puisque si on se donne $Z \sim \mathcal{U}([0,1])$ et on pose, pour
$i \in \mathbb{N}^*$, $X_i = \lfloor 2^i Z \rfloor - \lfloor 2^{i-1} Z \rfloor$ le i-ème indice du développement dyadique de Z, alors $Z = \sum_{i=1}^{+\infty} \frac{X^i}{2^i}$ p.s., et donc les $(X_i)_{i\in\mathbb{N}^*}$ ont la loi souhaitée.

PREUVE Soit $A = \{\forall n \in \mathbb{N}^*, \exists i \geq n \mid X_i = 0\}$. Supposons A p.s..
Les $(X_i)_{i\in\mathbb{N}^*}$ sont des variables aléatoires i.i.d.de loi $\mathcal{B}(1/2)$ équivaut à :

$$\begin{aligned}
&\forall n \in \mathbb{N}^*, \forall x_1, \dots, x_n \in \{0,1\}, \mathbb{P}\Big((X_1, \dots, X_n) = (x_1, \dots, x_n)\Big) = \frac{1}{2^n} \\
\iff &\forall n \in \mathbb{N}^*, \forall x_1, \dots, x_n \in \{0,1\}, \mathbb{P}\Big((X_1, \dots, X_n) = (x_1, \dots, x_n), Z = \sum_{i=1}^{+\infty} \frac{X_i}{2^i}, A\Big) = \frac{1}{2^n} \\
\iff &\forall n \in \mathbb{N}^*, \forall x_1, \dots, x_n \in \{0,1\}, \mathbb{P}\Big(Z = \sum_{i=1}^{n} \frac{x_i}{2^i} + \sum_{i=n+1}^{+\infty} \frac{X_i}{2^i}, Z = \sum_{i=1}^{+\infty} \frac{X_i}{2^i}, A\Big) = \frac{1}{2^n} \\
\iff &\forall n \in \mathbb{N}^*, \forall x_1, \dots, x_n \in \{0,1\}, \mathbb{P}\Big(Z \in \Big[\sum_{i=1}^{n} \frac{x_i}{2^i}, \sum_{i=1}^{n} \frac{x_i}{2^i} + \frac{1}{2^n}\Big[\Big) = \frac{1}{2^n}
\end{aligned}$$

Donc, par un argument de classes monotones, cela équivaut au fait que $Z \sim \mathcal{U}([0,1])$.

Vérifions que l'on a bien A p.s. quelque soit le côté de l'équivalence considéré :

- si les $(X_i)_{i\in\mathbb{N}^*}$ sont des variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(1/2)$, on a :

$$\mathbb{P}(A^c) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}^*} \bigcap_{i\geq n} X_i = 1\Big) \leq \sum_{n\in\mathbb{N}^*} \mathbb{P}\Big(\bigcap_{i\geq n} X_i = 1\Big) = 0$$

  puisque $\mathbb{P}\Big(\bigcap_{i\geq n} X_i = 1\Big) \leq \mathbb{P}\Big(\bigcap_{i=n}^{n+k-1} X_i = 1\Big) = \frac{1}{2^k} \underset{k\to+\infty}{\longrightarrow} 0.$

- si $Z \sim \mathcal{U}$:

$$\mathbb{P}(A^c) = \mathbb{P}\Big(A^c \bigcap \sum_{i=1}^{+\infty} \frac{X_i}{2^i}\Big) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}^*} Z = \sum_{i=1}^{n} \frac{X_i}{2^i} + \underbrace{\sum_{i=n+1}^{+\infty} \frac{1}{2^i}}_{\frac{1}{2^n}}\Big) \leq P(Z \in \mathbb{Q}) = 0$$

□

On peut généraliser la proposition avec ce résultat :

THÉORÈME 5.11. *Soient $b \geq 2$ un entier, $Z \in [0,1]$ p.s. et $(X_i)_{i\in\mathbb{N}^*}$ telles que $X_i \in [\![0, b-1]\!]$ p.s. et $Z = \sum_{i=1}^{+\infty} \frac{X_i}{b^i}$ p.s..*

Alors $Z \sim \mathcal{U}([0,1])$ si et seulement si $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{U}([\![0, b-1]\!])$.

III. Construction de $(U_i)_{i\in\mathbb{N}^*}$ variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{U}([0,1])$

Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(1/2)$. On pose, pour $i, j \in \mathbb{N}^*$, $Y_{i,j} = X_{2^{i-1}(2j-1)}$ (en remarquant que $(i,j) \longmapsto 2^{i-1}(2j-1)$ est une bijection de $(\mathbb{N}^*)^2$ sur $\mathbb{N}^*$).

Alors $(Y_{i,j})_{i,j\in\mathbb{N}^*}$ est une famille de variables aléatoires indépendantes et par le lemme de regroupement, $(\mathcal{B}_j = \sigma(\{Y_{i,j} \mid i \in \mathbb{N}^*\}))_{j\in\mathbb{N}^*}$ forme une famille de tribus indépendantes.
On peut donc poser, pour $j \in \mathbb{N}^*$, $U_j = \sum_{i=1}^{+\infty} \frac{Y_{i,j}}{2^i}$, variable aléatoire qui suit $\mathcal{B}(1/2)$ et qui est $\mathcal{B}_j$-mesurable.
Les $(U_j)_{j\in\mathbb{N}^*}$ sont donc indépendantes donc d'après la proposition précédente : $\forall j \in \mathbb{N}^*, U_j \sim \mathcal{U}([0,1])$.

IV. Construction de $(X_i)_{i\in\mathbb{N}^*}$ variables aléatoires indépendantes de lois $(\mu_i)_{i\in\mathbb{N}^*}$

Soit $(\mu_i)_{i\in\mathbb{N}^*}$ une suite de mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose construit les $(U_i)_{i\in\mathbb{N}^*}$ comme dans la partie précédente. On pose alors pour tout $i \in \mathbb{N}^*$, $X_i = F_{\mu_i}^{<-1>}(U_i)$. On sait d'après la première partie que $X_i \sim \mu_i$. Et comme $\sigma(X_i) \subseteq \sigma(U_i)$, on a que les $(X_i)_{i\in\mathbb{N}^*}$ sont indépendants puisque les $(U_i)_{i\in\mathbb{N}^*}$ le sont.

REMARQUE 5.12.

- Ici on peut tout construire sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$. Essentiellement, on peut faire toutes les probabilités classiques sur cet espace mais on continuera de préférer de ne pas expliciter $(\Omega, \mathcal{A}, \mathbb{P})$.
- Que dire de $X = (X_i)_{i\in\mathbb{N}^*} : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^{\mathbb{N}^*}$? Est-ce une variable aléatoire ? Il faudrait pour cela définir une tribu sur $\mathbb{R}^{\mathbb{N}^*}$!

DÉFINITION 5.13. [TRIBU CYLINDRIQUE]
Soit $(E, \mathcal{E})$ un espace mesurable et $\mathcal{T}$ un ensemble ($\mathcal{T}$ modélise le temps, donc est à valeurs dans $\mathbb{N}, \mathbb{N}^*, \mathbb{R}, \dots$).
On définit sur $E^{\mathcal{T}}$ les événements, pour $n \in \mathbb{N}^*$, $(t_i)_{1\le i\le n} \in \mathcal{T}^n$ et $(A_i)_{1\le i\le n} \in \mathcal{E}^n$:

$$A_{(t_1,\dots,t_n),(A_1,\dots,A_n)} = \left\{ f \in E^{\mathcal{T}} \mid \forall 1 \le i \le n, f_{t_i} \in A_i \right\}$$

puis on définit la *tribu cylindrique* :

$$\mathcal{C} = \sigma\Big(\left\{ A_{(t_1,\dots,t_n),(A_1,\dots,A_n)} \mid n \in \mathbb{N}^*, (t_i)_{1\le i\le n} \in \mathcal{T}^n, (A_i)_{1\le i\le n} \in \mathcal{E}^n \right\} \Big)$$

REMARQUE 5.14.

- La tribu cylindrique est donc par définition la plus petite tribu qui permet de parler d'un nombre fini de temps.
- Si $\mathcal{T} = \mathbb{N}^*$, on a $\mathcal{C} = \sigma(\{[\forall 1 \le i \le n, X_i \in A_i] \mid n \in \mathbb{N}^*, (A_i)_{1\le i\le n} \in \mathcal{E}^n\})$.

Quel est l'intérêt de la tribu cylindrique ?

PROPOSITION 5.15. *Soient $X = (X_i)_{i\in\mathbb{N}^*}$, E un ensemble tel que $(E, \mathcal{E})$ soit mesurable et $\mathcal{C}$ la tribu cylindrique sur $E^{\mathbb{N}^*}$.*

Alors X est une variable aléatoire à valeurs dans $(E^{\mathbb{N}^*}, \mathcal{C})$ si et seulement si pour tout $i \in \mathbb{N}^*$, X_i est une variable aléatoire à valeurs dans $(E, \mathcal{E})$.

PREUVE

$\Longrightarrow$ Si $B \in \mathcal{E}$, on a pour tout $i \in \mathbb{N}^*$: $\{X_i \in B\} = \{X \in A_{i,B}\} \in \mathcal{A}$ puisque $A_{i,B} \in \mathcal{C}$.

$\Longleftarrow$ $\left\{X \in A_{(t_1,\dots,t_n),(A_1,\dots,A_n)}\right\} = \bigcap_{i=1}^n (X_{t_i} \in B_i) \in \mathcal{A}$ puisque pour tout $i \in \mathbb{N}^*$, $(X_{t_i} \in B_i) \in \mathcal{A}$ comme X_{t_i} est une variable aléatoire.

□

Donc finalement, $X = (X_i)_{i\in\mathbb{N}^*}$ définie en début de partie est bien une variable aléatoire.

Il y a un théorème plus abstrait (non constructif) pour garantir l'existence de suites de variables aléatoires :

THÉORÈME 5.16. [THÉORÈME D'EXTENSION DE KOLMOGOROV]
Soit, pour $n \geq 1$, ν_n une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. On suppose que les $(\nu_n)_n$ sont compatibles, *c'est-à-dire que :*

$$\forall n \in \mathbb{N}^*, \forall (A_i)_{1\leq i\leq n} \in \mathcal{B}(\mathbb{R})^n, \nu_{n+1}(A_1 \times \cdots \times A_n \times \mathbb{R}) = \nu_n(A_1 \times \cdots \times A_n)$$

Alors il existe $X = (X_i)_{i\in\mathbb{N}^*}$ à valeurs dans $(\mathbb{R}^n, \mathcal{C})$ telle que :

$$\forall n \in \mathbb{N}^*, (X_1, \dots, X_n) \sim \nu_n$$

REMARQUE 5.17. On retrouve le résultat précédent en posant $\nu_n = \otimes_{i=1}^n \mu_i$.
On a compatibilité puisque :

$$(\otimes_{i=1}^{n+1}\mu_i)\left(\prod_{i=1}^n A_i \times \mathbb{R}\right) = \prod_{i=1}^n \mu_i(A_i) \times \underbrace{\mu(\mathbb{R})}_{=1} = (\otimes_{i=1}^n \mu_i)\left(\prod_{i=1}^n A_i\right)$$

CHAPITRE 6

Limites presque sûres

L'un des objectifs des probabilités est d'essayer d'identifier des événements qui ont lieu tout le temps à partir de l'observation de nombreux aléas $(X_i)_{i\in I}$.

On se place sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

I. Lemme de Borel-Cantelli

Définition 6.1. [Limite supérieure d'une famille d'événements]
Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements. On définit sa limite supérieure :

$$\begin{aligned}\limsup_{n\to+\infty} A_n &= \{\omega\in\Omega \,|\, \omega \text{ est dans une infinité des } (A_n)_{n\in\mathbb{N}}\}\\ &= \{\omega\in\Omega \,|\, \forall n\in\mathbb{N}, \exists p\geq n, \omega\in A_p\}\\ &= \bigcap_{n\geq 1}\bigcup_{p\geq n} A_p\end{aligned}$$

Remarque 6.2. On a donc :

$$\limsup_{n\to+\infty} A_n = \left\{\omega\in\Omega \,|\, \sum_{n=1}^{+\infty} \mathbb{1}_{\omega\in A_n} = +\infty\right\} = \left\{\sum_{n=1}^{+\infty}\mathbb{1}_{A_n} = +\infty\right\}$$

Lemme 6.3. [Lemme de Borel-Cantelli]
Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.

(1) Si $\sum_{n\in\mathbb{N}}\mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_{n\to+\infty} A_n) = 0$.
On a donc presque sûrement un nombre fini de $(A_n)_{n\in\mathbb{N}}$ réalisés.

*(2) Si $\sum_{n\in\mathbb{N}}\mathbb{P}(A_n) = +\infty$ et **si les $(A_n)_{n\in\mathbb{N}}$ sont indépendants**, alors on a $\limsup_{n\to+\infty} A_n$ p.s..*
On a donc presque sûrement un nombre infini de $(A_n)_{n\in\mathbb{N}}$ réalisés.

Preuve

(1) On a $\mathbb{E}[\sum_{n\in\mathbb{N}}\mathbb{1}_{A_n}] \underset{\text{Fubini}}{=} \sum_{n\in\mathbb{N}}\mathbb{E}[\mathbb{1}_{A_n}] = \sum_{n\in\mathbb{N}}\mathbb{P}(A_n) < +\infty$.
Donc $\sum_{n\in\mathbb{N}}\mathbb{1}_{A_n}$ est positive et d'espérance finie donc $\sum_{n\in\mathbb{N}}\mathbb{1}_{A_n} < +\infty$ p.s.. Ce qui conclut

puisque $(\sum_{n\in\mathbb{N}} \mathbb{1}_{A_n} < +\infty)$ est le complémentaire de $\limsup_{n\to+\infty} A_n$.

(2) On a $1 - \mathbb{P}(\limsup_{n\to+\infty} A_n) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \bigcap_{p\geq n} (A_p)^c\Big)$ puis :

$$\mathbb{P}\Big(\bigcap_{p\geq n} (A_p)^c\Big) \leq \mathbb{P}\Big(\bigcap_{p\geq n}^{k} (A_p)^c\Big) \underset{\text{indép.}}{=} \prod_{p=n}^{k} \mathbb{P}((A_p)^c) = \exp\Big(\sum_{n=p}^{k} \ln(1-\mathbb{P}(A_p))\Big)$$

Donc $\mathbb{P}(\bigcap_{p\geq n}(A_p)^c) \leq \exp\Big(\sum_{n=p}^{+\infty} \ln(1-\mathbb{P}(A_p))\Big) \leq \mathrm{e}^{-\sum_{p=n}^{+\infty} \mathbb{P}(A_p)} = \mathrm{e}^{-\infty} = 0.$

On en déduit $\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \bigcap_{p\geq n} (A_p)^c\Big) \leq \sum_{n\in\mathbb{N}} 0 = 0$ donc $\limsup_{n\to+\infty} A_n$ p.s..

□

I. A. Application aux nombres univers

Soit $b \geq 2$ un entier.
Soit x un nombre réel positif. On regarde sa décomposition en base b :

$$x = x_0 + \sum_{i\geq 1} \frac{x_i}{b^i} \quad x_0 \in \mathbb{N} \quad x_i \in \{0, \dots, b-1\}$$

où $x_0 \in \mathbb{N}$ et $\forall i \in \mathbb{N}^*, x_i \in [\![0, b-1]\!]$ et les $(x_i)_{i\in\mathbb{N}}$ sont non tous égaux à $b-1$ à partir d'un certain rang.
Soit $\ell \geq 1$. Tout élément m de $[\![0, b-1]\!]^\ell$ est un *mot* de longueur ℓ.
Fixons un tel m. On s'intéresse au nombre d'occurrences de m dans x :

$$N_m(x) = \mathrm{card}(\{i \in \mathbb{N}^* \mid (x_i, \dots, x_{i+\ell-1}) = m\})$$

On dit qu'un nombre x est *univers en base* b si :

$$\forall \ell \geq 1, \forall m \in [\![0, b-1]\!]^\ell, N_m(x) = +\infty$$

PROPOSITION 6.4. *Si $X \sim \mathcal{U}([0,1])$, alors on a $(\forall b \geq 2, X$ est univers en base $b)$ p.s..*

REMARQUE 6.5.

- Ce résultat dit que tout nombre assez générique est univers mais il est très dur d'en exhiber.
- Cela reste vrai pour n'importe quelle loi à densité par rapport à la mesure de LEBESGUE.
  Soit $A = \{x \mid x$ n'est pas univers en toute base$\}$.
  On a $\mathrm{Leb}(A \cap [0,1]) = 0$ et de même $\mathrm{Leb}(A) = 0$.
  Si μ est une probabilité de densité f, alors

$$\mu(\{x \mid x \text{ est univers en toute base}\}) = 1 - \mu(A) = 1 - \int_A f(x)\,dx = 1$$

PREUVE Il suffit de prouver que pour $b \geq 2$, puis $\ell \geq 1$, puis $m \in [\![0, b-1]\!]^\ell$ fixés, on a $N_m(X) = +\infty$ p.s., puisqu'alors comme toute intersection dénombrable d'événements presque sûrs est presque sûre, on aura $\bigcap_{b\geq 2} \bigcap_{\ell \geq 1} \bigcap_{m \in [\![0,b-1]\!]^\ell} \{N_m(X) = +\infty\}$ p.s..

RAPPEL 6.6. Soit, pour $i \in \mathbb{N}^*$, $X_i = \lfloor b^i X \rfloor - b \lfloor b^{i-1} X \rfloor$, alors les $(X_i)_{i\in\mathbb{N}^*}$ sont indépendantes et identiquement distribuées, de loi uniforme sur $[\![0, b-1]\!]$.

Pour $p \in \mathbb{N}$, on considère l'événement $A_p = \{(X_{p\ell+1}, \dots, X_{p\ell+\ell}) = m\}$. Les $(X_i)_{i\in\mathbb{N}^*}$ sont indépendants donc par regroupement les $(A_p)_{p\in\mathbb{N}}$ le sont aussi.
Comme A_p est l'événement "le mot apparaît en position $p\ell+1$", on a $\limsup_{p\to+\infty} A_p \subseteq \{N_m(X) = +\infty\}$.
Et puisque $\mathbb{P}(A_p) = \mathbb{P}((X_{p\ell+1}, \dots, X_{p\ell+\ell}) = m) = \frac{1}{b^\ell}$, on a $\sum_{p\in\mathbb{N}} \mathbb{P}(A_p) = +\infty$.
Donc par le lemme de BOREL-CANTELLI (2), on a $\mathbb{P}(\limsup_{p\to+\infty} A_p) = 1$. Ainsi, $N_m(x) = +\infty$ p.s..
□

II. Loi du 0-1 de KOLMOGOROV

On peut déduire du lemme de BOREL-CANTELLI que :

$$(A_i)_{i\in\mathbb{N}} \text{ indépendants} \quad \Longrightarrow \quad \mathbb{P}\Big(\limsup_{n\to+\infty} A_n\Big) \in \{0,1\}$$

(la valeur 0 ou 1 dépendant du critère $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ est fini ou non).

REMARQUE 6.7. $\mathbb{P}(\limsup_{n\to+\infty} A_n)$ ne dépend pas des N premiers événements pour tout $N \in \mathbb{N}$.

On cherche à généraliser ce résultat.

DÉFINITION 6.8. [TRIBU TERMINALE]

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires.
On définit, pour $k \in \mathbb{N}^*$, $\mathcal{T}_k = \sigma(\{X_n \mid n \geq k\})$ puis $\mathcal{T}$ la *tribu terminale* (ou *tribu de queue*) par :

$$\mathcal{T} = \bigcap_{k\in\mathbb{N}^*} \downarrow \mathcal{T}_k$$

$\mathcal{T}$ est bien une tribu (en tant qu'intersection de tribus), qui ne dépend pas d'un nombre fini de $(X_i)_{i\in\mathbb{N}}$.

THÉORÈME 6.9. [LOI DU 0-1 DE KOLMOGOROV]
Si les $(X_i)_{i\in\mathbb{N}}$ sont indépendants, alors :

$$\boxed{\forall A \in \mathcal{T}, \mathbb{P}(A) \in \{0,1\}}$$

EXEMPLE 6.10. Si les $(A_i)_{i\in\mathbb{N}}$ sont indépendants, on pose $X_i = \mathbb{1}_{A_i}$ et alors la loi du 0-1 s'applique et

$$\forall k, \left\{\limsup_{n\to+\infty} A_n\right\} = \left\{\sum_{i=1}^{+\infty} \mathbb{1}_{A_i} = +\infty\right\} = \left\{\sum_{i=k}^{+\infty} \mathbb{1}_{A_i} = +\infty\right\} \in \mathcal{T}_k$$

Donc $\limsup_{n\to+\infty} A_n \in \mathcal{T}$ et par la loi du 0-1 :

$$\mathbb{P}\Big(\limsup_{n\to+\infty} A_n\Big) \in \{0,1\}$$

EXEMPLE 6.11. [ÉVÉNEMENTS DANS $\mathcal{T}$]
Dans $\mathcal{T}$, on a par exemple $\limsup_{i\to+\infty} X_i \leq 0$, $\sum_{i\in\mathbb{N}} \frac{X^i}{2^i} < +\infty$ et $\mathrm{card}(\{i \in \mathbb{N} \mid X_i \in A\}) = +\infty$.

PREUVE Soit $A \in \mathcal{T}$. Soit $\mathcal{M} = \{B \in \mathcal{A} \mid B \text{ est indépendant de } A\}$.

On vérifie que $\mathcal{M}$ est une classe monotone (exercice).
Par regroupement, $\sigma(X_1,\dots,X_{k-1})$ est indépendante de $\mathcal{T}_k = \sigma(X_k, X_{k+1},\dots)$.
Or, $A \in \mathcal{T}_k$. Donc $\sigma(X_1,\dots,X_{k-1}) \subseteq \mathcal{M}$.
Ainsi $\mathcal{C} = \bigcup_{k\geq 1} \uparrow \sigma(X_1,\dots,X_k) \subseteq \mathcal{M}$, et $\mathcal{C}$ est stable par intersection finie, donc

$$\mathcal{T} \subseteq \sigma(\mathcal{C}) \underset{(1)}{=} \mathcal{M}(\mathcal{C}) \underset{(2)}{\subseteq} \mathcal{M}$$

par le lemme classes monotones (1) et puisque $\mathcal{M}$ est une classe monotone (2).
Ainsi $A \in \mathcal{M}$, donc A est indépendant de lui-même, donc $\mathbb{P}(A) \in \{0,1\}$. □

II. A. Application à la récurrence de la marche aléatoire simple sur $\mathbb{Z}$

Une *marche aléatoire* se construit à partir d'une suite de variables aléatoires $(X_i)_{i\in\mathbb{N}^*}$ indépendantes et identiquement distribuées, appelés *incréments de la marche*.
La marche aléatoire issue de x, d'incréments $(X_i)_{i\in\mathbb{N}^*}$ est le processus $S = (S_n)_{n\in\mathbb{N}}$ défini par $S_0 = x$ puis par récurrence sur $n \in \mathbb{N}$, $S_{n+1} = S_n + X_n$. On a alors :

$$\boxed{\begin{cases} S_0 = x \\ S_n = x + \sum_{i=1}^n X_i \quad \text{pour } n \in \mathbb{N}^* \end{cases}}$$

On parle de *marche aléatoire simple* sur $\mathbb{Z}$ si les $(X_i)_{i\in\mathbb{N}^*}$ sont des variables aléatoires indépendantes et identiquement distribuées de loi de RADEMACHER de paramètre $\frac{1}{2}$ ($\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$).

Soit S la marche aléatoire simple issue de 0 :

$$\begin{cases} S_0 = 0 \\ S_n = \sum_{i=1}^n X_i \quad \text{pour } n \in \mathbb{N}^* \end{cases}$$

L'objectif est de comprendre le comportement de S en temps long.
Que dire du nombre de passages en 0 de S ? Est-il infini ou fini ?

PROPOSITION 6.12.

$$(\limsup_{n\to+\infty} S_n = +\infty, \liminf_{n\to+\infty} S_n = -\infty) \text{ p.s.}$$

Cela signifie que la marche oscille. On en déduit :

COROLLAIRE 6.13.

$$(\forall x \in \mathbb{Z}, \mathrm{card}(\{n \in \mathbb{N} \,|\, S_n = x\}) = +\infty) \text{ p.s.}$$

On dit que la marche est *récurrente*.

PREUVE D'après la proposition, S passe une infinité de fois de valeurs supérieures à x à des valeurs inférieures à x, et donc passe une infinité de fois par x. □

PREUVE Soit B l'événement "la marche est bornée". Montrons que $\mathbb{P}(B) = 0$.

On a $B = (\limsup_{n\to+\infty} S_n < +\infty, \liminf_{n\to+\infty} S_n > -\infty) = \bigcup_{p\geq 1} \uparrow (\forall n \in \mathbb{N}, -p \leq S_n \leq p)$.
Or :

$$\begin{aligned}(\forall n \in \mathbb{N}, -p \leq S_n \leq p) &\subseteq (\forall n, m \in \mathbb{N}, S_n - S_m \leq 2p)\\ &\subseteq (\forall n \in \mathbb{N}, S_{n+2p+1} - S_n \neq 2p+1)\\ &= (\forall n \in \mathbb{N}, (X_{n+1} = \cdots = X_{n+2p+1}) \neq (1, \ldots, 1))\end{aligned}$$

Mais les $(X_i)_{i\in\mathbb{N}^*}$ sont des variables aléatoires indépendantes valant ± 1 avec probabilité $\frac{1}{2}$.
Ainsi, les événements $A_\ell = \{(X_{\ell(2p+1)+1}, \ldots, X_{\ell(2p+1)+2p+1}) = (1, \ldots, 1)\}$ sont indépendants et de probabilité $\frac{1}{2^{2p+1}}$, donc $\sum_{\ell\in\mathbb{N}} \mathbb{P}(A_\ell) = +\infty$.
Par le lemme de BOREL-CANTELLI, on a $\limsup_{\ell\to+\infty} A_\ell$ p.s..
Or, $\{\forall n \in \mathbb{N}, (X_{n+1}, \ldots, X_{n+2p+1}) \neq (1, \ldots, 1)\} \subseteq \limsup_{\ell\to+\infty} A_\ell^{\mathsf{c}}$, donc :

$$\mathbb{P}(\forall n \in \mathbb{N}, -p \leq S_n \leq p) = 0$$

Et donc $\mathbb{P}(B) = 0$.

Soit $A_+ = (\limsup_{n\to+\infty} S_n = +\infty)$. On a pour tout $k \in \mathbb{N}$:

$$A_+ = \Big(\limsup_{n\geq k} S_n = +\infty\Big) = \Big(\limsup_{n\geq k} (S_n - S_k) = +\infty\Big) = \Big(\limsup_{n\geq k} (X_{k+1} + \cdots + X_n) = +\infty\Big) \in \mathcal{T}_k$$

Donc $(\limsup_{n\to+\infty} S_n = +\infty) \in \mathcal{T}$, d'où $\mathbb{P}(\limsup_{n\to+\infty} S_n = +\infty) \in \{0, 1\}$.

Supposons $\mathbb{P}(\limsup_{n\to+\infty} S_n = +\infty) = 0$.
Les $(X_i)_{i\in\mathbb{N}^*}$ ont même loi que $(-X_i)_{i\in\mathbb{N}^*}$, donc S et $-S$ ont même loi. D'où

$$\mathbb{P}\Big(\liminf_{n\to+\infty} S_n = -\infty\Big) = \mathbb{P}\Big(\limsup_{n\to+\infty} -S_n = -\infty\Big) = \mathbb{P}\Big(\limsup_{n\to+\infty} S_n = +\infty\Big) = 0$$

Donc :

$$1 = \mathbb{P}(B^{\mathsf{c}}) = \mathbb{P}\Big(\limsup_{n\to+\infty} S_n = +\infty \text{ ou } \liminf_{n\to+\infty} S_n = -\infty\Big) \leq \mathbb{P}\Big(\limsup_{n\to+\infty} S_n = +\infty\Big) + \mathbb{P}\Big(\liminf_{n\to+\infty} S_n = -\infty\Big) = 0$$

Ce qui est absurde. Ainsi $\mathbb{P}(\limsup_{n\to+\infty} S_n = +\infty) = 1$. □

III. Loi forte des grands nombres

THÉORÈME 6.14. [LOI FORTE DES GRANDS NOMBRES]
Soit $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires réelles indépendantes et identiquement distribuées. On suppose que les $(X_i)_{i\in\mathbb{N}^*}$ admettent une espérance finie (c'est-à-dire sont L^1). Alors :

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X_1] \text{ p.s.}$$

REMARQUE 6.15.

- On dit que la *moyenne empirique* tend vers la *moyenne théorique*.
- La loi forte des grands nombres est vraie sur $\mathbb{R}^d$ (en l'appliquant sur chaque coordonnée).

PREUVE *Commençons par une preuve dans le cas où les variables aléatoires sont L^4.*
On peut supposer que pour tout $i \in \mathbb{N}^*$, $\mathbb{E}[X_i] = 0$ (quitte à considérer les $(X_i - \mathbb{E}[X_i])_{i\in\mathbb{N}^*}$, qui restent indépendantes et identiquement distribuées).
De plus on pose $\sigma_2 = \mathbb{E}[X_1^2]$, $m_4 = \mathbb{E}[X_1^4]$ et $S_n = \sum_{i=1}^n X_i$ pour tout entier n.

Pour tout $n \in \mathbb{N}$, on a $S_n \in L^4$ et :

$$\mathbb{E}[S_n^4] = \mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i\Big)^4\Big] = \sum_{1\le i_1<\cdots<i_4\le n} \mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}]$$

Or, si l'un des indices i_k n'apparaît qu'une fois : $\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}] \underset{\text{indép}}{=} \mathbb{E}[X_{i_k}]\mathbb{E}[\prod_{j\neq k} X_{i_j}] = 0$.
Ainsi les seuls termes non nuls de la somme sont ceux où les i_k sont tous égaux ou forment deux paires :

$$\mathbb{E}[S_n^4] = \sum_{1\le i\le n} \mathbb{E}[X_i^4] + 3\sum_{1\le i,j\le n \mid i\neq j} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \le nm_4 + 3n(n-1)\sigma_2^2 \le Cn^2$$

Alors :

$$\mathbb{E}\Big[\sum_{n\ge1}\Big(\frac{S_n}{n}\Big)^4\Big] \underset{\text{FUBINI}}{=} \sum_{n\ge1}\frac{\mathbb{E}[S_n^4]}{n^4} \le \sum_{n\ge1}\frac{C}{n^2} < +\infty$$

Donc $\sum_{n\ge1}\left(\frac{S_n}{n}\right)^4$ est une variable aléatoire positive d'espérance finie, donc $\sum_{n\ge1}\left(\frac{S_n}{n}\right)^4 < +\infty$ p.s..
Donc le terme général de la série tend vers 0 presque sûrement, c'est-à-dire $\frac{S_n}{n} \underset{n\to+\infty}{\longrightarrow} 0$ p.s.. □

PREUVE *Faisons la preuve dans le cas général.*

Il suffit de montrer que $X_1 + \cdots + X_n$ ne croît pas linéairement.
Pour $a > 0$, soit $M = \sup_{n\in\mathbb{N}}(X_1 + \cdots + X_n - na)$.

1. Montrons qu'il suffit de prouver que $\mathbb{P}(M < +\infty) > 0$. Si c'est le cas :

$$\begin{aligned}\{M < +\infty\} &= \Big\{\sup_{n\in\mathbb{N}}(X_1 + \cdots + X_n - na) < +\infty\Big\}\\ &= \Big\{\sup_{n\ge k}(X_1 + \cdots + X_n - na) - (X_1 + \cdots + X_k) < +\infty\Big\}\\ &= \Big\{\sup_{n\ge k}(X_{k+1} + \cdots + X_n - na) < +\infty\Big\}\end{aligned}$$

Donc $\{M < +\infty\} \in \mathcal{T}_k$. Ainsi $\{M < +\infty\} \in \mathcal{T}$ la tribu terminale. Mais les $(X_i)_{i\in\mathbb{N}^*}$ sont indépendantes donc la loi du 0-1 de KOLMOGOROV s'applique et $\mathbb{P}(M < +\infty) \in \{0, 1\}$.
Et donc par hypothèse $\mathbb{P}(M < +\infty) = 1$.

On a $X_1 + \cdots + X_n - na \leq M < +\infty$ p.s.. Donc $\frac{1}{n}\sum_{i=1}^n X_i \leq \frac{M}{n} + a$ p.s..
Ainsi $\limsup_{n\to+\infty} \frac{X_1+\cdots+X_n}{n} \leq a$ p.s. puisque $M < +\infty$ p.s..
Cela est vrai pour tout $a > 0$, c'est en particulier le cas pour $a_p = \frac{1}{p}$ pour tout $p \in \mathbb{N}^*$.
On en déduit que $\limsup_{n\to+\infty} \frac{X_1+\cdots+X_n}{n} \leq 0$ p.s. (en tant qu'intersection dénombrable des événements presque surs $(\limsup_{n\to+\infty} \frac{X_1+\cdots+X_n}{n} \leq a_p)_{p\in\mathbb{N}^*}$).

Le problème étant symétrique, on montre de même que $\liminf_{n\to+\infty} \frac{X_1+\cdots+X_n}{n} \geq 0$ p.s.. Et donc $\frac{X_1+\cdots+X_n}{n} \underset{n\to+\infty}{\longrightarrow} 0$ p.s..

2. Montrons donc que pour tout $a > 0$, M est fini avec probabilité positive.

NOTE 6.16. Moralement c'est vrai car $X_1+\cdots+X_n-na = (X_1-a)+\cdots+(X_n-a)$ donc à la n-ème étape, on rajoute un terme d'espérance $-a < 0$. La suite a donc tendance à diverger vers $-\infty$.

Supposons par l'absurde que cela ne soit pas le cas et donc que $M = +\infty$ p.s..
Posons $M_k = \sup_{0\leq n\leq k}(X_1 + \cdots + X_n - na)$.
$(M_k)_{k\in\mathbb{N}}$ est une suite croissante qui tend vers M. Donc $M_k \underset{k\to+\infty}{\longrightarrow} +\infty$ p.s..
Essayons de relier M_{k+1} à M_k :

$$\begin{aligned} M_{k+1} &= \max(0, X_1 - a, X_1 + X_2 - 2a, \ldots, X_1 + \cdots + X_{k+1} - (k+1)a) \\ &= \max\Big(0, X_1 - a + \underbrace{\sup_{0\leq n\leq k}(X_2 + \cdots + X_{n+1} - na)}_{=\tilde{M}_k}\Big) \end{aligned}$$

où $\tilde{M}_k$ s'obtient à partir de la suite $(X_i)_{i\geq 2}$ de la même manière que M_k s'obtient à partir des $(X_i)_{i\geq 1}$. Mais puisque $(X_i)_{i\geq 2} \overset{\mathcal{L}}{=} (X_i)_{i\geq 1}$, on a $\tilde{M}_k \overset{\mathcal{L}}{=} M_k$.
On en déduit que $\tilde{M} = \sup_{k\in\mathbb{N}} \tilde{M}_k = +\infty$ p.s.. On a 2 arguments possibles :

- on peut écrire que $(\tilde{M}_k)_{k\geq 0} = F((X_{i+1}))_{i\geq 1}$ où F est telle que $(M_k)_{k\geq 0} = F((X_i))_{i\geq 1}$, et donc $(\tilde{M}_k)_{k\geq 0} \overset{\mathcal{L}}{=} (M_k)_{k\geq 0}$ puis en passant au sup $\tilde{M} \overset{\mathcal{L}}{=} M$. Ainsi $\tilde{M} = +\infty$ p.s..
- Pour tout $t \in \mathbb{R}$, on a

$$\mathbb{P}(\tilde{M} \leq t) = \mathbb{P}\Big(\sup_{k\in\mathbb{N}} \tilde{M}_k \leq t\Big) = \lim_{k\in\mathbb{N}} \uparrow \mathbb{P}(\tilde{M}_k \leq t) = \lim_{k\in\mathbb{N}} \uparrow \mathbb{P}(M_k \leq t) = \mathbb{P}\Big(\sup_{k\in\mathbb{N}} M_k \leq t\Big) = 0$$

On a $M_{k+1} - \tilde{M}_k = \max(-\tilde{M}_k, X_1 - a)$. Et puisque

$$\mathbb{E}[|M_k|] = \mathbb{E}\Big[\Big|\sup_{0\leq n\leq k} X_1 + \cdots + X_n - na\Big|\Big] \leq \sum_{n=0}^{k} \mathbb{E}[|X_1 + \cdots + X_n - na|) < +\infty$$

, on peut prendre l'espérance et on a puisque $(M_k)_{k\in\mathbb{N}}$ est strictement croissante :

$$\mathbb{E}[M_{k+1} - \tilde{M}_k] = \mathbb{E}[M_{k+1}] - \mathbb{E}[\tilde{M}_k] = \mathbb{E}[M_{k+1}] - \mathbb{E}[M_k] = \mathbb{E}[M_{k+1} - M_k] \geq 0$$

Par ailleurs :

$$\max(-\tilde{M}_k, X_1 - a) \underset{k\to+\infty}{\longrightarrow} \max(-\tilde{M}, X_1 - a) = X_1 - a \text{ p.s.}$$

Or, $|\max(-\tilde{M}_k, X_1 - a)| \leq |X_1 - a|$ p.s., donc pas convergence dominée :

$$\mathbb{E}[\max(-\tilde{M}_k, X_1 - a)] \underset{k\to+\infty}{\longrightarrow} \mathbb{E}[X_1 - a] < 0$$

Ce qui est absurde puisqu'on a vu que le terme de gauche est positif. Donc $\mathbb{P}(M < +\infty) > 0$.

□

REMARQUE 6.17. La loi forte des grands nombres dit que des observations répétées d'une expérience donne accès à des informations sur la loi.

- En probabilité, on a un espace de probabilité donné ou un modèle (par exemple $(X_i)_{i\in\mathbb{N}^*}$ variables aléatoires indépendantes suivant une loi $\mathcal{B}(3/4)$) et on essaye d'en déduire des informations sur les $(X_i)_{i\in\mathbb{N}^*}$. On a alors $\frac{X_1+\cdots+X_n}{n} \underset{n\to+\infty}{\longrightarrow} \frac{3}{4}$ p.s..
- En statistique, la loi sur les espaces de probabilité est inconnue.
  Soient $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{B}(\theta)$ avec θ inconnu. On souhaite retrouver θ.
  Par exemple, pour $i \in \mathbb{N}^*$, $X_i = \begin{cases} 1 & \text{si le } i\text{-ème habitant veut voter "oui"} \\ 0 & \text{sinon} \end{cases}$.
  On connaît par contre un échantillon de taille n : $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n) \in \{0,1\}^n$.
  La loi forte des grands nombres nous dit que $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i \underset{n\to+\infty}{\longrightarrow} \theta$ p.s..
  On dit que $\hat{\theta}_n$ est un *estimateur* de $\hat{\theta}$.

L'hypothèse L^1 est nécessaire :

PROPOSITION 6.18. *Soient $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées telle que $\mathbb{E}[|X_1|] = +\infty$.*

Alors la probabilité que la limite de $\frac{X_1+\cdots+X_n}{n}$ existe dans $\mathbb{R}$ est nulle.

PREUVE Si $\frac{X_1+\cdots+X_n}{n}$ converge, alors $\lim_{n\to+\infty} \frac{X_1+\cdots+X_n}{n} - \frac{X_1+\cdots+X_{n-1}}{n-1}\frac{n-1}{n} = 0$, d'où $\lim_{n\to+\infty} \frac{X_n}{n} = 0$.
On a :

$$+\infty = \mathbb{E}[|X_1|] = \int_0^\infty \mathbb{P}(|X_1| \geq t)\,dt = \sum_{n=0}^{+\infty} \int_n^{n+1} \mathbb{P}(|X_1| \geq t)\,dt \leq \sum_{n=0}^{+\infty} \mathbb{P}(|X_1| \geq n) = \sum_{n=0}^{+\infty} \mathbb{P}(|X_{n+1}| \geq n)$$

D'après le lemme de BOREL-CANTELLI (les événements $(|X_{n+1}| \geq n)_{n\in\mathbb{N}^*}$ sont indépendants et la somme de leur probabilités diverge), on a $\limsup_{n\to+\infty} \{|X_{n+1}| \geq n\}$ p.s..
Donc $\left(\limsup_{n\to+\infty} \frac{|X_n|}{n} \geq 1\right)$ p.s. et alors $\mathbb{P}\left(\frac{X_n}{n} \underset{n\to+\infty}{\longrightarrow} 0\right) = 0$. □

III. A. Méthode d'intégration de MONTE-CARLO

Soit $f : [0,1] \longrightarrow \mathbb{R}$ de classe L^1. On souhaite calculer $\int_0^1 f(x)\,dx$ à partir de valeurs de f.
On simule $(U_i)_{i\in\mathbb{N}^*}$ variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[0,1]$ et on renvoie pour un certain entier n :

$$\frac{f(U_1) + \cdots + f(U_n)}{n}$$

Alors la loi forte des grands nombres assure que presque sûrement cette quantité converge vers $\mathbb{E}[f(U_1)] = \int_0^1 f(x)\,dx$. Donc pour n assez grand, on obtient une bonne estimation.

III. B. Application aux nombres normaux

Soit $b \geq 2$ une base. Soit $x \in \mathbb{R}$. On considère sa décomposition en base b : $x = x_0 + \sum_{i\geq 1} \frac{x_i}{b^i}$.
Pour $\ell \in \mathbb{N}^*$, soit $m \in [\![0, b-1]\!]^\ell$. On regarde $N_m^n(x) = \mathrm{card}(\{1 \leq i \leq n \,|\, (x_i, x_{i+1}, \ldots, x_{i+\ell-1}) = m\})$ le nombre d'occurrences de m dans x avant $n \in \mathbb{N}$.
On dit que x est *normal* en base b si $\forall \ell \in \mathbb{N}^*, \forall m \in [\![0, b-1]\!]^\ell, \frac{1}{n} N_m^n(x) \underset{n\to+\infty}{\longrightarrow} \frac{1}{b^\ell}$.

PROPOSITION 6.19. *Si $Z \sim \mathrm{Leb}_{|[0,1]}$ alors (Z est normal en toute base) p.s..*

REMARQUE 6.20. Comme pour les nombres univers ceci implique que si μ_Z a une densité par rapport à la mesure de LEBESGUE alors (Z est normal en toute base) p.s..

PREUVE Il suffit de prouver qu'à $b \geq 2$, puis $\ell \geq 1$, puis $m \in [\![0, b-1]\!]^\ell$ fixés, on a $\frac{N_m^n(Z)}{n} \underset{n\to+\infty}{\longrightarrow} \frac{1}{b^\ell}$ p.s. (le résultat en découle alors par intersections dénombrables d'événements presque surs).

Notons pour $i \in \mathbb{N}^*$, $X_i = \mathbb{1}_{m \text{ apparaît en position } i} = \mathbb{1}_{(Z_i,\ldots,Z_{i+\ell-1})=m}$.
Alors $\mathbb{P}(X_i = 1) = \mathbb{P}(Z_i, \ldots, Z_{i+\ell-1} = m) = \frac{1}{b^\ell}$, donc $X_i \sim \mathcal{B}(1/b^\ell)$.
Mais les $(X_i)_{i\geq 1}$ ne sont pas indépendantes.
A $n \in \mathbb{N}$ fixé, considérons alors $N_m^{n,r}(Z) = X_r + X_{r+\ell} + \cdots + X_{r+\ell(n-1)}$ pour $1 \leq r \leq \ell$.
On a $\sum_{1\leq r\leq \ell} N_m^{n,r}(Z) = N_m^{n\ell}(Z)$.
De plus, $N_m^{n,r}(Z)$ est composé de $(X_i)_{i\in\mathbb{N}^*}$ indépendants par regroupement.
Donc par la loi forte des grands nombres :

$$\frac{N_m^{n,r}(Z)}{n} \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X_r] = \frac{1}{b^\ell} \text{ p.s.}$$

Et en sommant il vient $\frac{N_m^{n\ell}(Z)}{n} \underset{n\to+\infty}{\longrightarrow} \frac{\ell}{b^\ell}$ p.s..
On veut la limite de $\frac{N_m^n(Z)}{n}$. On procède par encadrement :

$$\underbrace{\frac{\lfloor n/\ell \rfloor}{n}}_{\underset{n\to+\infty}{\longrightarrow} 1/\ell} \underbrace{\frac{N_m^{\lfloor n/\ell \rfloor \ell}(Z)}{\lfloor n/\ell \rfloor}}_{\underset{n\to+\infty}{\longrightarrow} \ell/b^\ell} \leq \frac{N_m^n(Z)}{n} \leq \underbrace{\frac{\lfloor n/\ell \rfloor + 1}{n}}_{\underset{n\to+\infty}{\longrightarrow} 1/\ell} \underbrace{\frac{N_m^{(\lfloor n/\ell \rfloor + 1)\ell}(Z)}{\lfloor n/\ell \rfloor + 1}}_{\underset{n\to+\infty}{\longrightarrow} \ell/b^\ell}$$

Et donc $\frac{N_m^n(Z)}{n} \underset{n\to+\infty}{\longrightarrow} \frac{1}{b^\ell}$ p.s.. □

EXEMPLE 6.21. [CONSTRUCTION D'UNE MESURE SINGULIÈRE]

On se donne $(X_i)_{i\in\mathbb{N}^*}$ suite de variables aléatoires i.i.d.de loi $\mathcal{B}(1/3)$. On pose $X = \sum_{i\geq 1} \frac{X_i}{2^i} \in [0, 1]$ p.s..

- La loi de X n'a pas d'atome puisque pour tout $x \in [0, 1]$ puis $k \in \mathbb{N}$:

$$\mathbb{P}(X = x) = \mathbb{P}(\forall i \in \mathbb{N}^*, X_i = x_i) \leq \mathbb{P}(\forall i \leq k, X_i = x_i) = \left(\frac{2}{3}\right)^k \underset{k\to+\infty}{\longrightarrow} 0$$

– Mais elle n'a pas non plus de densité par rapport à la mesure de LEBESGUE.
Posons $N = \{x \mid x \text{ est normal en base } 2\}$. On a

$$\frac{N_1^n(X)}{n} = \frac{X_1 + \cdots + X_n}{n} \underset{n\to+\infty}{\overset{\text{LFGN}}{\longrightarrow}} \frac{1}{3} \neq \frac{1}{2} \text{ p.s.}$$

Donc $\mu_X(N) = \mathbb{P}(X \in N) = 0$. Puis si $\mu_X \geq f \,\mathrm{Leb}$, alors :

$$0 = \mu_X(N) \geq \int_N f(x)\,dx \underset{\mathrm{Leb}(N^c)=0}{=} \int f(x)\,dx$$

Donc $f = 0$ p.p..

CHAPITRE 7

Convergences de variables aléatoires

I. Différents types de convergences

DÉFINITION 7.1. [TYPES DE CONVERGENCES]

Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires, toutes à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

- On dit que $(X_n)_{n\in\mathbb{N}}$ tend *presque sûrement* vers X et on note $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$ si $X_n \underset{n\to+\infty}{\longrightarrow} X$ p.s., c'est-à-dire si $\boxed{\mathbb{P}\Big(X_n \underset{n\to+\infty}{\longrightarrow} X\Big) = 1}$.
- Si $p \geq 1$, on dit que $(X_n)_{n\in\mathbb{N}}$ tend *dans* L^p vers X et on note $X_n \xrightarrow[n\to+\infty]{L^p} X$ si

$$\forall n \in \mathbb{N}, X_n \in L^p, \quad X \in L^p, \quad \text{et} \quad \boxed{\mathbb{E}[\|X_n - X\|^p] \underset{n\to+\infty}{\longrightarrow} 0}$$

- On dit que $(X_n)_{n\in\mathbb{N}}$ tend *en proba* vers X et on note $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ si

$$\forall \varepsilon > 0, \boxed{\mathbb{P}(\|X_n - X\| > \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0}$$

- On dit que $(X_n)_{n\in\mathbb{N}}$ tend *en loi* vers X et on note $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ si

$$\forall f : \mathbb{R}^d \longrightarrow \mathbb{R} \text{ continue bornée}, \boxed{\mathbb{E}[f(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)]}$$

REMARQUE 7.2.
- La convergence en loi est très différente des autres ! Les deux types de convergence les plus importants sont la convergence presque sûre et la convergence en loi.
- On peut définir ces convergences sur des espaces plus exotiques :
  - → la convergence p.s. pour des variables aléatoires à valeurs dans $(E, \mathcal{E})$ quelconque,
  - → la convergence L^p sur n'importe quel espace normé,
  - → la convergence en probabilité sur n'importe quel espace métrique,
  - → la convergence en loi sur n'importe quel espace topologique $\mathcal{X}$.

Il faut faire très attention à la convergence en loi !

$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ ne dit pas que pour n assez grand, X_n est proche de X, mais que μ_{X_n} est proche de μ_X :

$$\mathbb{E}[f(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f(X)] \iff \int f(x)\,d\mu_{X_n}(x) \xrightarrow[n\to+\infty]{} \int f(x)\,d\mu_X(x)$$

La convergence en loi de la variable aléatoire ne dépend donc que de sa loi !

EXEMPLE 7.3. Soit $X \sim \mathcal{B}(1/2)$. Posons $X_n = X$ pour $n \in \mathbb{N}$. Alors :

$$\forall n \in \mathbb{N}, \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] = \mathbb{E}[f(1-X)] \quad \text{car } 1-X \sim \mathcal{B}(1/2)$$

Donc $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} 1-X$ mais $\forall n \in \mathbb{N}, |X_n - (1-X)| = 1$ p.s. donc les variables aléatoires ne se rapprochent pas ! On a aussi $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$, donc on n'a pas unicité de la limite !

De plus, les propriétés usuelles échouent :

- Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$, on n'a pas forcément $X_n + Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} X + Y$ (alors que cela est vrai pour les autres convergences).
  Considérer dans l'exemple précédent $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} 1-X$ et $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ mais $2X_n \xrightarrow[n\to+\infty]{\mathcal{L}} 1$!
- Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$, on n'a pas forcément $X_n Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} XY$ (alors que cela est vrai pour la convergence p.s. et la convergence en proba),
- Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y$, on n'a pas forcément $(X_n, Y_n) \xrightarrow[n\to+\infty]{\mathcal{L}} (X, Y)$.

II. Liens entre les différents types de convergence

PROPOSITION 7.4. *Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires, toutes à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

*Soient $p, q \in [1, +\infty]$ tels que **p < q**. On a les implications suivantes :*

(1) $X_n \xrightarrow[n\to+\infty]{L^q} X \implies X_n \xrightarrow[n\to+\infty]{L^p} X$

(2) $X_n \xrightarrow[n\to+\infty]{L^p} X \implies X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$

(3) $X_n \xrightarrow[n\to+\infty]{p.s.} X \implies X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$

(4) $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \implies X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$

PREUVE [POINT (1) DE LA PROPOSITION 7.4]

Soit r tel que $\frac{1}{q/p} + \frac{1}{r} = 1$. On a, par l'inégalité de HÖLDER :

$$\mathbb{E}[\|X_n - X\|^p] = \mathbb{E}[\|X_n - X\|^p \times 1] \leq \mathbb{E}[\|X_n - X\|^{p(q/p)}]^{p/q} \mathbb{E}[1^r]^{1/r} = \mathbb{E}[\|X_n - X\|^q]^{p/q} \xrightarrow[n\to+\infty]{} 0$$

□

REMARQUE 7.5. $x \longmapsto |x|^{q/p}$ est convexe. On retrouve le résultat en appliquant l'inégalité de JENSEN :

$$\mathbb{E}[\|X_n - X\|^p] = (\mathbb{E}[\|X_n - X\|^p]^{q/p})^{p/q} \leq \mathbb{E}[\|X_n - X\|^{p(q/p)}]^{p/q} = \mathbb{E}[\|X_n - X\|^q]^{p/q} \underset{n\to+\infty}{\longrightarrow} 0$$

La réciproque de (1) est fausse en général :

EXEMPLE 7.6. Soit en effet $Y_n \sim \mathcal{B}(1/n)$ et $X_n = n^\alpha Y_n$ pour $n \in \mathbb{N}^*$ et pour un α à déterminer.

Alors $\forall n \in \mathbb{N}^*, \mathbb{E}[|X_n|^p] = n^{\alpha p - 1}$. Donc $X_n \underset{n\to+\infty}{\overset{L^p}{\longrightarrow}} 0 \iff \alpha < \frac{1}{p}$.
Pour $\frac{1}{q} < \alpha < \frac{1}{p}$, on a $X_n \underset{n\to+\infty}{\overset{L^p}{\longrightarrow}} 0$ mais $X_n \underset{n\to+\infty}{\overset{L^q}{\not\to}} 0$.

REMARQUE 7.7. On verra cependant que dans certains cas, la réciproque est vraie.

PREUVE [POINT (3) DE LA PROPOSITION 7.4]

On a par convergence dominée, puisque $\mathbb{1}_{\|X_n - X\| > \varepsilon} \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} 0$ et $\mathbb{1}_{\|X_n - X\| > \varepsilon} \leq 1$:

$$\mathbb{P}(\|X_n - X\| > \varepsilon) = \mathbb{E}[\mathbb{1}_{\|X_n - X\| > \varepsilon}] \underset{n\to+\infty}{\longrightarrow} 0$$

□

Là encore, la réciproque est fausse ! Donnons deux contre-exemples :

EXEMPLE 7.8.

- Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes telles que $X_n \sim \mathcal{B}(1/n)$ pour $n \in \mathbb{N}^*$.
  On a $\mathbb{P}(|X_n| > \varepsilon) \leq \mathbb{P}(X_n = 1) = \frac{1}{n} \underset{n\to+\infty}{\longrightarrow} 0$. Donc $X_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0$.
  Mais $\sum_{n\in\mathbb{N}} \mathbb{P}(X_n = 1) = \sum_{n\in\mathbb{N}} \frac{1}{n} = +\infty$. Par indépendance, le lemme de BOREL-CANTELLI donne $\limsup_{n\to+\infty} X_n = 1$ p.s., donc $X_n \underset{n\to+\infty}{\overset{\text{p.s.}}{\not\to}} 0$.
  Ce qui implique que $(X_n)_{n\in\mathbb{N}}$ n'a pas de limite p.s., puisque si $X_n \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} Z$ alors $X_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} Z$ et donc par unicité des limites en probabilité, on aurait $Z = 0$ p.s..
- Soit $U \sim \mathcal{U}([0,1])$. Posons $X_{2^n+k} = \mathbb{1}_{\frac{k}{2^n} \leq U < \frac{k+1}{2^n}}$ pour $n \geq 0$ et $0 \leq k \leq 2^n - 1$.
  On a $X_{2^n+k} \sim \mathcal{B}(1/2^n)$.
  Pour tout $i \geq 1$, il existe (n_i, k_i) tel que $i = 2^{n_i} + k_i$ (on notera alors que $n_i \underset{i\to+\infty}{\longrightarrow} +\infty$ puisque $2^{n_i} + k_i \leq 2 \times 2^{n_i}$). Donc $\mathbb{P}(|X_i| > \varepsilon) \leq \mathbb{P}(X_i = 1) = \frac{1}{2^{n_i}} \underset{i\to+\infty}{\longrightarrow} 0$.
  Or, pour tout $x \in [0,1[$, on a $\forall n \geq 1, \exists k \mid x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[$. Ainsi $\forall \omega, \forall n \geq 1, \exists k \mid X_{2^n+k} = 1$.
  Donc ($X_i = 1$ pour une infinité de i) p.s. donc $X_i \underset{i\to+\infty}{\overset{\text{p.s.}}{\not\to}} 0$.

Cependant, le résultat suivant affirme que l'on peut tout de même remonter un peu :

PROPOSITION 7.9. *Si $X_n \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} X$, alors il existe une sous-suite $(n_k)_{k\in\mathbb{N}}$ telle que $X_{n_k} \underset{k\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} X$.*

PREUVE

$$\forall k, \exists n_k \,|\, \forall n \geq n_k : \mathbb{P}(\|X_n - X\| \geq 1/k) \leq \frac{1}{k^2}$$

On peut choisir $n_{k+1} > n_k$. Alors $\sum_{k\geq 1} \mathbb{P}(\|X_{n_k} - X\| > 1/k) < +\infty$.
Donc par le lemme de BOREL-CANTELLI : (pour k assez grand, $\|X_{n_k} - X\| \leq \frac{1}{k}$) p.s..
Donc $(\limsup_{n\to+\infty} \|X_{n_k} - X\| = 0)$ p.s., c'est-à-dire $X_{n_k} \xrightarrow[k\to+\infty]{\text{p.s.}} X$. □

On va voir deux applications de cette proposition. Tout d'abord :

$$\left\{ \begin{array}{l} X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \\ C = \sup_{n\in\mathbb{N}} \mathbb{E}[\|X_n^p\|] < +\infty \end{array} \right. \implies X \in L^p \quad \text{et} \quad \mathbb{E}[\|X\|^p] \leq C$$

En effet, il existe une sous-suite $(n_k)_{k\in\mathbb{N}}$ telle que $X_{n_k} \xrightarrow[k\to+\infty]{\text{p.s.}} X$. Alors, par le lemme de FATOU :

$$\mathbb{E}[\|X\|^p] = \mathbb{E}\Big[\liminf_{k\in\mathbb{N}} \|X_{n_k}\|^p\Big] \leq \liminf_{k\in\mathbb{N}} \mathbb{E}[\|X_{n_k}\|^p] \leq C < +\infty$$

Ensuite, considérons $\mathcal{F} = \{f : [0,1] \longrightarrow [0,1] \text{ mesurable}\} / \sim$ où $\sim$ est la relation d'équivalence "égalité presque partout par rapport à la mesure de LEBESGUE".
Alors il n'existe pas de métrique sur $\mathcal{F}$ telle que :

$$f_n \xrightarrow[n\to+\infty]{\text{p.p.}} f \iff d(f_n, f) \xrightarrow[n\to+\infty]{} 0$$

NOTE 7.10. On a le même résultat si l'espace d'arrivée des fonctions est $\mathbb{R}$. L'obstruction vient même si l'on ne demande pas à d de vérifier l'inégalité triangulaire !

En effet, soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires de $([0,1], \mathcal{B}([0,1]), \text{Leb})$ dans $[0,1]$ telles que $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$ mais $X_n \xrightarrow[n\to+\infty]{\text{p.s.}\not\to} X$. Alors $d(X_n, X) \xrightarrow[n\to+\infty]{\not\to} 0$.
Donc il existe $\varepsilon > 0$ puis une suite strictement croissante d'entiers $(n_k)_{k\in\mathbb{N}}$ telle que

$$\forall k \in \mathbb{N}, d(X_{n_k}, X) > \varepsilon$$

Or, $X_{n_k} \xrightarrow[k\to+\infty]{\mathbb{P}} X$ donc il existe une sous-suite $(X_{n_{k(r)}})_{r\in\mathbb{N}}$ de $(X_{n_k})_{k\in\mathbb{N}}$ qui tend presque surement vers X. Alors $d(X_{n_{k(r)}}, X) \xrightarrow[r\to+\infty]{} 0$, ce qui est absurde. Donc d n'existe pas !

PREUVE [POINT (2) DE LA PROPOSITION 7.4]

Par le point (1), il suffit de le vérifier dans le cas $p = 1$. On a par l'inégalité de MARKOV :

$$\mathbb{P}(\|X_n - X\| \geq \varepsilon) \leq \frac{\mathbb{E}[\|X_n - X\|]}{\varepsilon} \xrightarrow[n\to+\infty]{} 0$$

□

A nouveau, la réciproque est fausse !
Il suffit de considérer $U \sim \mathcal{U}([0,1])$ et de poser $X_n = n\mathbb{1}_{U\leq\frac{1}{n}}$ pour $n \in \mathbb{N}^*$.
On a $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} 0$ mais $X_n \xrightarrow[n\to+\infty]{L^1\not\to} 0$ puisque $\mathbb{E}[|X_n|] = 1$.

PROPOSITION 7.11. *Soit $r > p > 1$. On a :*

$$\left\{ \begin{array}{l} X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \\ \sup_{n\in\mathbb{N}} \mathbb{E}[\|X_n\|^r] < +\infty \end{array} \right. \implies X_n \xrightarrow[n\to+\infty]{L^p} X$$

PREUVE On sait déjà que $X \in L^r$. De plus, pour tout entier n :

$$\begin{aligned} \mathbb{E}[\|X_n - X\|^p] &= \mathbb{E}[\|X_n - X\|^p \mathbb{1}_{\|X_n - X\| \leq \varepsilon}] + \mathbb{E}[\|X_n - X\|^p \mathbb{1}_{\|X_n - X\| > \varepsilon}] \\ &\leq \varepsilon^p + \mathbb{E}[\|X_n - X\|^{p(r/p)}]^{p/r} \mathbb{E}[\mathbb{1}_{\|X_n - X\| > \varepsilon}]^{\alpha}]^{1/\alpha} \end{aligned}$$

en appliquant l'inégalité de HÖLDER avec α vérifiant $\frac{p}{r} + \frac{1}{\alpha} = 1$.

Or, d'une part $\mathbb{E}[\|X_n - X\|^{p(r/p)}]^{p/r} = \mathbb{E}[\|X_n - X\|^r]^{p/r} \leq \|X_n - X\|_r \leq \|X_n\|_r + \|X\|_r \leq C$.
Et d'autre part $\mathbb{E}[\mathbb{1}_{\|X_n - X\| > \varepsilon}]^{\alpha}]^{1/\alpha} = \mathbb{P}(\|X_n - X\| > \varepsilon)^{1/\alpha} \xrightarrow[n\to+\infty]{} 0$.

Finalement, $\limsup_{n\to+\infty} \mathbb{E}[\|X_n - X\|^p] \leq \varepsilon^p$, donc $X_n \xrightarrow[n\to+\infty]{L^p} X$. □

On cherche un critère pour rendre équivalent la convergence en probabilité et la convergence L^1. On verra que cela serra très utile pour les martingales.

DÉFINITION 7.12. [VARIABLES ALÉATOIRES UNIFORMÉMENT INTÉGRABLES]
On dit qu'une suite de variables aléatoires réelles $(X_i)_{i\in I}$ de L^1 est uniformément intégrable si :

$$\lim_{L\to+\infty} \sup_{i\in I} \mathbb{E}[|X_i| \mathbb{1}_{|X_i|>L}] = 0$$

EXEMPLE 7.13. [VARIABLES ALÉATOIRES UNIFORMÉMENT INTÉGRABLES]

- Si $X \in L^1$, $\{X\}$ est uniformément intégrable puisque par convergence dominée :
$$\mathbb{E}[|X| \mathbb{1}_{|X|>L}] \xrightarrow[L\to+\infty]{} 0$$
car $|X| \mathbb{1}_{|X|>L} \xrightarrow[L\to+\infty]{} 0$ (comme $X \in L^1$, on a $|X| < +\infty$ p.s.) et $|X| \mathbb{1}_{|X|>L} \leq |X| \in L^1$.
- S'il existe $Y \in L^1$ telle que $\forall i \in I, |X_i| \leq Y$ p.s., alors $(X_i)_{i\in I}$ est uniformément intégrable puisque :
$$\mathbb{E}[|X_i| \mathbb{1}_{|X_i|>L}] \leq \mathbb{E}[|Y| \mathbb{1}_{|Y|>L}]$$
- Si I est fini et les $(X_i)_{i\in I}$ sont L^1 alors $(X_i)_{i\in I}$ est uniformément intégrable (prendre $Y = \sum_{i\in I} |X_i|$).

On a la propriété de caractérisation suivante :

PROPOSITION 7.14.

$$(X_i)_{i\in I} \text{ U.I.} \iff \left\{ \begin{array}{l} C = \sup_{i\in I} \mathbb{E}[|X_i|] < +\infty \\ \lim_{\delta\to 0} \sup_{A \,|\, \mathbb{P}(A)\leq\delta, i\in I} \mathbb{E}[|X_i| \mathbb{1}_A] = 0 \end{array} \right.$$

PREUVE

$\Longrightarrow$ Fixons $\varepsilon > 0$.

- $L \longmapsto \sup_{i\in I} \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|>L}]$ est une fonction décroissante qui tend vers 0 donc est inférieure à ε pour un certain L_ε, et alors :

$$\forall i \in I, \mathbb{E}[|X_i|] = \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|\leq L_\varepsilon}] + \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|>L_\varepsilon}] \leq L_\varepsilon + \varepsilon$$

- Si A est un événement tel que $\mathbb{P}(A) \leq \delta$, on a pour tout $i \in I$:

$$\mathbb{E}[|X_i|\,\mathbb{1}_A] = \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|\leq L}\mathbb{1}_A] + \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|>L}\mathbb{1}_A]$$

Donc $\sup_{i\in I} \mathbb{E}[|X_i|\,\mathbb{1}_A] \leq L\mathbb{P}(A) + \underbrace{\sup_{i\in I} \mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|>L}]}_{\leq \varepsilon \text{ pour } L\geq L_\varepsilon}$.

Prenant $L = L_\varepsilon$ et $\delta = \frac{\varepsilon}{L_\varepsilon}$, on obtient $\sup_{i\in I} \mathbb{E}[|X_i|\,\mathbb{1}_A] \leq 2\varepsilon$.

$\Longleftarrow$ Soit $\varepsilon > 0$. Il existe $\delta > 0$ tel que $\mathbb{P}(A) \leq \delta \Longrightarrow \forall i \in I, \mathbb{E}[|X_i|\,\mathbb{1}_A] \leq \varepsilon$.
Soit $i \in I$. Par l'inégalité de MARKOV, on a $\mathbb{P}(|X_i| > L) \leq \frac{\mathbb{E}[|X_i|]}{L} \leq \frac{C}{L}$. Posons donc $L_\varepsilon = \frac{C}{L}$.
Alors pour $L > L_\varepsilon$: $\mathbb{P}(|X_i| > L) \leq \delta$.
D'où $\mathbb{E}[|X_i|\,\mathbb{1}_{|X_i|>L}] \leq \varepsilon$ en appliquant avec $A = (|X_i| > L)$.

□

PROPOSITION 7.15. *Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires L^1 et X une variable aléatoire.*

Alors :

$$\boxed{X_n \xrightarrow[n\to+\infty]{L^1} X \iff X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \quad \text{et} \quad (X_n)_{n\in\mathbb{N}} \text{ U.I.}}$$

PREUVE

$\Longrightarrow$ On sait que la convergence L^1 implique la convergence en probabilité.
On a $\mathbb{E}[|X_n|] \leq \mathbb{E}[|X_n - X|] + \mathbb{E}[|X|] \longrightarrow_{n\to+\infty} \mathbb{E}[|X|]$, donc $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^1 et

$$\mathbb{E}[|X_n|\,\mathbb{1}_A] \leq \mathbb{E}[|X_n - X|\,\mathbb{1}_A] + \mathbb{E}[|X|\,\mathbb{1}_A]$$

Soit $\varepsilon > 0$. Il existe $n_0 \in \mathbb{N}$ tel que $\forall n \geq n_0, \mathbb{E}[|X_n - X|] \leq \varepsilon$.
Par ailleurs, la famille $(X_n)_{n<n_0} \cup \{X\}$ est uniformément intégrable car c'est une famille finie de L^1. Donc il existe δ tel que pour une variable aléatoire Y de cette famille : $\mathbb{P}(A) \leq \delta \Longrightarrow \mathbb{E}[|Y|\,\mathbb{1}_A] \leq \varepsilon$.
Alors si $\mathbb{P}(A) \leq \delta$:

$$\mathbb{E}[|X_n|\,\mathbb{1}_A] \leq \begin{cases} \varepsilon & \text{si } n < n_0 \\ \mathbb{E}[|X_n - X|\,\mathbb{1}_A] + \mathbb{E}[|X|\,\mathbb{1}_A] \leq 2\varepsilon & \text{sinon} \end{cases}$$

$\Longleftarrow$ $(X_n)_{n\in\mathbb{N}}$ et $\{X\}$ sont deux familles uniformément intégrables donc pour tout $\varepsilon > 0$:

$$\exists \delta > 0 \,|\, \forall A, \mathbb{P}(A) \leq \delta \Longrightarrow \forall n \in \mathbb{N}, \mathbb{E}[|X_n|\,\mathbb{1}_A] \leq \varepsilon \text{ et } \mathbb{E}[|X|\,\mathbb{1}_A] \leq \varepsilon$$

Il existe $n_0 \in \mathbb{N}$ tel que $\mathbb{P}(|X_n - X| > \varepsilon) \leq \delta$ pour $n \geq n_0$. Alors :

$$\begin{aligned}\mathbb{E}[|X_n - X|] &= \mathbb{E}[|X_n - X| \mathbb{1}_{|X_n - X| \leq \varepsilon}] + \mathbb{E}[|X_n - X| \mathbb{1}_{|X_n - X| > \varepsilon}] \\ &\leq \varepsilon + \mathbb{E}[|X_n| \mathbb{1}_{|X_n - X| > \varepsilon}] + \mathbb{E}[|X| \mathbb{1}_{|X_n - X| > \varepsilon}] \\ &\leq 3\varepsilon\end{aligned}$$

□

PREUVE [POINT (4) DE LA PROPOSITION 7.4]
On a pour tout $n \in \mathbb{N}$:

$$\begin{aligned}|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= \left|\mathbb{E}[f(X_n) - f(X)\mathbb{1}_{\|X_n - X\| > \varepsilon}] + \mathbb{E}[f(X_n) - f(X)\mathbb{1}_{\|X_n - X\| \leq \varepsilon}]\right| \\ &\leq 2\|f\|_\infty \mathbb{P}(\|X_n - X\| > \varepsilon) + \mathbb{E}[(f(X_n) - f(X))\mathbb{1}_{\|X_n - X\| \leq \varepsilon}(\mathbb{1}_{\|X\| \leq A} + \mathbb{1}_{\|X\| > A})]\end{aligned}$$

On peut supposer $\varepsilon < 1$. D'où :

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\|f\|_\infty \mathbb{P}(\|X_n - X\| > \varepsilon) + \sup_{\substack{x,y \\ \|x\| \leq A \\ \|x-y\| \leq \varepsilon}} |f(x) - f(y)| + 2\|f\|_\infty \mathbb{P}(\|X\| \geq A)$$

Donc finalement $\limsup_{n \to +\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \sup_{\substack{x,y \\ \|x\| \leq A \\ \|x-y\| \leq \varepsilon}} |f(x) - f(y)| + 2\|f\|_\infty \mathbb{P}(\|X\| \geq A)$.
En faisant tendre ε vers 0, on a le sup qui tend vers 0 par uniforme continuité de f sur $\{x \mid \|x\| \leq A + \varepsilon\}$ qui est compact. Puis en faisant tendre A vers $+\infty$, le second terme tend vers 0.
Ainsi $\mathbb{E}[f(X_n)] \underset{n \to +\infty}{\longrightarrow} \mathbb{E}[f(X)]$. □

La réciproque est encore fausse.
Considérons en effet $X \sim \mathcal{B}(1/2)$ et posons $X_n = X$ pour $n \in \mathbb{N}$.
Alors $X_n \underset{n \to +\infty}{\xrightarrow{\mathcal{L}}} 1 - X$ mais $X_n \underset{n \to +\infty}{\overset{\mathbb{P}}{\nrightarrow}} 1 - X$.

RAPPEL 7.16. On rappelle que $X_n \underset{n \to +\infty}{\xrightarrow{\mathcal{L}}} X$ signifie que pour toute f continue bornée, on a

$$\int f(x)\, d\mu_{X_n}(x) \underset{n \to +\infty}{\longrightarrow} \int f(x)\, d\mu_X(x)$$

Cela a encore un sens si les $(X_n)_{n \in \mathbb{N}}$ ne sont pas sur le même espace de probabilité !

Cependant, on peut faire un *couplage*, c'est-à-dire trouver des variables aléatoires $(Y_n)_{n \in \mathbb{N}}$ qui ont les lois de $(X_n)_{n \in \mathbb{N}}$ et pour lesquelles la convergence est plus forte :

THÉORÈME 7.17. *Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires réelles. On a équivalence entre :*

(i) $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$,

(ii) Pour tout t point de continuité de F_X, on a $F_{X_n}(t) \xrightarrow[n\to+\infty]{\mathcal{L}} F_X(t)$,

(iii) Il existe $(Y_n)_{n\in\mathbb{N}}$ et Y des variables aléatoires réelles telles que :

$$\begin{cases} \forall n\in\mathbb{N}, \mu_{Y_n} = \mu_{X_n} \\ \mu_Y = \mu_X \\ Y_n \xrightarrow[n\to+\infty]{p.s.} Y \end{cases}$$

REMARQUE 7.18.

- Les $(Y_n)_{n\in\mathbb{N}}$ et Y sont un couplage qui renforce la convergence.
- (i) $\iff$ (iii) est le **théorème de représentation de SKOROKHOD**.

PREUVE

(iii) $\implies$ (i) Soit f continue bornée. On a (la convergence presque sûre implique la convergence en loi) :

$$\mathbb{E}[f(X_n)] = \int f(x)\,d\mu_{X_n}(x) = \int f(x)\,d\mu_{Y_n}(x) \xrightarrow[n\to+\infty]{} \int f(x)\,d\mu_Y(x) = \int f(x)\,d\mu_X(x) = \mathbb{E}[f(X)]$$

(i) $\implies$ (ii) Posons, pour $p\in\mathbb{N}$, $f_p(x) = (1-p[x-(t-1/p)]_+)_+$ et $f_p(x) = (1-p(x-t)_+)_+$.
f_p et g_p sont continues bornées et on a $\forall x\in\mathbb{R}, f_p(x) \le \mathbb{1}_{x\le t} \le g_p(x)$. D'où

$$\mathbb{E}[f_p(X_n)] \le \mathbb{E}[\mathbb{1}_{X_n\le t}] \le \mathbb{E}[g_p(X_n)]$$

Or on a $\begin{cases} \mathbb{E}[f_p(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f_p(X)] \ge \mathbb{E}[\mathbb{1}_{X\le t-1/p}] \\ \mathbb{E}[\mathbb{1}_{X_n\le t}] = F_{X_n}(t) \\ \mathbb{E}[g_p(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[g_p(X)] \le \mathbb{E}[\mathbb{1}_{X\le t+1/p}] \end{cases}$

Donc $F_X(t-1/p) \le \liminf_{n\to+\infty} F_{X_n}(t) \le \limsup_{n\to+\infty} F_{X_n}(t) \le F_X(t+1/p)$.
En faisant tendre p vers $+\infty$, et puisque t est un point de continuité, on obtient : $F_{X_n}(t) \xrightarrow[n\to+\infty]{} F_X(t)$.

(ii) $\implies$ (iii) On sait que $F_{X_n} \xrightarrow[n\to+\infty]{} F_X$ sur les points de continuité de F_X.
On se donne $U\sim\mathcal{U}([0,1])$ et on pose $Y_n = F_{X_n}^{<-1>}(U)$ pour $n\in\mathbb{N}$ puis $Y = F_X^{<-1>}(U)$.
On a vu que $Y_n \overset{\mathcal{L}}{=} X_n$ et $Y \overset{\mathcal{L}}{=} X$.

RAPPEL 7.19. $F^{<-1>}(u) = \inf\{t \mid F(t)\ge u\}$ pour $0<u<1$. On a $F(t)\ge u \iff t\ge F^{<-1>}(u)$, ou encore $F(t)<u \iff t<F^{<-1>}(u)$ (contraposée).

REMARQUE 7.20. F est une fonction croissante bornée donc il y a un nombre au plus dénombrable de points de discontinuité. Si $A_p = \{t \mid F(t)\ge F(t^-)+1/p\}$, alors

$$\mathrm{card}(A_p)\times 1/p \le \sum_{t\in A_p} F(t)-F(t^-) \le \lim_{+\infty} F - \lim_{-\infty} F = 1-0 = 1$$

Donc A_p est fini et $\bigcup_{p\in\mathbb{N}^*} A_p = \{\text{points de discontinuité}\}$ est au plus dénombrable.

Donc on peut trouver Γ un ensemble dénombrable de points de continuité dense dans $\mathbb{R}$.

Soit $t \in \Gamma$, regardons l'événement :

$$\begin{aligned}\{Y > t\} = \left\{F_X^{<-1>}(U) > t\right\} &= \{U > F_X(t)\} = \left\{U > \lim_{n\to+\infty} F_{X_n}(t)\right\} \\ &\subseteq \{U > F_{X_n}(t) \text{ pour } n \text{ assez grand}\} = \left\{Y_n = F_{X_n}^{<-1>}(U) > t \text{ pour } n \text{ assez grand}\right\} \\ &\subseteq \left\{\liminf_{n\to+\infty} Y_n \geq t\right\}\end{aligned}$$

Et comme Γ est dense :

$$\left\{\liminf_{n\to+\infty} Y_n < Y\right\} = \bigcup_{t\in\Gamma}\left\{\liminf_{n\to+\infty} Y_n < t < Y\right\} = \bigcup_{t\in\Gamma}\left\{\liminf_{n\to+\infty} Y_n < t, Y > t\right\} = \varnothing$$

Donc $\underline{\liminf_{n\to+\infty} Y_n \geq Y \text{ p.s.}}$.

Regardons l'événement :

$$\begin{aligned}\{Y \leq t\} = \left\{F_X^{<-1>}(U) \leq t\right\} &= \{U \leq F_X(t)\} = \{U = F_X(t)\} \cup \{U < F_X(t)\} \\ &= \{U = F_X(t)\} \cup \{U < F_{X_n}(t) \text{ pour } n \text{ assez grand}\} \\ &\subseteq \{U = F_X(t)\} \cup \{U \leq F_{X_n}(t) \text{ pour } n \text{ assez grand}\} \\ &= \{U = F_X(t)\} \cup \{Y_n \leq t \text{ pour } n \text{ assez grand}\} \\ &\subseteq \{U = F_X(t)\} \cup \left\{\limsup_{n\to+\infty} Y_n \leq t\right\}\end{aligned}$$

Ainsi :

$$\begin{aligned}\left\{\limsup_{n\to+\infty} Y_n > Y\right\} &= \bigcup_{t\in\Gamma}\left\{\limsup_{n\to+\infty} Y_n > t\right\} \bigcap \{t \geq Y\} \\ &\subseteq \bigcup_{t\in\Gamma}\left\{\limsup_{n\to+\infty} Y_n > t\right\} \bigcap \left(\{U = F_X(t)\} \cup \left\{\limsup_{n\to+\infty} Y_n \leq t\right\}\right) \\ &\subset \underbrace{\{U = F_X(t)\}}_{\text{proba nulle}} \bigcup \underbrace{\left\{\limsup_{n\to+\infty} Y_n > t, \limsup_{n\to+\infty} Y_n \leq t\right\}}_{\varnothing}\end{aligned}$$

Donc $\underline{Y \geq \limsup_{n\to+\infty} Y_n \text{ p.s.}}$.

Finalement, on a $Y \leq \liminf_{n\to+\infty} Y_n \leq \limsup_{n\to+\infty} Y_n \leq Y$. Donc $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y$.

□

III. Théorèmes autour de la convergence en loi

III. A. Le théorème de Lévy

On peut caractériser la convergence en loi :

THÉORÈME 7.21. [THÉORÈME DE LÉVY (FAIBLE)]

Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors :

$$\boxed{X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X \iff \forall \lambda \in \mathbb{R}^d, \phi_{X_n}(\lambda) \xrightarrow[n\to+\infty]{} \phi_X(\lambda)}$$

THÉORÈME 7.22. [THÉORÈME DE LÉVY (FORT)]

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

S'il existe $\Psi : \mathbb{R}^d \longrightarrow \mathbb{C}$ continue en 0 et telle que $\forall \lambda \in \mathbb{R}^d, \phi_{X_n}(\lambda) \longrightarrow_{n\to+\infty} \Psi(\lambda)$, alors il existe une variable aléatoire X telle que $\Psi = \phi_X$ et $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$.

REMARQUE 7.23. Dans la version faible, le sens $\Longrightarrow$ est trivial, puisque $\mathrm{e}^{i\langle\lambda|x\rangle}$ est continue bornée.

On va utiliser le lemme suivant qui est un outil pour se ramener à un compact :

LEMME 7.24. *Soit μ une mesure de probabilité sur $\mathbb{R}$. Alors pour tout $A > 0$:*

$$\mu(\{x \mid |x| \geq A\}) \leq \frac{A}{2} \int_{-2/A}^{2/A} |1 - \phi_\mu(t)| \, dt$$

PREUVE Soit $c \geq 0$. On calcule :

$$\begin{aligned}
\int_{-c}^{c} (1 - \phi_\mu(t)) \, dt &= \int_{-c}^{c} \left(1 - \int \mathrm{e}^{itx} \, d\mu(x)\right) dt \\
&= \int_{\substack{-c\leq t\leq c \\ x\in\mathbb{R}}} (1 - \mathrm{e}^{itx}) \, d\mu(x) \, dt \\
&= \int_x \left(\int_{-c}^{c} (1 - \mathrm{e}^{itx}) \, dt\right) d\mu(x) \qquad \text{par le théorème de FUBINI} \\
&= \int_x \left(2c - \left[\frac{e^{itx}}{ix}\right]_{-c}^{c}\right) \mathbb{1}_{x\neq 0} \, d\mu(x) \\
&= 2c \int_x \left(1 - \frac{\sin(cx)}{cx}\right) d\mu(x)
\end{aligned}$$

Donc

$$\begin{aligned}
\int_{-c}^{c} |1 - \phi_\mu(t)| \, dt \geq \left|\int_{-c}^{c} (1 - \phi_\mu(t)) \, dt\right| &= 2c \left|\int_x \left(1 - \frac{\sin(cx)}{cx}\right) d\mu(x)\right| \\
&\geq 2c \int_x \underbrace{\left(1 - \left|\frac{\sin(cx)}{cx}\right|\right)}_{\geq 1/2 \text{ si } |cx|\geq 2} d\mu(x) \geq c \int_x \mathbb{1}_{|cx|\geq 2} \, d\mu(x)
\end{aligned}$$

Ainsi $\mu(\{x \mid |cx| \geq 2\}) \leq \frac{1}{c} \int_{-c}^{c} |1 - \phi_\mu(t)| \, dt$. Prenant $A = 2/c$, on obtient le résultat. □

PREUVE [THÉORÈME 7.21]

Supposons que $\phi_{X_n} \underset{n\to+\infty}{\longrightarrow} \phi_X$ simplement.
Soit $\varepsilon > 0$. On notera $X_n = (X_n(1), \dots, X_n(d))^\top \in \mathbb{R}^d$. On a par le lemme précédent :

$$\mathbb{P}(\underbrace{\|X_n\|_\infty > A}_{=\bigcup_{i=1}^n |X_n(i)>A|}) \leq \sum_{i=1}^d \mathbb{P}(|X_n(i)| > A) \leq \frac{A}{2} \sum_{i=1}^d \int_{-A/2}^{A/2} |1 - \phi_{X_n}(t)| \, dt \tag{7.1}$$

Il existe $A > 0$ tel que $\forall i \in [\![1, d]\!], \forall t \in [-2/A, 2/A], \phi_{X(i)}(t) \geq 1-\varepsilon$ par continuité sous l'espérance (puisque $\mathbb{E}[e^{itx}]$ est continue en 0, de limite 1).
Fixons ce A et regardons la limite du terme de droite dans (7.1) lorsque $n \to +\infty$. Par convergence dominée, cette limite vaut

$$\sum_{i=1}^d \frac{A}{2} \int_{-A/2}^{A/2} |1 - \phi_X(t)| \, dt \leq d2\varepsilon$$

Donc $\mathbb{P}(\|X\|_\infty > A) \leq 2d\varepsilon$. Et en particulier pour n assez grand : $\mathbb{P}(\|X_n\|_\infty > A) \leq 4d\varepsilon$.

Soit f fonction continue bornée. Il existe P_ε polynôme trigonométrique $2A$-périodique (qui dépend de A) tel que $\|f - P_\varepsilon\|_{\infty,[-A,A]} \leq \varepsilon$. Il vient :

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| = |\mathbb{E}[\underbrace{(f-P_\varepsilon)}_{\leq\varepsilon}(X_n)\mathbb{1}_{\|X_n\|\leq A}] + \mathbb{E}[\underbrace{(f-P_\varepsilon)}_{\|f\|_\infty+\|P_\varepsilon\|\leq 2\|f\|_\infty+\varepsilon}(X_n)\mathbb{1}_{\|X_n\|>A}] + \underbrace{\mathbb{E}[P_\varepsilon(X_n)]}_{=\sum_{\lambda_i}\phi_{X_n}(\lambda_i)\underset{n\to+\infty}{\longrightarrow}}$$
$$-\mathbb{E}[P_\varepsilon(X)] - \mathbb{E}[(f-P_\varepsilon)(X)\mathbb{1}_{\|X\|>A}] - \mathbb{E}[(f-P_\varepsilon)(X)\mathbb{1}_{\|X\|\leq A}]|$$

Donc :

$$\limsup_{n\to+\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\varepsilon + (2\|f\|_\infty + \varepsilon)\Big(\limsup_{n\to+\infty} \mathbb{P}(\|X_n\|_\infty > A) + \mathbb{P}(\|X\|_\infty > A)\Big)$$
$$\leq 2\varepsilon + (2\|f\|_\infty + \varepsilon)(4d\varepsilon + 2d\varepsilon) = O(\varepsilon)$$

□

REMARQUE 7.25. Pour la version forte, il manque un théorème qui dit que l'on peut extraire une sous-suite convergente de le suite de $(\mu_{X_n})_{n\in\mathbb{N}}$:

THÉORÈME 7.26. **[THÉORÈME DE PROKHOROV]**
Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de probabilités.
Alors $(\mu_n)_{n\in\mathbb{N}}$ est tendue (c'est-à-dire $\forall \varepsilon > 0, \exists K_\varepsilon$ compact $\mid \sup_{n\in\mathbb{N}} \mu_n(K_\varepsilon^c) \leq \varepsilon$) si et seulement si de toute sous-suite on peut extraire une suite convergente pour la topologie faible.

III. B. Application : le théorème central limite

THÉORÈME 7.27. [THÉORÈME CENTRAL LIMITE]
Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées de carré intégrable. Notons $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathrm{Var}(X_1)$. Alors :

$$\boxed{\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - m) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Z \sim \mathcal{N}(0, \sigma^2)}$$

REMARQUE 7.28. La loi forte des grands nombres donne que $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{\text{p.s.}} m$.

PREUVE Si $Y \in L^2$ est réelle, alors $t \longmapsto \mathrm{e}^{iYt}$ (intégrable puisque $|.| \leq 1$) est deux fois dérivable, de dérivée $t \longmapsto iY\,\mathrm{e}^{iYt}$ (intégrable puisque $|.| \leq |Y|$) et de dérivée seconde $t \longmapsto -Y^2\,\mathrm{e}^{iYt}$ (intégrable puisque $|.| \leq Y^2$).
Donc on peut dériver deux fois sous l'espérance $\phi_Y : t \longmapsto \mathbb{E}[\mathrm{e}^{iYt}]$ et on a

$$\forall t \in \mathbb{R}, \phi_Y'(t) = i\mathbb{E}[Y\,\mathrm{e}^{iYt}] \quad \text{et} \quad \phi_Y''(t) = -\mathbb{E}[Y^2\,\mathrm{e}^{iYt}]$$

On a alors le développement limité :

$$\phi_Y(t) \underset{t\to 0}{=} 1 + it\mathbb{E}[Y] - \frac{t^2}{2}\mathbb{E}[Y^2] + o(t^2)$$

Ici, posons $S = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - m)$. On a, pour $t \in \mathbb{R}$:

$$\begin{aligned}
\phi_S(t) = \mathbb{E}[\mathrm{e}^{itS}] &= \prod_{i=1}^n \mathbb{E}\Big[\,\mathrm{e}^{it\frac{(X_i-m)}{\sqrt{n}}}\,\Big] && \text{par indépendance}\\
&= \left(\phi_{X_i-m}\left(\frac{t}{\sqrt{n}}\right)\right)^n\\
&\underset{n\to+\infty}{=} \left(1 + \frac{it}{\sqrt{n}}\mathbb{E}[X_i - m] - \frac{t^2}{2n}\mathbb{E}[(X_i-m)^2] + o(1/n)\right)^n\\
&= \exp\left(n\ln\left(1 - \frac{t^2}{2n}\sigma^2 + o(1/n)\right)\right) = \mathrm{e}^{-\frac{t^2\sigma^2}{2}+o(1)}\\
&\xrightarrow[n\to+\infty]{} \mathrm{e}^{-\frac{t^2\sigma^2}{2}} = \phi_{\mathcal{N}(0,\sigma^2)}(t)
\end{aligned}$$

Le théorème de LÉVY donne alors le résultat. □

EXEMPLE 7.29. [APPLICATION EN STATISTIQUE]

On a une population de N individus. Une proportion p inconnue est "pour", le reste est "contre".

Pour estimer p on interroge n personnes choisies de manière indépendantes et uniforme, et on pose $X_i = 1$ si la personne est "pour", $X_i = 0$ sinon, pour $i \in [\![1, n]\!]$.
Les $(X_i)_{1\leq i\leq n}$ sont indépendantes et identiquement distribuées de loi $\mathcal{B}(p)$. Par la loi forte des grands nombres :

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to+\infty]{\text{p.s.}} p$$

On se pose alors la question du nombre minimal n d'échantillons permettant d'assurer que l'on fait une erreur de plus de 3% sur p dans au plus un sondage sur 20 ?
Mathématiquement, cela se traduit par la recherche de n minimal tel que $\mathbb{P}(|\hat{p}_n - p| > 0,03) \leq 0,05$.

Or, on a :

$$\begin{aligned}\mathbb{P}(|\hat{p}_n - p| > 0,03) &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > 0,03\right)\\ &= \mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)\right| \times \frac{1}{\sqrt{p(1-p)}} > \frac{0,03\sqrt{n}}{\sqrt{p(1-p)}}\right)\\ &\leq \mathbb{P}\Bigg(\Bigg|\underbrace{\frac{1}{\sqrt{n}\sqrt{p(1-p)}}\sum_{i=1}^{n}(X_i - m)}_{Y_n}\Bigg| > \frac{0,03\sqrt{n}}{\sqrt{1/4}}\Bigg)\end{aligned}$$

D'après le théorème central limite, on a $Y_n \underset{n\to+\infty}{\longrightarrow} Z \sim \mathcal{N}(0,1)$. Donc :

$$\begin{aligned}\mathbb{P}(|\hat{p}_n - p| > 0,03) &\leq \mathbb{P}(|Y_n| > 0,06\sqrt{n}) = 1 - F_{Y_n}(0,06\sqrt{n}) + F_{Y_n}(-0,06\sqrt{n})\\ &\underset{n \text{ grand}}{\simeq} 1 - F_Z(0,06\sqrt{n}) + F_Z(-0,06\sqrt{n}) = 2\mathbb{P}(Z > 0,06\sqrt{n})\end{aligned}$$

On veut $2(1 - F_{\mathcal{N}(0,1)}(0,06\sqrt{n})) \leq 0,05$, c'est-à-dire $F_{\mathcal{N}(0,1)}(0,06\sqrt{n}) \leq 0,975$.

Dans les tables, on obtient alors $\boxed{n \geq \left(\frac{1,96}{0,06}\right)^2 \simeq 1067}$.

REMARQUE 7.30. n ne dépend pas de la taille de la population N ! Cela vient du fait que $n\hat{p}_n \sim \mathcal{B}(n,p)$.

THÉORÈME 7.31. [THÉORÈME CENTRAL LIMITE EN DIMENSION d]
Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ indépendantes et identiquement distribuées de carré intégrable. Notons $m = \mathbb{E}[X_i] \in \mathbb{R}^d$ et $\Gamma = \mathrm{Cov}(X_1(i), X_1(j))_{1\leq i,j\leq d} \in \mathcal{M}_d(\mathbb{R})$. Alors :

$$\boxed{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} Z \sim \mathcal{N}_{\mathbb{R}^d}(0,\Gamma)}$$

PREUVE Posons $S = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)$. On a :

$$\forall \lambda \in \mathbb{R}^d, \phi_S(\lambda) = \mathbb{E}\left[\exp\left(i\langle\lambda \mid \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)\rangle\right)\right] = \phi_{\langle\lambda|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-m)\rangle}(1)$$

Or, $\langle\lambda \mid \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)\rangle = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\langle\lambda \mid X_i\rangle - n\mathbb{E}[\langle\lambda \mid X_i\rangle]\right)$.
Les $(\langle\lambda \mid X_i\rangle)_{1\leq i\leq n}$ sont des variables aléatoires indépendantes et identiquement distribuées L^2.
On a :

$$\forall 1 \leq i \leq n, \begin{cases}\mathbb{E}(\langle\lambda \mid X_i\rangle) = \sum_k \lambda(k)\mathbb{E}[X_i(k)] = \langle\lambda \mid m\rangle\\ \mathbb{E}(\langle\lambda \mid X_i\rangle) = \sum_{k,\ell}\lambda(k)\lambda(\ell)\,\mathrm{Cov}(X_i(k), X_i(\ell)) = {}^t\lambda\Gamma\lambda\end{cases}$$

On applique le théorème central limite dans $\mathbb{R}$:

$$\langle\lambda \mid \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m)\rangle \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} Z \sim \mathcal{N}(0, {}^t\lambda\Gamma\lambda)$$

Donc $\forall \lambda \in \mathbb{R}^d, \phi_S(\lambda) \underset{n\to+\infty}{\longrightarrow} \phi_{Z_\lambda}(1) = \mathrm{e}^{-{}^t\lambda\Gamma\lambda/2} = \phi_{\mathcal{N}(0,\Gamma)}(\lambda)$. On conclut par le théorème de Lévy :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - m) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Z \sim \mathcal{N}_{\mathbb{R}^d}(0, \Gamma)$$

□

EXEMPLE 7.32. [COMPORTEMENT DE MARCHES ALÉATOIRES]

On reprend les notations du paragraphe II. A. dans la section II. du chapitre 6, en supposant désormais que les $(X_i)_{i\geq 1}$ sont à valeurs dans $\mathbb{R}^d$ où d est la *dimension* (fixe) de la marche. On fait de plus l'hypothèse que les $(X_i)_{i\geq 1}$ sont L^2 et on note $m = \mathbb{E}[X_1]$ et $\Gamma = \mathrm{Cov}(X_1(i), X_1(j))_{1\leq i,j\leq d}$. On considère S la marche aléatoire issue de 0. On rappelle que S est définie par

$$\begin{cases} S_0 = 0 \\ S_n = \sum_{i=1}^{n} X_i \quad \text{pour } n \in \mathbb{N}^* \end{cases}$$

Quel est le comportement de S_n lorsque n est grand ?

$\underline{m \neq 0}$ On a $\frac{S_n}{n} \underset{n\to+\infty}{\longrightarrow} m$ p.s. donc $S_n \sim nm$ p.s..
On dit que l'on a un comportement *ballistique*. En particulier $\|S_n\| \underset{n\to+\infty}{\longrightarrow} +\infty$ p.s. et

$$\forall R > 0, \mathrm{card}(\{n \mid \|S_n\| \leq R\}) < +\infty \text{ p.s.}$$

Le nombre de visites dans la boule de rayon R est fini presque sûrement.

$\underline{m = 0}$ La loi forte des grands nombres ne suffit pas puisque l'on a juste que $\frac{S_n}{n} \underset{n\to+\infty}{\longrightarrow} 0$ p.s.. On a par le théorème central limite $\frac{S_n}{\sqrt{n}} \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Z \sim \mathcal{N}(0, \Gamma)$, donc $\left\|\frac{S_n}{\sqrt{n}}\right\| \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \|Z\|$ (par continuité de la norme).
L'ordre de grandeur de S_n est donc $\sqrt{n}$, mais cela ne signifie pas que $\|S_n\| \underset{n\to+\infty}{\longrightarrow} +\infty$ p.s..
On ne peut donc pas conclure sur le nombre de visites dans $B(0, R)$.

CHAPITRE 8

Espérance conditionnelle

I. Echauffements

I. A. Mesurabilité de variables aléatoires

Si $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ est une variable aléatoire, la plus petite tribu (incluse dans $\mathcal{A}$) telle que X soit mesurable est $\sigma(X) = X^{-1}(\mathcal{E}) = \{X^{-1}(C) \,|\, C \in \mathcal{E}\} = \{X \in C \,|\, C \in \mathcal{E}\}$.

On dit que X est *$\mathcal{B}$-mesurable* si $X : (\Omega, \mathcal{B}) \longrightarrow (E, \mathcal{E})$ est mesurable, c'est-à-dire si $\sigma(X) \subset \mathcal{B}$.
On dit que X est *Y-mesurable* si X est $\sigma(Y)$-mesurable, c'est-à-dire si $\sigma(X) \subset \sigma(Y)$.

LEMME 8.1. *Soient X une variable aléatoire réelle et Y une variable aléatoire à valeurs dans $(E, \mathcal{E})$.*
Alors X est Y-mesurable si et seulement si il existe $h : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable telle que

$$\boxed{X = h(Y)}$$

REMARQUE 8.2.

- C'est un lemme crucial qui provient de la théorie de la mesure (on n'a pas besoin d'avoir une mesure de probabilité),
- Le lemme est vrai dans $\mathbb{R}^d$ (il suffit d'appliquer coordonnée par coordonnée).
- On peut ainsi parler de mesurabilité sans parler de tribu.

PREUVE

$\Longleftarrow$ Si $X = h(Y)$ alors (l'inclusion découlant du fait que h soit mesurable) :

$$\sigma(X) = \sigma(h(Y)) = \{h(Y) \in B \,|\, B \in \mathcal{B}(\mathbb{R})\} = \{Y \in h^{-1}(B) \,|\, B \in \mathcal{B}(\mathbb{R})\} \subseteq \{Y \in C \,|\, C \in \mathcal{E}\} = \sigma(Y)$$

$\Longrightarrow$ On procède en plusieurs étapes :

- Si $X = \mathbb{1}_A$ pour un événement A : on a $A = X^{-1}(\{1\}) \in \sigma(X) \subseteq \sigma(Y)$.
  Donc il existe $C \in \mathcal{E}$ tel que $A = Y^{-1}(C)$.
  On pose $h : E \longrightarrow \mathbb{R}$ définie par $h = \mathbb{1}_C$. Alors h est mesurable et $h(Y) = \mathbb{1}_A = X$.

- Si $X = \sum_{i=1}^{d} \lambda_i \mathbb{1}_{A_i}$ est une variable aléatoire étagée.
  On peut supposer les $(A_i)_{1\leq i\leq d}$ disjoints. En effet si $A_i \cap A_j \neq \varnothing$ pour un couple $i \neq j$, on a
  $$\lambda_i \mathbb{1}_{A_i} + \lambda_j \mathbb{1}_{A_j} = \lambda_i \mathbb{1}_{A_i \setminus A_j} + \lambda_j \mathbb{1}_{A_j \setminus A_j} + (\lambda_i + \lambda_j)\mathbb{1}_{A_i \cap A_j}$$
  En remplaçant A_i et A_j par $A_i \setminus A_j$, $A_j \setminus A_i$ et $A_i \cap A_j$, on obtient une famille disjointe et la fonction reste étagée.
  On peut aussi supposer les $(\lambda_i)_{1\leq i\leq d}$ deux à deux disjoints : en effet si $\lambda_i = \lambda_j$ pour un couple $i \neq j$, on a $\lambda_i \mathbb{1}_{A_i} + \lambda_j \mathbb{1}_{A_j} = \lambda_i \mathbb{1}_{A_i \cup A_j}$, donc c'est bon si on remplace A_i et A_j par $A_i \cup A_j$.
  Soit $i \in [\![1, d]\!]$. On a donc $A_i = X^{-1}(\{\lambda_i\}) \in \sigma(X) \subseteq \sigma(Y)$, donc il existe $C_i \in \mathcal{E}$ tel que $A_i = Y^{-1}(C_i)$.
  On peut choisir les $(C_i)_{1\leq i\leq d}$ disjoints quitte à remplacer C_i par $C_i \setminus (\bigcap_{j=0}^{i-1} C_j)$ (en construisant par récurrence), puisqu'en effet si $y \in C_i \cap C_j$ est dans l'image de Y, alors on aurait $Y^{-1}(\{y\}) \subset Y^{-1}(C_i) \cap Y^{-1}(C_j) = A_i \cap A_j = \varnothing$, ce qui est impossible. Donc on a bien $Y^{-1}(C_i) = Y^{-1}\Big(C_i \setminus (\bigcap_{j=0}^{i-1} C_j)\Big)$.
  Posons alors $h : y \longmapsto \sum_{i=1}^{d} \lambda_i \mathbb{1}_{y\in C_i}$. On a
  $$\omega \in A_i \Longrightarrow Y(\omega) \in C_i \Longrightarrow h(Y(\omega)) = \lambda_i = X(\omega)$$
- Si X est positive : alors $X : (\Omega, \sigma(Y)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est limite croissante de fonctions étagées $(X_n : (\Omega, \sigma(Y)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})))_{n\in\mathbb{N}}$. Donc il existe $(h_n)_{n\in\mathbb{N}}$ fonctions mesurables étagées telles que pour tout entier n, $X_n = h_n(Y)$. On pose alors
  $$h : y \longmapsto \begin{cases} \lim_{n\to+\infty} \uparrow h_n(y) & \text{lorsqu'elle existe} \\ 0 & \text{sinon} \end{cases}$$
  h est alors mesurable en tant que limite simple de mesurables, et pour tout $\omega \in \Omega$:
  $$X(\omega) = \lim_{n\to+\infty} \uparrow X_n(\omega) = \lim_{n\to+\infty} \uparrow h_n(Y(\omega)) = h(Y(\omega))$$
  (on remarque que la limite existe pour $y = Y(\omega)$).
- Si X est quelconque, on décompose X sous la forme $X = X^+ - X^-$ et on obtient le résultat en utilisant le point précédent.

□

I. B. Intuition sur l'espérance conditionnelle

L'objectif est de définir $\mathbb{E}[X \mid Y]$, appelé *espérance conditionnelle de X sachant Y* qui n'est pas un nombre comme $\mathbb{E}[X]$ mais la variable aléatoire Y-mesurable la plus proche possible de X. C'est donc une fonction de Y (que l'on notera $h(Y)$) qui est la meilleure prédiction possible sur X.

Exemple 8.3. Soit Y une variable aléatoire discrète ($Y \in (y_i)_{i\in I}$ avec I fini ou dénombrable et $\mathbb{P}(Y = y_i) > 0$ pour tout $i \in I$). Soit X une variable aléatoire réelle L^2.

On cherche le $h(Y)$ le plus proche de X dans L^2. C'est-à-dire on cherche h tel que $\mathbb{E}[|h(Y) - X|^2]$ soit minimale (si $Y : \Omega \longrightarrow (E, \mathcal{E})$, on a $h : E \longrightarrow \mathbb{R}$). On a :

$$\mathbb{E}[|h(Y) - X|^2] = \sum_{i\in I} \mathbb{E}[\mathbb{1}_{Y=y_i}(h(y_i) - X)^2] = \sum_{i\in I} h(y_i)^2 \mathbb{P}(Y = y_i) - 2h(y_i)\mathbb{E}[X\mathbb{1}_{Y=y_i}] + \mathbb{E}[X^2 \mathbb{1}_{Y=y_i}]$$

Choisissons chaque $h(y_i)$ en minimisant le polynôme de degré 2. On prend $h(y_i) = \frac{\mathbb{E}[X\mathbb{1}_{Y=y_i}]}{\mathbb{P}(Y=y_i)}$.

REMARQUE 8.4. Cela revient en fait à conditionner par rapport à un événement B tel que $\mathbb{P}(B) > 0$.

- $A \longmapsto \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$ est une mesure de probabilité.
- On définit $\mathbb{E}[. \mid B]$ l'espérance sous $\mathbb{P}(. \mid B)$. On a $\boxed{\mathbb{E}[X \mid B] = \mathbb{E}[X\mathbb{1}_B]/\mathbb{P}(B)}$.
  En effet, on vérifie que c'est vrai pour les fonctions indicatrices, puis par linéarité pour les fonctions étagées, et enfin par limite croissante pour toute fonction.

On en déduit qu'une fonction h telle que $\mathbb{E}[|h(Y) - X|^2]$ soit minimale est :

$$h : y \longmapsto \begin{cases} \mathbb{E}[X \mid Y = y_i] & \text{si } y = y_i \\ 0 & \text{sinon} \end{cases}$$

On verra que dans ce cas $\mathbb{E}[X \mid Y] = h(Y)$.
Ainsi, $\mathbb{E}[X \mid Y]$ fait une moyenne de X sur l'ensemble où Y a une valeur fixée.

REMARQUE 8.5. $h(Y)$ ainsi défini a un sens dès que $X \in L^1$. On va donc vouloir définir $\mathbb{E}[X \mid Y]$ pour toute variable aléatoire $X \in L^1$.

II. Définition / construction

PROPOSITION 8.6. [PROPRIÉTÉ CARACTÉRISTIQUE DE L'ESPÉRANCE CONDITIONNELLE]
Soient $\mathcal{B}$ une tribu et X une variable aléatoire réelle L^1. Alors il existe une unique (à un ensemble négligeable près) variable aléatoire réelle Z telle que Z est $\mathcal{B}$-mesurable et pour toute variable aléatoire W réelle $\mathcal{B}$-mesurable bornée, on a $\boxed{\mathbb{E}[XW] = \mathbb{E}[ZW]}$.

Z est alors appelée espérance conditionnelle de X sachant $\mathcal{B}$ *et on la note* $\mathbb{E}[X \mid \mathcal{B}]$.

Pour une variable aléatoire Y, on note alors $\mathbb{E}[X \mid Y] = \mathbb{E}[X \mid \sigma(Y)]$.

L'intuition derrière cette définition est qu'à côté d'une variable aléatoire $\mathcal{B}$-mesurable W, mettre X ou Z donne le même résultat. Donc Z est proche de X parmi les $\mathcal{B}$-mesurables.

PREUVE Commençons par montrer l'unicité. Soient Z_1 et Z_2 vérifiant la propriété caractéristique.
Z_1 et Z_2 sont $\mathcal{B}$-mesurables donc $W = \mathbb{1}_{Z_1>Z_2}$ aussi et alors $\mathbb{E}[Z_1W] = \mathbb{E}[XW] = \mathbb{E}[Z_2W]$.
D'où $\mathbb{E}[(Z_1 - Z_2)\mathbb{1}_{Z_1>Z_2}] = 0$.
Comme la variable aléatoire sous l'espérance est positive, on a $(Z_1 - Z_2)\mathbb{1}_{Z_1>Z_2} = 0$ p.s..
Donc $Z_1 \leq Z_2$ p.s.. Par symétrie, on obtient que $Z_1 = Z_2$ p.s.. D'où l'unicité.

Montrons l'existence. On va procéder en plusieurs étapes :

- Si $X \in L^2$.
  Rappelons que $L^2(\mathcal{A}) = \{X \mid \mathbb{E}[|X|^2] < +\infty\}/\sim$ (où $\sim$ est la relation d'équivalence "égalité presque sure"), muni de $\langle X \mid Y\rangle = \mathbb{E}[XY]$, est un espace de HILBERT.
  Et comme $\mathcal{B} \subseteq \mathcal{A}$, $L^2(\mathcal{B})$ est un sous-espace fermé de $L^2(\mathcal{A})$.
  Notons alors π la projection orthogonale sur $L^2(\mathcal{B})$ et posons $Z = \pi X$. Vérifions que Z satisfait la propriété caractéristique :

- $Z \in L^2(\mathcal{B})$ donc est $\mathcal{B}$-mesurable.
- Si W est $\mathcal{B}$-mesurable bornée, $W \in L^2(\mathcal{B})$ et $Z - X = \pi X - X \in L^2(\mathcal{B})^\perp$, donc $\langle Z - X \mid W \rangle = 0$. Ainsi $\langle Z \mid W \rangle = \langle X \mid W \rangle$, ou encore $\mathbb{E}[ZW] = \mathbb{E}[XW]$.

REMARQUE 8.7. Donc sur L^2, on a $\mathbb{E}[. \mid \mathcal{B}] = \pi_{L^2(\mathcal{B})}$. On a alors les propriétés suivantes :

- $\mathbb{E}[. \mid \mathcal{B}]$ est linéaire.
- Si $X \in L^2$ est positive, alors $\mathbb{E}[X \mid \mathcal{B}] \geq 0$ p.s..
  En effet $\mathbb{E}[X \mid \mathcal{B}]$ est $\mathcal{B}$-mesurable, donc $W = \mathbb{1}_{\mathbb{E}[X \mid \mathcal{B}]<0}$ est $\mathcal{B}$-mesurable bornée, la propriété caractéristique s'applique et on a $\mathbb{E}[\mathbb{1}_{\mathbb{E}[X \mid \mathcal{B}]<0}\mathbb{E}[X \mid \mathcal{B}]] = \mathbb{E}[\mathbb{1}_{\mathbb{E}[X \mid \mathcal{B}]<0}X] \geq 0$
  On en déduit $\mathbb{1}_{\mathbb{E}[X \mid \mathcal{B}]<0}\mathbb{E}[X \mid \mathcal{B}] = 0$ p.s. et donc $\mathbb{E}[X \mid \mathcal{B}] \geq 0$ p.s..
- Si $X, Y \in L^2$, on a par linéarité et positivité : $\boxed{X \leq Y \text{ p.s.} \implies \mathbb{E}[X \mid \mathcal{B}] \leq \mathbb{E}[Y \mid \mathcal{B}] \text{ p.s.}}$.

- Si $X \geq 0$ p.s. : pour $n \in \mathbb{N}$, posons $X_n = \min(X, n) \in L^2$.
  Par le point précédent, les $(\mathbb{E}[X_n \mid \mathcal{B}])_{n\in\mathbb{N}}$ existent et la suite est même croissante puisque $(X_n)_{n\in\mathbb{N}}$ l'est. Soit alors $Z = \lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \mid \mathcal{B}]$. Z est $\mathcal{B}$-mesurable.
  Si $W = \mathbb{1}_B$ avec $B \in \mathcal{B}$, on a :

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}\Big[\lim_{n\to+\infty} \uparrow \min(X, n)W\Big] \\
&= \lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \mid W] && \text{par convergence monotone} \\
&= \lim_{n\to+\infty} \uparrow \mathbb{E}[\mathbb{E}[X_n \mid \mathcal{B}]W] && \text{par la propriété caractéristique} \\
&= \mathbb{E}\Big[\lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \mid \mathcal{B}]W\Big] && \text{par convergence monotone} \\
&= \mathbb{E}[ZW]
\end{aligned}$$

  On a donc construit Z $\mathcal{B}$-mesurable tel que pour tout W $\mathcal{B}$-mesurable positive, on a $\mathbb{E}[XW] = \mathbb{E}[ZW]$. C'est ce que l'on prend pour définition de $\mathbb{E}[X \mid \mathcal{B}]$ lorsque $X \geq 0$ p.s..
- Enfin, si $X \in L^1$. On décompose X sous la forme $X = X^+ - X^-$. On peut construire par le point précédent $Z^+ = \mathbb{E}[X^+ \mid \mathcal{B}]$ et $Z^- = \mathbb{E}[X^- \mid \mathcal{B}]$.
  Posons $Z = Z^+ - Z^-$ qui est $\mathcal{B}$-mesurable, et vérifions que Z convient.
  On remarque que $\mathbb{E}[0 \mid \mathcal{B}] = 0$ puisque 0 est $\mathcal{B}$-mesurable et pour tout W, $\mathbb{E}[0W] = \mathbb{E}[0W] = 0$ et que pour $X \geq 0 \in L^1$, on a $X^- = 0$ p.s. et donc $Z = Z^+ - 0 = \mathbb{E}[X \mid \mathcal{B}]$: la définition coïncide avec celle de la partie précédente.
  De plus, dans la construction de $\mathbb{E}[. \mid \mathcal{B}]$ pour les variables aléatoires positives, on avait $\forall n \in \mathbb{N}, \mathbb{E}[X_n \mid \mathcal{B}] \geq \mathbb{E}[0 \mid \mathcal{B}] = 0$ p.s., donc $\mathbb{E}[X \mid \mathcal{B}] \geq 0$ p.s.. Ainsi $Z^+, Z^- \geq 0$ p.s..
  La propriété caractéristique avec $W = 1$ donne alors $\mathbb{E}[Z^+] = \mathbb{E}[X^+] < +\infty$ et $\mathbb{E}[Z^-] = \mathbb{E}[X^-] < +\infty$. Donc $Z \in L^1$.
  De plus si W est $\mathcal{B}$-mesurable bornée :

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}[(X^+ - X^-)(W^+ - W^-)] \\
&= \mathbb{E}[X^+W^+] - \mathbb{E}[X^+W^-] - \mathbb{E}[X^-W^+] + \mathbb{E}[X^-W^-] \\
&= \mathbb{E}[Z^+W^+] - \mathbb{E}[Z^+W^-] - \mathbb{E}[Z^-W^+] + \mathbb{E}[Z^-W^-] \\
&= \mathbb{E}[(Z^+ - Z^-)(W^+ - W^-)] \\
&= \mathbb{E}[ZW]
\end{aligned}$$

□

REMARQUE 8.8. On avait défini $\mathbb{E}[.]$ pour des variables aléatoires positives (espérance à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$) ainsi que les variables aléatoires L^1 (espérance à valeurs dans $\mathbb{R}$).
De même ici, on a défini $\mathbb{E}[X \mid \mathcal{B}]$ pour les variables aléatoires positives (l'espérance conditionnelle est une variable aléatoire à valeurs positives) ainsi que les variables aléatoires L^1 (l'espérance conditionnelle est une variable aléatoire L^1).
Là encore, la notion la plus importante est celle sur les variables aléatoires L^1.

REMARQUE 8.9.

- Pour $X \in L^1$, on a la formulation équivalente de la proposition caractéristique :

$$\forall W\ \mathcal{B}\text{-mesurable bornée}, \mathbb{E}[XW] = \mathbb{E}[ZW] \iff \boxed{\forall B \in \mathcal{B}, \mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[Z\mathbb{1}_B]}$$

- Dans le cas de l'espérance conditionnelle par rapport à une variable aléatoire Y, la proposition caractéristique est équivalente à :

$$\forall W\ Y\text{-mes. bornée}, \mathbb{E}[XW] = \mathbb{E}[ZW] \iff \boxed{\forall \phi \text{ mesurable bornée}, \mathbb{E}[X\phi(Y)] = \mathbb{E}[Z\phi(Y)]}$$

EXEMPLE 8.10. Revenons au cas de la section précédente où Y est discrète (on rappelle $Y \in (y_i)_{i\in I}$ avec I fini ou dénombrable et $\mathbb{P}(Y = y_i) > 0$ pour tout $i \in I$).
Soit $X \in L^1$. Pour tout ϕ mesurable bornée, on a :

$$\begin{aligned}
\mathbb{E}[X\phi(Y)] &= \mathbb{E}\Big[X \sum_{i\in I} \phi(y_i)\mathbb{1}_{Y=y_i}\Big] \\
&= \sum_{i\in I} \phi(y_i)\mathbb{E}[X\mathbb{1}_{Y=y_i}] && \text{par le théorème de FUBINI car } \phi \text{ est bornée} \\
&= \sum_{i\in I} \phi(y_i)h(y_i)\mathbb{P}(Y = y_i) && \text{où } h(y) = \mathbb{E}[X \mid Y = y] \\
&= \mathbb{E}[h(Y)\phi(Y)]
\end{aligned}$$

Ceci étant vrai pour toute ϕ mesurable bornée, on a par la remarque précédente : $\boxed{\mathbb{E}[X \mid Y] = h(Y)}$.

EXEMPLE 8.11. Regardons le cas particulier où $X = \psi(U)$ où $U \sim \mathcal{U}([0,1])$ et ψ est mesurable.
On a $\mathbb{E}[|X|] = \mathbb{E}[|\psi(U)|] = \int_0^1 |\psi(u)|\ du$. Donc si $\psi \in L^1([0,1])$, alors $X \in L^1$.
Soit $k \in \mathbb{N}^*$. On note $\mathcal{B}_k = \sigma\left(\left\{U \in \left[\frac{p}{k}, \frac{p+1}{k}\right[\right\}_{p\in\mathbb{Z}}\right)$. Déterminons $\mathbb{E}[X \mid \mathcal{B}_k]$.

Pour tout $p \in \mathbb{Z}$, on a $\left\{U \in \left[\frac{p}{k}, \frac{p+1}{k}\right[\right\} = \{\lfloor kU \rfloor = p\}$ donc en fait $\mathcal{B}_k = \sigma(\lfloor kU \rfloor)$ (on va voir que c'est beaucoup plus agréable à traiter).
En effet, $\lfloor kU \rfloor$ est une variable aléatoire discrète. On peut donc appliquer l'exemple précédent :

$$\mathbb{E}[\psi(U) \mid \mathcal{B}_k] = \mathbb{E}[\psi(U) \mid \lfloor kU \rfloor] = h(\lfloor kU \rfloor)$$

avec, pour $p \in [\![0, k-1]\!]$:

$$h(p) = \mathbb{E}[\psi(U) \mid \lfloor kU \rfloor = p] = \frac{\mathbb{E}[\psi(U)\mathbb{1}_{\lfloor kU \rfloor = p}]}{\mathbb{P}(\lfloor kU \rfloor = p)} = \frac{\int_0^1 \psi(u)\mathbb{1}_{\frac{p}{k} \le u < \frac{p+1}{k}}\ du}{\mathbb{P}(\frac{p}{k} \le u < \frac{p+1}{k})} = \int_{\frac{p}{k}}^{\frac{p+1}{k}} \psi(u)\frac{du}{1/k}$$

$h(p)$ est donc la valeur moyenne de ψ sur $\left[\frac{p}{k}, \frac{p+1}{k}\right[$. Alors :

$$\boxed{\mathbb{E}[\psi(U) \mid \mathcal{B}_k] = \int_{\frac{\lfloor kU \rfloor}{k}}^{\frac{\lfloor kU \rfloor + 1}{k}} \psi(u) \frac{du}{1/k}}$$

$\mathbb{E}[\psi(U) \mid \mathcal{B}_k]$ est donc la valeur moyenne de ψ sur $\left[\frac{\lfloor kU \rfloor}{k}, \frac{\lfloor kU \rfloor + 1}{k}\right[$.

L'effet de $\mathbb{E}[. \mid \mathcal{B}_k]$ est donc une perte d'information :

- avec $X = \psi(U)$, on a toute l'information sur X,
- avec $\mathbb{E}[X \mid \mathcal{B}_k]$, on remplace la valeur de X par sa moyenne sur un ensemble,
- avec $\mathbb{E}[X]$, on n'a que la moyenne globale, on a perdu toute l'information sur X.

Appliquons ce résultat à $\psi : x \longmapsto \frac{1}{x}$. $\mathbb{E}[\frac{1}{U} \mid \mathcal{B}_k]$ existe puisque $\frac{1}{U} \geq 0$ p.s. et on a de même :

$$\mathbb{E}\Big[\frac{1}{U} \mid \mathcal{B}_k\Big] = \int_{\frac{\lfloor kU \rfloor}{k}}^{\frac{\lfloor kU \rfloor + 1}{k}} \frac{1}{u} \frac{du}{1/k} = \begin{cases} +\infty & \text{si } U < \frac{1}{k} \\ \int_{\frac{1}{k}}^{\frac{2}{k}} \frac{1}{u} \frac{du}{1/k} & \text{si } \frac{1}{k} \leq U \leq \frac{2}{k} \end{cases} \quad \text{p.s.}$$

On a donc $0 \leq \frac{1}{U} < +\infty$ p.s. mais $\mathbb{P}\Big(\mathbb{E}\Big[\frac{1}{U} \mid \mathcal{B}_k\Big] = \infty\Big) > 0$.

PROPOSITION 8.12.

(i) $\mathbb{E}[.\,|\,\mathcal{B}] : L^1 \longrightarrow L^1$ *est linéaire : pour toutes variables aléatoires* $X, Y \in L^1$ *et* $\lambda \in \mathbb{R}$ *:*

$$\mathbb{E}[X+Y\,|\,\mathcal{B}] = \mathbb{E}[X\,|\,\mathcal{B}] + \mathbb{E}[Y\,|\,\mathcal{B}] \quad \text{et} \quad \mathbb{E}[\lambda X\,|\,\mathcal{B}] = \lambda\mathbb{E}[X\,|\,\mathcal{B}]$$

(ii) Si X *est* L^1 *et* $\mathcal{B}$*-mesurable :*

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

En particulier $\mathbb{E}[c\,|\,\mathcal{B}] = c$ *et* $\mathbb{E}[\phi(Y)\,|\,Y] = \phi(Y)$*.*

(iii) Si X *est* L^1 *et est indépendante de* $\mathcal{B}$ *(c'est-à-dire* $\sigma(X)$ *et* $\mathcal{B}$ *sont indépendantes), alors :*

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

En particulier X *et* Y *sont indépendantes, alors* $\mathbb{E}[X\,|\,Y] = \mathbb{E}[X]$*.*

(iv) Si $\mathcal{B} \subseteq \mathcal{C}$ *sont 2 sous-tribus de* $\mathcal{A}$*, on a pour tout* $X \in L^1$ *:*

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

(v) **[INÉGALITÉ DE JENSEN]**
Si X *et* $\phi(X)$ *sont* L^1 *avec* ϕ *convexe, alors :*

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

En particulier $|\mathbb{E}[X\,|\,\mathcal{B}]| \leq \mathbb{E}[|X|\;|\,\mathcal{B}]$ *: on dit que* $\mathbb{E}[.\,|\,\mathcal{B}] : L^1 \longrightarrow L^1$ *est un* opérateur contractant *de norme inférieure ou égale à 1.*

(vi) Si X *et* Y *sont* L^1*, alors :*

$$\boxed{X \leq Y \text{ p.s.} \Longrightarrow \mathbb{E}[X\,|\,\mathcal{B}] \leq \mathbb{E}[Y\,|\,\mathcal{B}] \text{ p.s.}}$$

(vii) Si X *est* L^1*,* XY *est* L^1 *et* Y *est* $\mathcal{B}$*-mesurable, alors :*

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

REMARQUE 8.13.

1. Le point (ii) implique que $\mathbb{E}[X\,|\,\mathcal{A}] = X$.
   Le point (iii) implique que $\mathbb{E}[X\,|\,\{\varnothing, \Omega\}] = \mathbb{E}[X]$.
   $\mathbb{E}[X\,|\,\mathcal{B}]$ est une interpolation entre ces 2 extrêmes : on fait une *moyenne partielle*.
2. Certaines de ces propriétés sont triviales pour les variables aléatoires L^2.
   En effet, on a vu que $\mathbb{E}[.\,|\,\mathcal{B}]_{|L^2} = \pi_{L^2(\mathcal{B})}$. Cela donne directement (i) et (ii) sur L^2, mais aussi (iv) puisque si $\mathcal{B} \subseteq \mathcal{C}$ alors $L^2(\mathcal{B}) \subseteq L^2(\mathcal{C})$ et donc $\pi_{L^2(\mathcal{B})} \circ \pi_{L^2(\mathcal{C})} = \pi_{L^2(\mathcal{C})} \circ \pi_{L^2(\mathcal{B})} = \pi_{L^2(\mathcal{B})}$.

PREUVE

(i) Pour tout W $\mathcal{B}$-mesurable borné, on a :

$$\begin{aligned}\mathbb{E}[(\lambda X+Y)W] &= \lambda\mathbb{E}[XW]+\mathbb{E}[YW] = \lambda\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{B}]W]+\mathbb{E}[\mathbb{E}[Y\,|\,\mathcal{B}]W]\\ &= \mathbb{E}[(\lambda\mathbb{E}[X\,|\,\mathcal{B}]+\mathbb{E}[Y\,|\,\mathcal{B}])W]\end{aligned}$$

Ainsi $(\lambda\mathbb{E}[X\,|\,\mathcal{B}]+\mathbb{E}[Y\,|\,\mathcal{B}])$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique de l'espérance conditionnelle de $\lambda X+Y$ par rapport à $\mathcal{B}$. donc :

$$\mathbb{E}[\lambda X+Y\,|\,\mathcal{B}] = \lambda\mathbb{E}[X\,|\,\mathcal{B}]+\mathbb{E}[Y\,|\,\mathcal{B}]$$

(ii) Si X est $\mathcal{B}$-mesurable, X satisfait la propriété caractéristique donc $\mathbb{E}[X\,|\,\mathcal{B}] = X$.

(iii) Pour tout W $\mathcal{B}$-mesurable borné, on a :

$$\mathbb{E}[XW] = \mathbb{E}[X]\mathbb{E}[W] = \mathbb{E}[\mathbb{E}[X]W]$$

Ainsi $\mathbb{E}[X]$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique donc $\mathbb{E}[X\,|\,\mathcal{B}] = \mathbb{E}[X]$.

(iv) Soient $\mathcal{B}\subseteq\mathcal{C}$. D'après (ii) : $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{B}]\,|\,\mathcal{C}] = \mathbb{E}[X\,|\,\mathcal{B}]$ p.s..
Puis pour tout W $\mathcal{B}$-mesurable borné, on a :

$$\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{C}]W] \underset{\substack{\text{prop. caract.}\\ \text{de } \mathbb{E}[.\,|\,\mathcal{C}]}}{=} \mathbb{E}[XW] \underset{\substack{\text{prop. caract.}\\ \text{de } \mathbb{E}[.\,|\,\mathcal{B}]}}{=} \mathbb{E}[\mathbb{E}[X\,|\,\mathcal{B}]W]$$

Ainsi $\mathbb{E}[X\,|\,\mathcal{B}]$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique, donc $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{C}]\,|\,\mathcal{B}] = \mathbb{E}[X\,|\,\mathcal{B}]$.

(v) La preuve est similaire à la preuve de ce résultat dans le cas de l'espérance.
On a vu que $\phi : x\longmapsto \sup_{a,b\in\mathbb{R}}(\{ax+b\,|\,\forall y, ay+b\le\phi(y)\})$. D'où :

$$\begin{aligned}\phi(\mathbb{E}[X\,|\,\mathcal{B}]) &= \sup_{a,b\in\mathbb{R}}\{a\mathbb{E}[X\,|\,\mathcal{B}]+b\,|\,\forall y, ay+b\le\phi(y)\}\\ &= \sup_{a,b\in\mathbb{R}}\{\mathbb{E}[aX+b\,|\,\mathcal{B}]\,|\,\forall y, ay+b\le\phi(y)\}\\ &\le \sup_{a,b\in\mathbb{R}}\{\mathbb{E}[\phi(X)\,|\,\mathcal{B}]\,|\,\forall y, ay+b\le\phi(y)\} \qquad \text{par le point (vi)}\\ &= \mathbb{E}[\phi(X)\,|\,\mathcal{B}]\end{aligned}$$

(vi) La preuve est identique au cas L^2. On pose $W = \mathbb{1}_{\mathbb{E}[Y\,|\,\mathcal{B}]<\mathbb{E}[X\,|\,\mathcal{B}]}$. Alors

$$\mathbb{E}[\underbrace{(X-Y)W}_{\le 0 \text{ p.s.}}] = \mathbb{E}[XW]-\mathbb{E}[YW] = \mathbb{E}[\mathbb{E}[X\,|\,\mathcal{B}]W]-\mathbb{E}[\mathbb{E}[Y\,|\,\mathcal{B}]W] = \mathbb{E}[\underbrace{(\mathbb{E}[X\,|\,\mathcal{B}]-\mathbb{E}[Y\,|\,\mathcal{B}])W}_{\ge 0 \text{ p.s.}}]$$

Donc la variable aléatoire dans la deuxième espérance est nulle p.s., donc $\mathbb{E}[X\,|\,\mathcal{B}]\le\mathbb{E}[Y\,|\,\mathcal{B}]$ p.s..

(vii) – Si $X\ge 0$ p.s. et $Y\ge 0$ p.s.. Pour tout W $\mathcal{B}$-mesurable positif, on a :

$$\mathbb{E}[(XY)W] = \mathbb{E}[X\underbrace{(YW)}_{\mathcal{B}\text{-mes pos}}] = \mathbb{E}[(\mathbb{E}[X\,|\,\mathcal{B}]Y)W]$$

Ainsi $\mathbb{E}[X\,|\,\mathcal{B}]Y$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique (pour une variable aléatoire positive). Donc $\mathbb{E}[XY\,|\,\mathcal{B}] = \mathbb{E}[X\,|\,\mathcal{B}]Y\in L^1$.

– Dans le cas général : on décompose $X = X^+ - X^-$ et $Y = Y^+ - Y^-$ et on développe les espérances, ce qui donne aisément le résultat.

□

EXEMPLE 8.14. [CAS À DENSITÉ]
Soient (X, Y) des variables aléatoires à valeurs dans $\mathbb{R}$ telles que $d\mu_{(X,Y)}(x, y) = \rho(x, y)\, dx\, dy$. Soit $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}[|\psi(X, Y)|] < +\infty$. Que vaut $h(Y) = \mathbb{E}[\psi(X, Y) \mid Y]$?

On rappelle que h est caractérisé par :

$$\forall \phi \text{ mesurable bornée }, \mathbb{E}[\psi(X, Y)\phi(Y)] = \mathbb{E}[h(Y)\phi(Y)]$$

Or,

$$\begin{aligned}
\mathbb{E}[\psi(X, Y)\phi(Y)] &= \int_{\mathbb{R}^2} \psi(x, y)\phi(y)\rho(x, y)\, dx\, dy = \int_{\mathbb{R}} \phi(y) \left(\int_{\mathbb{R}} \psi(x, y)\rho(x, y)\, dx \right) dy \\
&\underset{(\star)}{=} \int_{\mathbb{R}} \phi(y) \underbrace{\frac{\int_{\mathbb{R}} \psi(x, y)\rho(x, y)\, dx}{\int_{\mathbb{R}} \rho(x, y)\, dx} \mathbb{1}_{\int_{\mathbb{R}} \rho(x,y)\, dx \neq 0}}_{h(y)} \underbrace{\left(\int_{\mathbb{R}} \rho(x, y)\, dx \right) dy}_{d\mu_Y(y)} \\
&= \mathbb{E}\left[\frac{\int_{\mathbb{R}} \psi(x, Y)\rho(x, Y)\, dx}{\int_{\mathbb{R}} \rho(x, Y)\, dx} \mathbb{1}_{\int_{\mathbb{R}} \rho(x,Y)\, dx \neq 0} \phi(Y) \right]
\end{aligned}$$

Pour justifier l'indicatrice dans $(\star)$, on remarque que si $\int_{\mathbb{R}} \rho(x, y)\, dx = 0$, alors $\int_{\mathbb{R}} \psi(x, y)\rho(x, y)\, dx = 0$.

Ainsi $\boxed{\mathbb{E}[\psi(X, Y) \mid Y] = \frac{\int_{\mathbb{R}} \psi(x, Y)\rho(x, Y)\, dx}{\int_{\mathbb{R}} \rho(x, Y)\, dx} \mathbb{1}_{\int_{\mathbb{R}} \rho(x,Y)\, dx \neq 0}}$ où $\mathbb{E}[|\psi(X, Y)|] < +\infty$.

EXEMPLE 8.15. Soit $\mathcal{B}$ une tribu. Soient X indépendante de $\mathcal{B}$ et Y $\mathcal{B}$-mesurable. Que vaut $\mathbb{E}[\psi(X, Y) \mid \mathcal{B}]$?

Soit W $\mathcal{B}$-mesurable borné. X est indépendante de (Y, W), d'où :

$$\begin{aligned}
\mathbb{E}[\psi(X, Y)W] &= \int \psi(x, y)w\, d\mu_{(X,Y,W)}(x, y, w) && \text{par le lemme de transfert} \\
&= \int \psi(x, y)w\, d\mu_X(x)\, d\mu_{(Y,W)}(y, w) && \text{par indépendance} \\
&= \int \left(\int \psi(x, y)\, d\mu_X(x) \right) w\, d\mu_{(Y,W)}(y, w) \\
&= \mathbb{E}\left[\left(\int \psi(x, Y)\, d\mu_X(x) \right) W \right]
\end{aligned}$$

Ainsi $\int \psi(x, Y)\, d\mu_X(x)$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique. Donc :

$$\boxed{\mathbb{E}[\psi(X, Y) \mid \mathcal{B}] = \int \psi(x, Y)\, d\mu_X(x) = h(Y)}$$

où $h(y) = \mathbb{E}[\psi(X, y)]$.

On a les mêmes résultats de convergence que pour l'espérance :

THÉORÈME 8.16.

- **[LEMME DE FATOU]**
  Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires positives presque sûrement. Alors :

$$\boxed{\mathbb{E}\Big[\liminf_{n\to+\infty} X_n \,|\, \mathcal{B}\Big] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n \,|\, \mathcal{B}]}$$

- **[THÉORÈME DE CONVERGENCE MONOTONE]**
  Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires positives presque sûrement telles que $\lim_{n\to+\infty} \uparrow X_n = X$ p.s.. Alors :

$$\boxed{\lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \,|\, \mathcal{B}] = \mathbb{E}[X \,|\, \mathcal{B}]}$$

- **[THÉORÈME DE CONVERGENCE DOMINÉE]**
  Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telles que $\forall n \in \mathbb{N}, |X_n| \leq Z$ p.s. avec $Z \in L^1$ et $X_n \xrightarrow[n\to+\infty]{p.s.} X$. Alors :

$$\boxed{\mathbb{E}[X_n \,|\, \mathcal{B}] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X \,|\, \mathcal{B}]}$$

PREUVE

- On a :

$$\begin{aligned}
\mathbb{E}\Big[\liminf_{n\to+\infty} X_n \,|\, \mathcal{B}\Big] &= \mathbb{E}\Big[\lim_{n\to+\infty} \uparrow \inf_{k\geq n} X_k \,|\, \mathcal{B}\Big] \\
&= \lim_{n\to+\infty} \mathbb{E}\Big[\underbrace{\inf_{k\geq n} X_k}_{\leq X_k \forall k\geq n} \,|\, \mathcal{B}\Big] && \text{par le TCM conditionnel} \\
&\leq \lim_{n\to+\infty} \inf_{k\geq n} \mathbb{E}[X_k \,|\, \mathcal{B}] \\
&= \liminf_{n\to+\infty} \mathbb{E}[X_n \,|\, \mathcal{B}]
\end{aligned}$$

- Pour tout W $\mathcal{B}$-mesurable bornée, on a :

$$\begin{aligned}
\mathbb{E}[XW] &= \mathbb{E}\Big[\lim_{n\to+\infty} \uparrow X_n W\Big] \\
&= \lim_{n\to+\infty} \uparrow \mathbb{E}[X_n W] && \text{par le TCM classique} \\
&= \lim_{n\to+\infty} \uparrow \mathbb{E}[\mathbb{E}[X_n \,|\, \mathcal{B}]W] && \text{par la propriété caractéristique} \\
&= \mathbb{E}\Big[\lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \,|\, \mathcal{B}]W\Big] && \text{par le TCM classique}
\end{aligned}$$

Ainsi $\lim_{n\to+\infty} \uparrow \mathbb{E}[X_n \,|\, \mathcal{B}]$ est $\mathcal{B}$-mesurable et satisfait la propriété caractéristique, d'où le résultat.

– On a $\forall n \in \mathbb{N}, -Z \leq X_n \leq Z$ p.s.. Donc $\forall n \in \mathbb{N}, Z + X_n \geq 0$ p.s., d'où :

$$\begin{aligned}
\mathbb{E}[Z \mid \mathcal{B}] + \mathbb{E}\Big[\liminf_{n\to+\infty} X_n \mid \mathcal{B}\Big] &= \mathbb{E}\Big[\Big(Z + \liminf_{n\to+\infty} X_n\Big) \mid \mathcal{B}\Big] \\
&= \mathbb{E}\Big[\liminf_{n\to+\infty}(Z + X_n) \mid \mathcal{B}\Big] \\
&\leq \liminf_{n\to+\infty} \mathbb{E}[Z + X_n \mid \mathcal{B}] \qquad \text{par le lemme de FATOU conditionnel} \\
&= \mathbb{E}[Z \mid \mathcal{B}] + \liminf_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{B}]
\end{aligned}$$

Donc $\mathbb{E}[X \mid \mathcal{B}] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{B}]$.
En appliquant le même raisonnement avec $(-X_n)_{n\in\mathbb{N}}$, on obtient $\mathbb{E}[-X \mid \mathcal{B}] \leq \liminf_{n\to+\infty} \mathbb{E}[-X_n \mid \mathcal{B}]$.
Ou encore $\limsup_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{B}] \leq \mathbb{E}[X \mid \mathcal{B}]$.
D'où finalement $\mathbb{E}[X_n \mid \mathcal{B}] \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}[X \mid \mathcal{B}]$.
Par le théorème de convergence dominée classique, on a $\mathbb{E}[|\mathbb{E}[X_n \mid \mathcal{B}] - \mathbb{E}[X \mid \mathcal{B}]|] \xrightarrow[n\to+\infty]{} 0$,
la domination provenant du fait que pour tout entier n :

$$|\mathbb{E}[X_n \mid \mathcal{B}] - \mathbb{E}[X \mid \mathcal{B}]| \leq |\mathbb{E}[X_n \mid \mathcal{B}]| + |\mathbb{E}[X \mid \mathcal{B}]| \underset{\text{JENSEN}}{\leq} \mathbb{E}[|X_n \mid \mathcal{B}]| + \mathbb{E}[|X \mid \mathcal{B}|] \leq 2\mathbb{E}[Z \mid \mathcal{B}] \in L^1$$

□

REMARQUE 8.17. On a vu plein de propriétés de $\mathbb{E}[. \mid \mathcal{B}]$ pour des variables aléatoires L^1. Elles sont également vraies pour des variables aléatoires positives. On étend effectivement facilement ces résultats grâce au théorème de convergence monotone.

III. Liens entre indépendance et orthogonalité

On note parfois l'indépendance $\perp$.

On rappelle que L^2 est muni du produit scalaire $\langle X \mid Y \rangle = \mathbb{E}[XY]$.
Si X et Y sont indépendants et L^2, on a $\langle X \mid Y \rangle = \mathbb{E}[X]\mathbb{E}[Y] \neq 0$ en général.
Cela a alors un peu plus de sens sur $A = \{X \in L^2 \mid \mathbb{E}[X] = 0\}$. En effet, pour $X, Y \in A$, on a $\langle X \mid Y \rangle = \mathrm{Cov}(X, Y)$, donc $\langle X \mid Y \rangle = 0$.
La réciproque est fausse en général. (on a vu qu'elle était vraie lorsque $(X, Y)^\top$ est un vecteur gaussien).

On a aussi vu que si X et Y sont indépendants, alors $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ p.s., d'où $\mathbb{E}[(X - \mathbb{E}[X]) \mid Y] = 0$ p.s. et donc $\pi_{L^2(\sigma(Y))} X - \mathbb{E}[X] = 0$ p.s.. Ainsi $X - \mathbb{E}[X] \in L^2(\sigma(Y))$.

PROPOSITION 8.18. *Soient $\mathcal{B}$ et $\mathcal{C}$ deux sous-tribus de $\mathcal{A}$. Les propositions suivantes sont équivalentes :*

(i) *$\mathcal{B}$ et $\mathcal{C}$ sont indépendantes,*

(ii) $\{X \in L^2(\mathcal{B}) \mid \mathbb{E}[X] = 0\} \perp \{Y \in L^2(\mathcal{C}) \mid \mathbb{E}[Y] = 0\}$,

(iii) $\forall X \in L^1(\mathcal{B}), \mathbb{E}[X \mid \mathcal{C}] = \mathbb{E}[X]$.

PREUVE

(i) $\Longrightarrow$ (iii) C'est une propriété de $\mathbb{E}[.\,|\,\mathcal{C}]$.

(ii) $\Longrightarrow$ (i) Soient $B \in \mathcal{B}$ et $C \in \mathcal{C}$.
Considèrons $X = \mathbb{1}_B - \mathbb{P}(B) \in L^2(\mathcal{B})$, qui vérifie $\mathbb{E}[X] = 0$, et $Y = \mathbb{1}_C - \mathbb{P}(C) \in L^2(\mathcal{C})$, qui vérifie $\mathbb{E}[Y] = 0$. Alors :

$$0 = \langle X \mid Y \rangle = \mathrm{Cov}(X, Y) = \mathrm{Cov}(\mathbb{1}_B, \mathbb{1}_C) = \mathbb{E}[\mathbb{1}_B \mathbb{1}_C] - \mathbb{E}[\mathbb{1}_B]\mathbb{E}[\mathbb{1}_C] = \mathbb{P}(B \cap C) - \mathbb{P}(B)\mathbb{P}(C)$$

Donc $B \perp C$. On en déduit que $\mathcal{B} \perp \mathcal{C}$.

(iii) $\Longrightarrow$ (ii) Soient $X \in L^2(\mathcal{B})$ et $Y \in L^2(\mathcal{C})$ telles que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.
Si Y est bornée, on a :

$$\langle X \mid Y \rangle = \mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{C}]Y] = \mathbb{E}[\mathbb{E}[X]Y] = 0$$

Si Y n'est pas bornée, on considère les variables aléatoires Y_n qui tronquent Y à $[-n, n]$. Il vient alors par convergence dominée :

$$\langle X \mid Y \rangle = \mathbb{E}[XY] = \lim_{n \to +\infty} \mathbb{E}[\underbrace{XY_n}_{|.| \leq |XY|}] = \lim_{n \to +\infty} 0 = 0$$

□

THÉORÈME 8.19. *Soit $Z = (X, Y_1, \ldots, Y_n)$ un vecteur gaussien. Alors :*

$$\boxed{\mathbb{E}[X \,|\, (Y_1, \ldots, Y_n)] = \pi^{\perp}_{\mathrm{Vect}(1, Y_1, \ldots, Y_n)} X}$$

En particulier : $\exists \lambda_0, \ldots, \lambda_n \in \mathbb{R} \,|\, \mathbb{E}[X \,|\, (Y_1, \ldots, Y_n)] = \lambda_0 + \sum_{i=1}^n \lambda_i Y_i$

REMARQUE 8.20.

- Ainsi $\mathbb{E}[X \,|\, Y_1, \ldots, Y_n]$ est une variable aléatoire gaussienne,
- $\mathrm{Vect}(1, Y_1, \ldots, Y_n)$ est beaucoup plus petit que $L^2(Y_1, \ldots, Y_n) = L^2(\sigma(\{Y_1, \ldots, Y_n\}))$. Le théorème permet donc de réduire significativement la recherche de $\mathbb{E}[X \,|\, (Y_1, \ldots, Y_n)]$.

PREUVE Posons $V = \pi^{\perp}_{\mathrm{Vect}(1, Y_1, \ldots, Y_n)} X = \lambda_0 + \sum_{i=1}^n \lambda_i Y_i$.
V est $(Y_1, \ldots, Y_n)$-mesurable. Considérons $\tilde{Z} = (V - X, Y_1, \ldots, Y_n)$.
Une combinaison linéaire de coordonnées de $\tilde{Z}$ est de la forme

$$\mu_0(V - X) + \sum_{i=1}^n \mu_i Y_i = -\mu_0 X + \mu_0 \lambda_0 + \sum_{i=1}^n (\mu_i + \mu_0 \lambda_i) Y_i$$

C'est une combinaison affine des $(Y_i)_{0 \leq i \leq n}$ et de $\{X\}$, donc une variable aléatoire gaussienne puisque Z est un vecteur gaussien. On en déduit que $\tilde{Z}$ est un vecteur gaussien.

On a $V - X \in \mathrm{Vect}(1, Y_1, \ldots, Y_n)^{\perp}$, d'où $\mathbb{E}[V - X] = \langle V - X \mid 1 \rangle = 0$ et pour $1 \leq i \leq n$, on a $\langle V - X \mid Y_i \rangle = 0$.

La matrice de covariance de $\tilde{Z}$ est donc de la forme $\begin{pmatrix} \star & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & \star & \\ 0 & & & \end{pmatrix}$ (symétrique de taille $n+1$).

Comme $\tilde{Z}$ est un vecteur gaussien, on en déduit que $V - X$ est indépendant de $(Y_1, \dots, Y_n)$.
Donc si W est $(Y_1, \dots, Y_n)$-mesurable bornée :

$$\mathbb{E}[VW] = \mathbb{E}[(V-X)W] + \mathbb{E}[XW] \underset{\text{indép.}}{=} \underbrace{\mathbb{E}[V-X]}_{=0}\mathbb{E}[W] + \mathbb{E}[XW]$$

Ainsi $\mathbb{E}[VW] = \mathbb{E}[XW]$ et V satisfait la propriété caractéristique de $\mathbb{E}[. \mid Y_1, \dots, Y_n]$. □

Exemple 8.21. Considèrons $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix}\right)$ et cherchons $\mathbb{E}[X \mid (Y,Z)]$.

Posons $\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix}\right)$. On a :

$$\begin{aligned} \mathbb{E}[X \mid (Y,Z)] &= \mathbb{E}[X_0 + 1 \mid (Y_0, Z_0)] && \text{car } \sigma(Y,Z) = \sigma(Y_0, Z_0) \\ &= 1 + \mathbb{E}[X_0 \mid (Y_0 + Z_0)] \end{aligned}$$

et par le théorème, il existe $\lambda_0, \lambda_1, \lambda_2 \in \mathbb{R}$ tels que $\mathbb{E}[X_0 \mid (Y_0 + Z_0)] = \lambda_0 + \lambda_1 Y_0 + \lambda_2 Z_0$.

Pour trouver les coefficients on regarde $\mathbb{E}[X_0]$, $\mathbb{E}[X_0 Y_0]$ et $\mathbb{E}[X_0 Z_0]$:

$$\begin{cases} \mathbb{E}[X_0] = 0 \\ \mathbb{E}[X_0] = \mathbb{E}[\mathbb{E}[X_0 \mid (Y_0, Z_0)]] = \mathbb{E}[\lambda_0 + \lambda_1 Y_0 + \lambda_2 Z_0] = \lambda_0 \end{cases} \implies \underline{\lambda_0 = 0}$$

$$\begin{cases} \mathbb{E}[X_0 Y_0] = \operatorname{Cov}(X_0, Y_0) = 0 \\ \mathbb{E}[X_0 Y_0] = \mathbb{E}[(\lambda_1 Y_0 + \lambda_2 Z_0) Y_0] = \lambda_1 \operatorname{Var}(Y_0) + \lambda_2 \operatorname{Cov}(Z_0, Y_0) = 5\lambda_1 + 3\lambda_2 \end{cases} \implies \underline{5\lambda_1 + 3\lambda_2 = 0}$$

$$\begin{cases} \mathbb{E}[X_0 Z_0] = \operatorname{Cov}(X_0, Z_0) = -1 \\ \mathbb{E}[X_0 Z_0] = \mathbb{E}[(\lambda_1 Y_0 + \lambda_2 Z_0) Z_0] = \lambda_1 \operatorname{Cov}(Y_0, Z_0) + \lambda_2 \operatorname{Var}(Z_0) = 3\lambda_1 + 4\lambda_2 \end{cases} \implies \underline{3\lambda_1 + 4\lambda_2 = -1}$$

On en déduit $(\lambda_0, \lambda_1, \lambda_2) = (0, 3/11, -5/11)$ et donc :

$$\boxed{\mathbb{E}[X \mid (Y,Z)] = 1 + \frac{3}{11}Y - \frac{5}{11}(Z-2) = \frac{21}{11} + \frac{3}{11}Y - \frac{5}{11}Z}$$

IV. Lois conditionnelles

Les lois conditionnelles sont des objets plus généraux que les espérances conditionnelles mais moins maniables …

Définition 8.22. [Noyau de transition]
On dit que $\begin{array}{rccc} \nu : & \mathbb{R} \times \mathcal{B}(\mathbb{R}) & \longrightarrow & [0,1] \\ & (y, A) & \longmapsto & \nu_y(A) \end{array}$ est un *noyau de transition* si :

- pour tout $A \in \mathcal{B}(\mathbb{R})$, $y \longmapsto \nu_y(A)$ est mesurable,
- pour tout $y \in \mathbb{R}$, $A \longmapsto \nu_y(A)$ est une mesure de probabilité.

$(\nu_y)_{y \in \mathbb{R}}$ est ainsi une famille de probabilités.

DÉFINITION 8.23. [LOI CONDITIONNELLE]
Soient X, Y deux variables aléatoires réelles. On dit que *la loi de X conditionnellement à Y* est le noyau de transition ν_Y, et on note $\mathcal{L}(X \mid Y) = \nu_Y$, si :

$$\forall \phi \text{ mesurable positive}, \mathbb{E}[\phi(X) \mid Y] = \int \phi(x)\, d\nu_Y(x) \text{ p.s.}$$

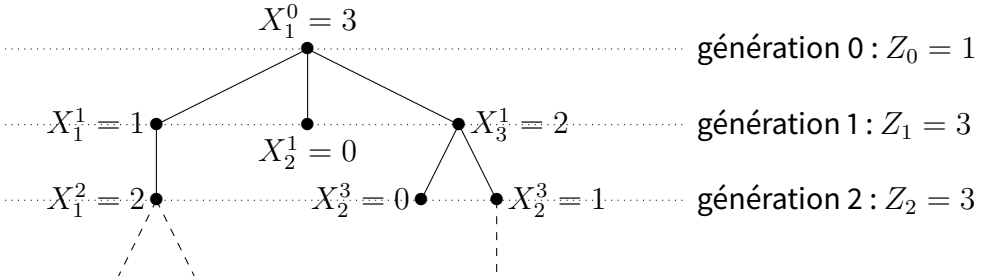
REMARQUE 8.24.
- On admet que le couple (X, Y) étant donné, $\mathcal{L}(X \mid Y)$ existe.
- Cela généralise $\mathbb{E}[X \mid Y]$ lorsque l'on sait que $\mathbb{E}[X^+ \mid Y]$ et $\mathbb{E}[X^- \mid Y]$ sont finies presque sûrement.
  On aura en effet $\boxed{\mathbb{E}[X \mid Y] = \int x\, d\nu_Y(x)}$.

EXEMPLE 8.25. [LOIS CONDITIONNELLES]

- Si $X \perp Y$, on a, pour ϕ mesurable positive :
$$\mathbb{E}[\phi(X) \mid Y] = \mathbb{E}[\phi(X)] = \int \phi(x)\, d\mu_X(x)$$
Donc $\mathcal{L}(X \mid Y)$ existe et $\boxed{\mathcal{L}(X \mid Y) = \mu_X}$ (c'est-à-dire pour tout $y \in \mathbb{R}$, $\nu_y = \mu_X$).
- Si X est Y-mesurable, c'est-à-dire $X = h(Y)$ pour un h mesurable, on a :
$$\mathbb{E}[\phi(X) \mid Y] = \mathbb{E}[\phi(h(Y)) \mid Y] = \phi(h(Y)) = \int \phi(x)\, d\delta_{h(Y)}(x)$$
Donc $\boxed{\mathcal{L}(X \mid Y) = \delta_{h(Y)}}$.
- Soit $(X, Y_1, \ldots, Y_n)$ un vecteur gaussien. Que vaut $\mathcal{L}(X \mid Y_1, \ldots, Y_n)$?
$$\mathbb{E}[\phi(X) \mid (Y_1, \ldots, Y_n)] = \mathbb{E}[\phi(V + \tilde{V}) \mid (Y_1, \ldots, Y_n)]$$
où $V = \mathbb{E}[X \mid (Y_1, \ldots, Y_n)]$ qui est $(Y_1, \ldots, Y_n)$-mesurable et $\tilde{V} = X - V$.
On a vu que $\tilde{V} \sim \mathcal{N}(0, \sigma^2)$ est une gaussienne indépendante de $(Y_1, \ldots, Y_n)$.
Et comme on a $\mathbb{E}[f(X, Y) \mid \mathcal{B}] = \int f(x, Y)\, d\mu_X(x)$ lorsque X est indépendante de $\mathcal{B}$ et Y est $\mathcal{B}$-mesurable, il vient :
$$\mathbb{E}[\phi(X) \mid (Y_1, \ldots, Y_n)] = \int \phi(V + x)\, d\mu_{\mathcal{N}(0,\sigma^2)}(x) = \int \phi(y)\, d\mu_{\mathcal{N}(V,\sigma^2)}(y)$$
Donc $\boxed{\mathcal{L}(X \mid (Y_1, \ldots, Y_n)) = \mathcal{N}(\mathbb{E}[X \mid (Y_1, \ldots, Y_n)], \sigma^2)}$ (σ pouvant être calculé).

V. Introduction au processus de branchement : le processus de GALTON-WATSON

On souhaite modéliser la descendance d'un individu. La motivation initiale était de s'intéresser à la survie des noms de Lord anglais, mais on peut aussi modéliser des bactéries qui se multiplient par division.
On s'intéresse au nombre Z_j d'individus à la génération j. On note X_i^j le nombre d'enfants de l'individu i de la génération $j-1$. On peut représenter la situation par un arbre :



Modélisation : on se donne μ une probabilité sur $\mathbb{N}$ appelée *loi de reproduction* qui va permettre de simuler aléatoirement le nombre d'enfants d'un individu. On se donne ainsi $(X_i^j)_{i,j\geq 1}$ des variables aléatoires indépendantes et identiquement distribuées de loi μ et on définit $(Z_n)_{n\in\mathbb{N}}$ la suite caractérisant le nombre d'individus par :

$$\boxed{\begin{cases} Z_0 = 1 \\ Z_n = \sum_{i=1}^{Z_{n-1}} X_i^n \quad \text{pour } n \in \mathbb{N} \end{cases}}$$

REMARQUE 8.26. Les hypothèses d'indépendance et de distribution identique sont des hypothèses fortes et très simplificatrices.

NOTE 8.27.
- On a $Z_n = \sum_{i\geq 1} X_i^n \mathbb{1}_{i\leq Z_{n-1}}$ donc Z_n est bien défini.
- On remarque que Z_n est défini à partir de Z_{n-1}. Comme Z_{n-1} est discrète, on a :

$$\mathbb{E}[\phi(Z_n) \mid \underbrace{Z_{n-1}}_{\text{discrète}}] = h(Z_{n-1})$$

où $h(z) = \mathbb{E}[\phi(Z_n) \mid Z_{n-1} = z] = \mathbb{E}\Big[\phi\Big(\sum_{i=1}^z X_i^n\Big) \mid Z_{n-1} = z\Big]$.
On a $(Z_{n-1} = z) \in \sigma((X_i^k)_{i\geq 1, k\leq n-1})$ donc $Z_{n-1} = z$ et $(X_i^n)_{i\geq 1}$ sont indépendants par regroupement, d'où :

$$h(z) = \mathbb{E}\Big[\phi\Big(\underbrace{\sum_{i=1}^z X_i^n}_{\sim \mu^{*z}}\Big)\Big] = \int \phi(u)\, d\mu^{*z}(u)$$

où μ^{*z} est la loi μ convoluée z fois avec elle-même.
Ainsi $\mathbb{E}[\phi(Z_n) \mid Z_{n-1}] = \int \phi(z)\, d\mu^{*Z_{n-1}}(z)$ et $\mathcal{L}(Z_n \mid Z_{n-1}) = d\mu^{*Z_{n-1}}$.

On se demande si la descendance est finie ou non ?

REMARQUE 8.28. Si $\mu(\{0\}) = 0$, alors la population est croissante et donc la dépendance est infinie.

On a $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$ donc si A est l'événement "on a extinction de la population", alors on a $A = \bigcup_{n\geq 1} \uparrow \{Z_n = 0\}$. Notre but est de trouver $\rho = \mathbb{P}(A) = \lim_{n\to+\infty} \uparrow \mathbb{P}(Z_n = 0)$.

On va utiliser les fonctions génératrices, qui jouent le rôle des fonctions caractéristiques pour les variables aléatoires à valeurs dans $\mathbb{N}$:

DÉFINITION 8.29. [FONCTION GÉNÉRATRICE]
Si $Z \in \mathbb{N}$ p.s., on définit sa *fonction génératrice* par :

$$\begin{array}{rccc} g_Z : & [0,1] & \longrightarrow & [0,1] \\ & s & \longmapsto & \mathbb{E}[s^Z] = \sum_{k\geq 0} p_k s^k \end{array}$$

où $p_k = \mathbb{P}(Z = k) \geq 0$.

PROPOSITION 8.30. *Soit $Z \in \mathbb{N}$ p.s.. Alors :*

- *g_Z est analytique sur $[0,1[$,*
- *g_Z est croissante et convexe, et même strictement convexe si $\mathbb{P}(Z \geq 2) > 0$,*
- *$g_Z(1) = 1$ et $g'_Z(1^-) = \mathbb{E}[Z]$ (donc $g'_Z(1^-) < +\infty$ si $Z \in L^1$).*

PREUVE

- Si $s \in [0,t]$ avec $t < 1$, on a :
$$g_Z(s) = \sum_{k\geq 0} \underbrace{p_k s^k}_{|\cdot| \leq t^k}$$
donc on a convergence normale sur $[0,t]$.
- On a croissance puisque les $(p_k)_{k\in\mathbb{N}}$ sont positifs. En effet $g''_Z(s) = \sum_{k\geq 2} k(k-1)p_k s^{k-2} \geq 0$.
On a stricte inégalité dans cette dernière inégalité s'il existe $k \geq 2$ tel que $p_k > 0$.
- On a $g'_Z(s) = \sum_{k\geq 1} k p_k s^{k-1} \xrightarrow[s\to 1^-]{\uparrow} \sum_{k\geq 1} k p_k = \mathbb{E}[Z]$.

□

Calculons les fonctions génératrices $(g_{Z_n})_{n\in\mathbb{N}}$:

- On a $g_{Z_0}(s) = s^1 = s$.
- Pour $n \geq 1$, on a :
$$\begin{aligned} g_{Z_n}(s) = \mathbb{E}[s^{Z_n}] &= \mathbb{E}\left[s^{\sum_{i=1}^{Z_{n-1}} X_i^n}\right] \\ &= \mathbb{E}\left[s^{\sum_{i=1}^{Z_{n-1}} X_i^n} \sum_{z=0}^{+\infty} \mathbb{1}_{Z_{n-1}=z}\right] \\ &= \sum_{z=0}^{+\infty} \mathbb{E}\left[s^{\sum_{i=1}^{z} X_i^n} \mathbb{1}_{Z_{n-1}=z}\right] && \text{par le théorème de FUBINI positif} \\ &= \sum_{z=0}^{+\infty} \prod_{i=1}^{z} \mathbb{E}[s^{X_i^n}]\mathbb{P}(Z_{n-1} = z) && \text{par indépendance} \\ &= \sum_{z=0}^{+\infty} (g_\mu(s))^z \mathbb{P}(Z_{n-1} = z) \end{aligned}$$
Et donc finalement $g_{Z_n}(s) = g_{Z_{n-1}}(g_\mu(s))$, ce qui entraîne : $g_{Z_n}(s) = g_{Z_0}(g_\mu^{\circ n}(s))$.

Ainsi

$$\boxed{\forall n \in \mathbb{N}, g_{Z_n}(s) = g_\mu^{\circ n}(s)}$$

PROPOSITION 8.31. *La probabilité d'extinction $\rho = \lim_{n\to+\infty} \uparrow \mathbb{P}(Z_n = 0)$ est le plus petit point fixe de g_μ sur $[0,1]$.*

REMARQUE 8.32. On a $g_\mu(1) = 1$ donc il y a au moins un point fixe.

PREUVE En remarquant que $\mathbb{P}(Z_n = 0) = g_{Z_n}(0) = g_\mu^{\circ n}(0)$, il vient :

$$\rho = \lim_{n\to+\infty} g_\mu^{\circ n}(0) = g_\mu\Big(\lim_{n\to+\infty} g_\mu^{\circ n}(0)\Big) = g_\mu(\rho)$$

Donc ρ est un point fixe. Par ailleurs, si z est un autre point fixe. Alors $0 \leq z$ donc par croissance de g_μ :

$$\rho = \lim_{n\to+\infty} g_\mu^{\circ n}(0) \leq \lim_{n\to+\infty} g_\mu^{\circ n}(z) = z$$

Donc ρ est bien le plus petit point fixe de g_μ. □

THÉORÈME 8.33. *Si $\mu \neq \delta_1$ et si $m = \mathbb{E}[X_1^1] = \int x\, d\mu(x)$, alors :*

- *lorsque $m \leq 1$, $\rho = 1$ et on a extinction presque sûrement,*
- *lorsque $m > 1$, $\rho < 1$ et on a survie avec probabilité strictement positive.*

REMARQUE 8.34. On parle de "transition de phase" en $m = 1$. D'un côté, l'arbre est fini presque sûrement. De l'autre, la probabilité qu'il soit fini est positive strictement.

PREUVE

- Si $m > 1$:
  On a $g_\mu(1) = 1$ et $g'_\mu(1) = m > 1$ donc $\exists \varepsilon > 0 \,|\, \forall y \in]1-\varepsilon, 1[, g_\mu(y) < y \quad (\iff g_\mu(y) - y < 0)$.
  Comme $g_\mu(0) \geq 0$, on a $g_\mu(0) - 0 > 0$.
  Par continuité de $y \mapsto g_\mu(y) - y$, la fonction s'annule en un point de $]0, 1-\varepsilon[$, donc il existe un point fixe strictement inférieur à 1.
- Si $m < 1$:
  On a $g'_\mu(1) = m < 1$ et g_μ est convexe donc reste supérieure à la tangente en 1 donc strictement supérieure à la droite $y = x$. Ainsi le seul point fixe est 1, donc $\rho = 1$.
- Si $m = 1$:
  On a $g'_\mu(1) = m = 1$ donc la tangente est la diagonale $y = x$.
  Si $\mathbb{P}(X \geq 2) = 0$, alors $p_1 = 1$ donc $\mu = \delta_1$, ce qui est faux par hypothèse.
  Donc $\mathbb{P}(X \geq 2) > 0$ et par stricte convexité on a le résultat.

□

Deuxième partie

Processus

CHAPITRE 9

Martingales

I. Théorie générale des processus

Définition 9.1. [Processus]

Un *processus* X sur l'espace de temps $\mathcal{T}$ à valeurs dans l'espace $(E, \mathcal{E})$ est une variable aléatoire $X = (X_t)_{t\in\mathcal{T}}$ à valeurs dans $(E^{\mathcal{T}}, \mathcal{C})$ où $\mathcal{C}$ est la tribu cylindrique.

Un processus X est donc une fonction $\mathcal{T} \longrightarrow E$ aléatoire. On a souvent $\mathcal{T} = \mathbb{R}^+, \mathbb{N}, \mathbb{R}, \mathbb{Z}, \ldots$.

Dans ce cours, on considérera uniquement $\boxed{\mathcal{T} = \mathbb{N}}$ et la *tribu cylindrique* engendrée par les événements du type $\{x : \mathcal{T} \longrightarrow E \,|\, \forall i \in [\![1, n]\!], x(t_i) \in A_i\}$ où $n \in \mathbb{N}$, $(t_i)_{1\leq i\leq n} \in \mathcal{T}^n$ et $(A_i)_{1\leq i\leq n} \in E^n$ sont fixés.

Exemple 9.2. Les marches aléatoires sont des processus.

Exemple 9.3. [Processus de Poisson]

Soit $(Y_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{E}(1)$. Chaque Y_i modélise par exemple le temps d'apparition d'une panne d'une machine.
On pose pour $t \geq 0$, $N_t =$ nombre de pannes avant $t = \sup(\{n \in \mathbb{N} \,|\, \sum_{i=1}^n Y_i \leq t\})$.

N est un processus sur $\mathbb{R}^+$ appelé processus de Poisson (car $N_t \sim \mathcal{P}(t)$).

Exemple 9.4. [Mouvement brownien]

Il existe une variable aléatoire $B = (B_t)_{t\geq 0} \in \mathbb{R}^{\mathbb{R}^+}$ telle que :

$$\begin{cases} B_0 = 0 \text{ p.s.} & \\ B_t - B_s \sim \mathcal{N}(0, t-s) & \text{(\textit{accroissements stationnaires})} \\ \forall t_0 < t_1 < \cdots < t_n, \text{ les } (B_{t_{i+1}} - B_{t_i})_{0\leq i\leq n-1} \text{ sont indépendants} & \text{(\textit{accroissements indépendants})} \\ t \longmapsto B_t \text{ est continue p.s.} & \end{cases}$$

On l'appelle le *mouvement brownien*. Son existence est compliquée (surtout à cause de la propriété de continuité qui n'est pas une propriété de la loi de B sur $(\mathbb{R}^+, \mathcal{C})$ où $\mathcal{C}$ est la tribu cylindrique).

Désormais, l'espace de temps $\mathcal{T}$ sera $\mathbb{N}$ ou $\mathbb{N}^*$.

DÉFINITION 9.5. [FILTRATION]

- Une *filtration* $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ est une suite croissante de tribus de $\mathcal{A}$.
- Si $X = (X_n)_{n\in\mathbb{N}}$ est un processus, on dit qu'il est *adapté* à la filiation $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ si pour tout $n \in \mathbb{N}$, X_n est $\mathcal{F}_n$-mesurable.
- Si $X = (X_n)_{n\in\mathbb{N}}$ est un processus, sa *filtration canonique* est $\mathcal{F}^X = (\mathcal{F}_n^X)_{n\in\mathbb{N}}$ définie par

$$\forall n \in \mathbb{N}, \mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$$

On a bien sûr que X est adapté à $\mathcal{F}^X$, et que toute filtration adaptée à X contient $\mathcal{F}^X$. $\mathcal{F}^X$ est donc la plus petite filtration adaptée à X.

EXEMPLE 9.6. [MARCHE ALÉATOIRE]

La marche aléatoire S est adaptée à la filtration canonique : $\mathcal{F}_n^S = \mathcal{F}_n^X = \sigma(\{X_0, \dots, X_n\})$. En effet :

- $S_n^x = x + \sum_{i=1}^n X_i$ est $\sigma(X_0, \dots, X_n)$-mesurable.
- $X_n = S_n - S_{n-1}$ est $\sigma(S_0, \dots, S_n)$-mesurable donc $\mathcal{F}^X \subseteq \mathcal{F}^S$.

EXEMPLE 9.7. [PROCESSUS DE GALTON-WATSON]

On reprend les notations de la section V. du chapitre 8 : on rappelle que l'on a définit Z par :

$$\begin{cases} Z_0 = 1 \\ Z_n = \sum_{i=1}^{Z_{n-1}} X_i^n \quad \text{pour } n \in \mathbb{N} \end{cases}$$

où les $(X_j^i)_{i,j\geq 1}$ sont des variables aléatoires i.i.d. de loi μ.

Alors Z est adapté à la filtration définie par $\mathcal{F}_n = \sigma(\{X_j^i \mid i \geq 1 \text{ et } 1 \leq j \leq n\})$ pour $n \in \mathbb{N}$.

L'intérêt de la filtration est qu'elle permet de *mesurer le temps*, dans le sens où, si l'on est au temps n :

- être $\mathcal{F}_n$-mesurable signifie que l'on est une fonction du passé,
- être $\mathcal{F}_{n+1}$-mesurable signifie que l'on est prévisible (on pouvait deviner la valeur avant).

II. Définition des martingales

DÉFINITION 9.8. [MARTINGALE]

Soit $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration et $M = (M_n)_{n\in\mathbb{N}}$ un processus adapté à $\mathcal{F}$ tel que $\forall n \in \mathbb{N}, M_n \in L^1$.

- On dit que M est une *$\mathcal{F}$-martingale* si :

$$\boxed{\forall n \in \mathbb{N}, \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n \text{ p.s.}}$$

- On dit que M est une *$\mathcal{F}$-sous-martingale* si :

$$\forall n \in \mathbb{N}, \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \geq M_n \text{ p.s.}$$

- On dit que M est une *$\mathcal{F}$-sur-martingale* si :

$$\forall n \in \mathbb{N}, \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \leq M_n \text{ p.s.}$$

Comment interpréter une martingale ?

- $\mathbb{E}[M_{n+1} \mid \mathcal{F}_n]$ est la meilleure prédiction sur le processus au temps $n+1$ sachant ce qu'il s'est passé au temps n. Pour une martingale, en moyenne, il n'y a pas d'évolution prévisible au temps n.
- On imagine un joueur dans un casino pour lequel on note M_n sa fortune au bout de n parties. A la $(n+1)$-ème partie, ses gains sont $M_{n+1} - M_n$.
  Au temps présent n, il peut estimer ce gain par :

$$\mathbb{E}[M_{n+1} - M_n \mid \mathcal{F}_n] = \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] - M_n$$

  Si le jeu du casino est équilibré, alors M est une martingale.

EXEMPLE 9.9. [MARCHE ALÉATOIRE]

Considérons $(S_n^x)_{n\in\mathbb{N}}$ une marche aléatoire issue de x et notons $\mathcal{F} = \mathcal{F}^X$. On suppose les $(X_i)_{i\geq 1}$ de classe L^1. Alors :

$$\mathbb{E}[S_{n+1}^x \mid \mathcal{F}_n] = \mathbb{E}[S_n^x + X_{n+1} \mid \mathcal{F}_n] = S_n^x + \mathbb{E}[X_{n+1} \mid \mathcal{F}_n]$$

Donc la marche aléatoire est :

- une martingale si $\mathbb{E}[X_1] = 0$,
- une sous-martingale si $\mathbb{E}[X_1] \geq 0$,
- une sur-martingale si $\mathbb{E}[X_1] \leq 0$.

REMARQUE 9.10.

- S'il n'y a pas d'ambiguïté, on ne reprécise par la filtration et on dit martingale au lieu de $\mathcal{F}$-martingale.
- Si la filtration n'est pas précisée, on utilise la filtration canonique.

On a bien sûr :

- M est une martingale si et seulement si M est une sous-martingale et une sur-martingale,
- M est une sous-martingale si et seulement si $-M$ est une sur-martingale,
- Les notions de martingale, sous-martingale et sur-martingale sont stables par somme.

PROPOSITION 9.11. *Soit $M = (M_n)_{n\in\mathbb{N}}$ un processus adapté à $\mathcal{F}$ (où $\forall n \in \mathbb{N}, M_n \in L^1$). Alors :*

- M *est une martingale* $\iff \forall k \in \mathbb{N}, \mathbb{E}[M_{n+k} \mid \mathcal{F}_n] = M_n$ *p.s.*
- M *est une sous-martingale* $\iff \forall k \in \mathbb{N}, \mathbb{E}[M_{n+k} \mid \mathcal{F}_n] \geq M_n$ *p.s.*
- M *est une sur-martingale* $\iff \forall k \in \mathbb{N}, \mathbb{E}[M_{n+k} \mid \mathcal{F}_n] \leq M_n$ *p.s.*

PREUVE

$\Longleftarrow$ On applique avec $k = 1$.

$\Longrightarrow$ On procède par récurrence. Par exemple pour le deuxième point (les autres sont similaires) :

- Si $k = 0$, on a $\mathbb{E}[M_n \mid \mathcal{F}_n] = M_n \geq M_n$ p.s..
- Si le résultat est vrai au rang k, alors :

$$\mathbb{E}[M_{n+k+1} \mid \mathcal{F}_n] = \mathbb{E}[\underbrace{\mathbb{E}[M_{n+k+1} \mid \mathcal{F}_{n+k}]}_{\geq M_{n+k} \text{ p.s.}} \mid \mathcal{F}_n] \geq \mathbb{E}[M_{n+k} \mid \mathcal{F}_n] \geq M_n$$

où l'on a utilisé pour la première inégalité la croissance de l'espérance conditionnelle, et pour la seconde l'hypothèse de récurrence.
D'où le résultat au rang $k + 1$.

□

EXEMPLE 9.12. [PROCESSUS DE GALTON-WATSON]

Supposons que les $(X_j^i)_{i,j\geq 1}$ sont L^1 et posons $m = \mathbb{E}[X_1^1]$. On a :

$$\begin{aligned}
\mathbb{E}[Z_n] &= \mathbb{E}\Big[\sum_{k\geq 0} Z_n \mathbb{1}_{Z_{n-1}=k}\Big] \\
&= \mathbb{E}\Big[\sum_{k\geq 0} (X_n^1 + \cdots + X_n^k)\mathbb{1}_{Z_{n-1}=k}\Big] \\
&= \sum_{k\geq 0} \mathbb{E}\Big[\mathbb{1}_{Z_{n-1}=k} \sum_{\ell=1}^{k} X_n^\ell\Big] && \text{par le théorème de FUBINI} \\
&= \sum_{k\geq 0} \mathbb{E}\Big[\sum_{\ell=1}^{k} X_n^\ell\Big]\mathbb{P}(Z_{n-1} = k) && \text{par indépendance} \\
&= \Big(\sum_{k\geq 0} k\mathbb{P}(Z_{n-1} = k)\Big)m = \mathbb{E}[Z_{n-1}]m
\end{aligned}$$

Par une récurrence immédiate, on obtient :

$$\mathbb{E}[Z_n] = m^n$$

REMARQUE 9.13.

- Si M est une martingale, alors $(\mathbb{E}[M_n])_{n\in\mathbb{N}}$ est constante.
- Si M est une sous-martingale, alors $(\mathbb{E}[M_n])_{n\in\mathbb{N}}$ est croissante.
- Si M est une sur-martingale, alors $(\mathbb{E}[M_n])_{n\in\mathbb{N}}$ est décroissante.

Posons $Y_n = \dfrac{Z_n}{m^n}$ pour $n \in \mathbb{N}$. Alors on a pour tout $n \in \mathbb{N}$: $\mathbb{E}[Y_n] = 1$.

PROPOSITION 9.14. *$Y = (Y_n)_{n\in\mathbb{N}}$ est une martingale.*

PREUVE On calcule :

$$\begin{aligned}
\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] &= \frac{1}{m^{n+1}}\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] \\
&= \frac{1}{m^{n+1}}\mathbb{E}\Big[\sum_{k\geq 0} \mathbb{1}_{Z_n=k} \sum_{\ell=1}^{k} X_{n+1}^{\ell} \mid \mathcal{F}_n\Big] \\
&= \frac{1}{m^{n+1}}\sum_{k\geq 0} \mathbb{E}\Big[\mathbb{1}_{Z_n=k} \sum_{\ell=1}^{k} X_{n+1}^{\ell} \mid \mathcal{F}_n\Big] && \text{par le TCM conditionnel} \\
&= \frac{1}{m^{n+1}}\sum_{k\geq 0} \mathbb{1}_{Z_n=k}\mathbb{E}\Big[\sum_{\ell=1}^{k} X_{n+1}^{\ell} \mid \mathcal{F}_n\Big] && \text{puisque } \mathbb{1}_{Z_n=k} \text{ est } \mathcal{F}_n\text{-mesurable} \\
&= \frac{1}{m^{n+1}}\sum_{k\geq 0} km\mathbb{1}_{Z_n=k} = \frac{1}{m^n}\sum_{k\geq 0} k\mathbb{1}_{Z_n=k} \\
&= \frac{1}{m^n}Z_n = Y_n
\end{aligned}$$

□

II. A. Construction de martingales en pariant

On considère une $\mathcal{F}$-martingale M et un joueur de casino dont les gains à la $(n+1)$-ème partie jouée sont représentés par $M_{n+1} - M_n$. On suppose que la joueur parie la quantité 1 à chaque partie.
Un autre joueur parie la quantité H_{n+1} sur le résultat de cette partie $(n+1)$-ème partie. Ses gains lors de cette partie sont $H_{n+1}(M_{n+1} - M_n)$.

NOTE 9.15. On a par exemple :

- Si $H_{n+1} = 1$, le deuxième joueur gagne les gains normaux,
- Si $H_{n+1} = 2$, il gagne le double des gains,
- Si $H_{n+1} = 0$, il ne joue pas,
- Si $H_{n+1} = -1$, il joue pour le casino / contre le premier joueur.

Les gains cumulés par le second joueur au bout de n parties valent $\sum_{i=1}^{n} H_i(M_i - M_{i-1})$.
Le joueur ne voit pas le résultat de chaque partie à l'avance donc on demande que pour tout $n \in \mathbb{N}$, H_n soit $\mathcal{F}_{n-1}$-mesurable.

DÉFINITION 9.16. Soit $\mathcal{F}$ une filtration.

- On dit que $(H_n)_{n\in\mathbb{N}^*}$ est *prévisible* si pour tout $n \in \mathbb{N}^*$, H_n est $\mathcal{F}_{n-1}$-mesurable,
- Si $H = (H_n)_{n\in\mathbb{N}^*}$ est prévisible et $M = (M_n)_{n\in\mathbb{N}}$ est adapté, on définit le processus adapté $(H.M)$ par :

$$\begin{cases} (H.M)_0 = 0 \\ (H.M)_n = \sum_{i=1}^{n} H_i(M_i - M_{i-1}) \quad \text{pour } n \in \mathbb{N}^* \end{cases}$$

THÉORÈME 9.17. *Soit $\mathcal{F}$ une filtration.*

- *Si H est prévisible et borné ($\forall n \in \mathbb{N}, |H_n| \leq C_n$ p.s.) et M est une martingale, alors $(H.M)$ est une martingale.*
- *Si H est prévisible, borné et positif et M est une sur-martingale, alors $(H.M)$ est une sur-martingale.*
- *Si H est prévisible, borné et positif et M est une sous-martingale, alors $(H.M)$ est une sous-martingale.*

REMARQUE 9.18. On retourne au casino. Si H est équilibré et M est une martingale, ceci signifie que quelque soit la stratégie de mise, H_n peut être n'importe quelle fonction $f_n(M_0, \dots, M_{n-1})$, alors $(H.M)$ est une martingale. En particulier, $\mathbb{E}[(H.M)_n] = 0$. Ainsi, il n'y a pas de gain positif en espérance ("on ne peut pas battre le casino").
Réciproquement, si pour tout H prévisible, M est adapté et pour tout $n \in \mathbb{N}$, $\mathbb{E}[(H.M)_n] = 0$, alors M est une martingale.

PREUVE

On remarque d'abord que si H est borné et les $(M_n)_{n\in\mathbb{N}}$ sont L^1, alors pour tout $n \in \mathbb{N}$, $(H.M)_n$ est L^1.

- Pour le premier point, on a :

$$\begin{aligned} \mathbb{E}[(H.M)_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}[(H.M)_n + H_{n+1}(M_{n+1} - M_n) \mid \mathcal{F}_n] \\ &= (H.M)_n + H_{n+1}(\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] - M_n) \\ &= (H.M)_n \qquad \text{car } M \text{ est une martingale} \end{aligned}$$

- Les deux autres points sont similaires.

□

III. Temps d'arrêt

On va introduire la notion de *temps d'arrêt*, qui est aussi utile en dehors du cadre des martingales.

DÉFINITION 9.19. [TEMPS D'ARRÊT]

Soit $\mathcal{F}$ une filtration. On dit que $T : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ est un *temps d'arrêt* si pour tout $n \in \mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$.

REMARQUE 9.20. Un joueur peut jouer jusqu'à un instant T qu'il doit pouvoir décider uniquement en fonction de ce qu'il s'est passé jusque là donc l'événement "le joueur arrête au temps n" doit être $\mathcal{F}_n$-mesurable.

On a :

$$\begin{aligned} T \text{ est un temps d'arrêt} &\iff \forall n \in \mathbb{N}, \{T = n\} \in \mathcal{F}_n \\ &\iff \forall n \in \mathbb{N}, \{T \leq n\} \in \mathcal{F}_n \\ &\iff \forall n \in \mathbb{N}, \{T > n\} \in \mathcal{F}_n \end{aligned}$$

EXEMPLE 9.21. Soient X un processus à valeurs dans $\mathbb{R}$ adapté à $\mathcal{F}$, $A \in \mathcal{B}(\mathbb{R})$ et $T_A = \inf\{n \in \mathbb{N} \mid X_n \in A\}$ le temps d'attente de A (avec la convention $\inf \varnothing = +\infty$).
Alors T_A est un temps d'arrêt. En effet, on a pour $n \in \mathbb{N}$ puisque $\mathcal{F}_n$ est adapté :

$$\{T_A > n\} = \{\forall i \leq n, X_i \notin A\} \in \mathcal{F}_n$$

Pour la marche aléatoire simple sur $\mathbb{Z}$, on peut considérer, pour $L \in \mathbb{N}$ fixé, T_L le plus petit entier n tel que $S_n = L$. On sait que $T_L < +\infty$ p.s. puisque $\limsup_{n\to+\infty} S_n = +\infty$ p.s.. On s'intéressera alors à la suite à $\mathbb{E}[T_L]$.

Soit $T = \inf\{n \in \mathbb{N} \mid X_n < 0 \text{ et } \exists 0 \leq p < n \mid X_p > 10\}$ le premier passage strictement négatif après un passage strictement supérieur à 10.
Alors $\{T > n\} = \{T_{\{10\}} > n\} \cup \bigcup_{p=0}^{n-1} \{X_p > 10 \text{ et } \forall j \in [\![p+1, n]\!], X_j \geq 0\} \in \mathcal{F}_n$.

PROPOSITION 9.22.

- *Si S et T sont deux temps d'arrêt, alors $S \wedge T = \min(S, T)$ et $S \vee T = \max(S, T)$ sont des temps d'arrêt.*
- *Si $(T_k)_{k\in\mathbb{N}}$ est une suite de temps d'arrêts alors $\sup_{k\in\mathbb{N}} T_k$ et $\inf_{k\in\mathbb{N}} T_k$ sont des temps d'arrêt.*

PREUVE Montrons par exemple que $\sup_{k\in\mathbb{N}} T_k$ est un temps d'arrêt. On a pour $n \in \mathbb{N}$:

$$\left\{\sup_{k\in\mathbb{N}} T_k \leq n\right\} = \bigcap_{k\in\mathbb{N}} \{T_k \leq n\} \in \mathcal{F}_n$$

□

On définit, pour T temps d'arrêt :

$$\boxed{\mathcal{F}_T = \{A \in \mathcal{F}_\infty \mid \forall n \in \mathbb{N}, A \cap \{T = n\} \in \mathcal{F}_n\}}$$

où $\mathcal{F}_\infty = \sigma(\bigcup_{n\in\mathbb{N}} \uparrow \mathcal{F}_n)$.

Soit un joueur dans un casino dont la fortune est modélisée par le processus X.
S'il joue jusqu'au temps d'arrêt T, ses gains sont donnés par le "processus arrêté" $(X_{n\wedge T})_{n\in\mathbb{N}}$ et sa fortune finale sera (si $T < +\infty$ p.s.) X_T.

PROPOSITION 9.23. *Soit X adapté et T un temps d'arrêt. Alors T et X_T sont $\mathcal{F}_T$-mesurables.*

PREUVE

- T est $\mathcal{F}_T$-mesurable :
  On a $\sigma(T) = \sigma(\{\{T = k\} \mid k \in \mathbb{N}\})$. Fixons $k \in \mathbb{N}$ et vérifions que $\{T = k\} \in \mathcal{F}_T$, c'est-à-dire que pour tout $n \in \mathbb{N}$, $\{T = k\} \cap \{T = n\} \in \mathcal{F}_n$:
  $$\{T = k\} \cap \{T = n\} = \begin{cases} \{T = n\} \in \mathcal{F}_n & \text{si } n = k \\ \varnothing \in \mathcal{F}_n & \text{sinon} \end{cases}$$
  Donc T est $\mathcal{F}_T$-mesurable.
- X_T est $\mathcal{F}_T$-mesurable :
  On a $\sigma(X_T) = \sigma(\{X_T \in A\} \mid A \in \mathcal{B}(\mathbb{R}))$. Or, pour $n \in \mathbb{N}$:
  $$\{X_T \in A\} \cap \{T = n\} = \underbrace{\{X_n \in A\}}_{\in \mathcal{F}_n \text{ car } X \text{ adapté}} \cap \{T = n\} \in \mathcal{F}_n$$

□

LEMME 9.24. *Soient S et T des temps d'arrêt tels que $S \leq T$. Alors $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

PREUVE Soit $A \in \mathcal{F}_S$. On a pour $n \in \mathbb{N}$:
$$A \cap \{T = n\} = \bigcup_{p=0}^{n} \underbrace{A \cap \{S = p\}}_{\in \mathcal{F}_p \subset \mathcal{F}_n} \cap \{T = n\} \in \mathcal{F}_n$$

□

REMARQUE 9.25. $T = n$ est un temps d'arrêt. On a bien $\mathcal{F}_T = \mathcal{F}_n$.

REMARQUE 9.26. Soient $(X_i)_{i \in \mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées L^1 telles que $\mathbb{E}[X_i] > 0$. On sait que $\frac{1}{n}\sum_{i=1}^n X_i \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X_1] > 0$ p.s. donc $\sum_{i=1}^n X_i \underset{n\to+\infty}{\longrightarrow} +\infty$ p.s..
Alors $T = \sup\{n \in \mathbb{N} \mid \sum_{i=1}^n X_i < 0\}$ est fini presque sûrement. Cependant, ce n'est pas un temps d'arrêt en général !

THÉORÈME 9.27. [THÉORÈME D'ARRÊT (FAIBLE)]

Soit $\mathcal{F}$ une filtration, M une martingale et T un temps d'arrêt. Alors :

- *$(M_{n\wedge T})_{n\in\mathbb{N}}$ est une martingale, appelée* martingale arrêtée.
- *s'il existe $C < +\infty$ tel que $T < C$ p.s., alors $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.*

NOTE 9.28. Le premier point est très utile. Le second est peu utilisé puisque l'hypothèse sur T est assez contraignante.

PREUVE

- Soit, pour $n \in \mathbb{N}$, $H_n = \mathbb{1}_{n \leq T}$. Alors H_n est $\mathcal{F}_{n-1}$-mesurable puisque $\{n \leq T\} = \{T > n-1\}$. Donc H est prévisible borné.

Donc $(H.M)$ est une martingale. Or, on a pour $n \in \mathbb{N}$:

$$(H.M)_n = \sum_{i=1}^{n} H_i(M_i - M_{i-1}) = \sum_{i=1}^{n\wedge T}(M_i - M_{i-1}) = M_{n\wedge T} - M_0$$

Le processus constant $(M_0)_{n\in\mathbb{N}}$ est une martingale, donc $(M_{n\wedge T})_{n\in\mathbb{N}}$ est une martingale en tant que somme de martingales.

– $(M_{n\wedge T})_{n\in\mathbb{N}}$ est une martingale donc :

$$\mathbb{E}[M_{n\wedge T}] = \mathbb{E}[M_{0\wedge T}] = \mathbb{E}[M_0]$$

Or, pour $n > C$, on a $n \wedge T = T$ p.s., donc $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

□

III. A. Application à la ruine du joueur

a. Cas d'un casino équilibré

Soient $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées de loi de RADEMACHER :

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

On considère un joueur qui commence avec la somme $x \in \mathbb{N}^*$ et dont on modélise le gain à la i-ème partie par X_i. Le joueur s'arrête s'il obtient une somme $L \in \mathbb{N}$ avec $L > x$ ou s'il est ruiné. Sa fortune au temps $n \in \mathbb{N}$ (s'il ne s'est pas arrêté avant) est $S_n^x = x + \sum_{i=1}^n X_i$.
On pose $T_{\{0,L\}} = \inf\{n \in \mathbb{N} \mid S_n = 0 \text{ ou } S_n = L\}$. C'est un temps d'arrêt.

On cherche la probabilité de faire fortune et le temps de jeu.

Puisque $\limsup_{n\to+\infty} S_n = +\infty$ p.s. et $\liminf_{n\to+\infty} S_n = -\infty$ p.s., on a $T_{\{0,L\}} < +\infty$ p.s..

– Probabilité de faire fortune. Notons $A = \{S_{T_{\{0,L\}}} = L\}$ l'événement "le joueur fait fortune". On sait que $(S^x_{n\wedge T_{\{0,L\}}})_{n\in\mathbb{N}}$ est une martingale, donc :

$$\mathbb{E}[S^x_{n\wedge T_{\{0,L\}}}] = \mathbb{E}[S^x_{0\wedge T_{\{0,L\}}}] = x$$

Ainsi, par convergence dominée, puisque $S^x_{n\wedge T_{\{0,L\}}} \leq L$ et $S^x_{n\wedge T_{\{0,L\}}} \xrightarrow[n\to+\infty]{\text{p.s.}} S^x_{T_{\{0,L\}}}$:

$$\mathbb{E}[S^x_{T_{\{0,L\}}}] = x$$

Ainsi $x = \mathbb{E}[\mathbb{1}_A L + \mathbb{1}_{A^c} 0] = L\mathbb{P}(A)$. On en déduit la probabilité de A :

$$\boxed{\mathbb{P}(\text{"le joueur fait fortune"}) = \frac{x}{L}}$$

REMARQUE 9.29. – Le résultat est assez intuitif puisque la probabilité tend vers 1 lorsque $x \to L$ et vers 0 lorsque $x \to 0$.

- on utilise très peu de choses sur S. On peut montrer que le résultat reste vrai si l'on autorise des paris $H : M_n = x + (H.S)_n$ où H est prévisible, entier, prend une infinité de fois des valeurs non nulles presque sûrement et vérifie $\forall n \in \mathbb{N}, |H_n| \leq \min(M_n, L - M_n)$.

– Temps de jeu : on s'intéresse au temps de jeu moyen : que vaut $\mathbb{E}[T_{\{0,L\}}]$?
On a, pour $n \in \mathbb{N}$:

$$\mathbb{E}[(S^x_{n+1})^2 \,|\, \mathcal{F}_n] = \mathbb{E}[(S^x_n)^2 + 2X_{n+1}S^x_n + X^2_{n+1} \,|\, \mathcal{F}_n] = (S^x_n)^2 + 2\underbrace{\mathbb{E}[X_{n+1} \,|\, F_n]}_{=\mathbb{E}[X_{n+1}]=0} S^x_n + 1 = (S^x_n)^2 + 1$$

Posons alors $Y_n = (S^x_n)^2 - n$ pour $n \in \mathbb{N}$. On a :

$$\mathbb{E}[Y_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[(S^x_{n+1})^2 \,|\, \mathcal{F}_n] - n - 1 = (S^x_n)^2 + 1 - n - 1 = Y_n$$

$(Y_{n\wedge T_{\{0,L\}}})_{n\in\mathbb{N}}$ est donc une martingale. D'où $\mathbb{E}[Y_{n\wedge T_{\{0,L\}}}] = \mathbb{E}[Y_{0\wedge T_{\{0,L\}}}]$.
Or, d'une part $\mathbb{E}[Y_{0\wedge T_{\{0,L\}}}] = \mathbb{E}[(S^x_0)^2 - 0] = x^2$.
Et d'autre part $\mathbb{E}[Y_{n\wedge T_{\{0,L\}}}] = \mathbb{E}[(S^x_{n\wedge T_{\{0,L\}}})^2] - \mathbb{E}[n \wedge T_{\{0,L\}}]$. Le premier terme tend vers $\mathbb{E}[(S^x_{T_{\{0,L\}}})^2]$ par convergence dominée et le second vers $\mathbb{E}[T_{\{0,L\}}]$ par convergence monotone.

Finalement on obtient $x^2 = \mathbb{E}[(S^x_{T_{\{0,L\}}})^2] - \mathbb{E}[T_{\{0,L\}}]$, et puisque $(S^x_{T_{\{0,L\}}})^2 = L^2\mathbb{1}_A + \mathbb{1}_{A^c}0$, il vient $\mathbb{E}[T_{\{0,L\}}] = L^2\mathbb{P}(A)$. D'où :

$$\boxed{\mathbb{E}[T_{\{0,L\}}] = (L - x)x}$$

REMARQUE 9.30. La durée de jeu la plus longue correspond donc à une mise initiale qui vaut la moitié de la somme souhaitée. C'est assez logique puisque l'on a une symétrie.

REMARQUE 9.31. On peut avoir un temps d'arrêt fini presque sûrement mais d'espérance infinie. Soit par exemple la marche aléatoire qui part de 1. On sait que $T_0 = \inf\{n \in \mathbb{N} \,|\, S^1_n = 0\} < +\infty$ p.s..
Or, $T_0 = \lim_{L\to+\infty} \uparrow T_{\{0,L\}}$, donc par convergence monotone :

$$\mathbb{E}[T_0] = \lim_{L\to\infty} \uparrow \mathbb{E}[T_{\{0,L\}}] = \lim_{L\to\infty} L - 1 = +\infty$$

b. Cas d'un casino déséquilibré

Soient $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées telles que :

$$\mathbb{P}(X_i = 1) = p \quad \text{et} \quad \mathbb{P}(X_i = -1) = q = 1 - p \quad \text{avec } 0 < p < \frac{1}{2} < q < 1$$

$(S_n)_{n\in\mathbb{N}}$ n'est pas une martingale : on cherche alors une fonction f telle que $(f(S_n))_{n\in\mathbb{N}}$ soit une martingale.
Cherchons $\alpha > 0$ tel que $(\alpha^{S^x_n})_{n\in\mathbb{N}}$ soit une martingale ($\alpha \neq 1$ si possible) et posons, pour $n \in \mathbb{N}$, $Y_n = \alpha^{S^x_n}$. On a :

$$\mathbb{E}[Y_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[\alpha^{S^x_n}\alpha^{X_{n+1}} \,|\, \mathcal{F}_n] = \alpha^{S^x_n}\mathbb{E}[\alpha^{X_{n+1}}] = \alpha^{S^x_n}(p\alpha + q\alpha^{-1})$$

Donc $(Y_n)_{n\in\mathbb{N}}$ est une martingale si et seulement si $p\alpha + q\alpha^{-1} = 1 \iff \alpha = 1$ ou $\alpha = \frac{q}{p}$.
Prenons donc $\alpha = \frac{q}{p}$. On a que $(Y_{n\wedge T_{\{0,L\}}})_{n\in\mathbb{N}}$ est une martingale, donc :

$$\mathbb{E}[Y_{n\wedge T_{\{0,L\}}}] = \mathbb{E}[Y_{0\wedge T_{\{0,L\}}}] = \left(\frac{q}{p}\right)^x$$

Or, en utilisant le lemme de BOREL-CANTELLI, on vérifie que $\{T_{\{0,L\}} < +\infty\}$ p.s..
Et puisque $Y_{n\wedge T_{\{0,L\}}} \underset{n\to+\infty}{\longrightarrow} Y_{T_{\{0,L\}}}$, on a par convergence dominée $\mathbb{E}[Y_{T_{\{0,L\}}}] = \left(\frac{q}{p}\right)^x$.
Ainsi $\left(\frac{q}{p}\right)^x = \mathbb{E}\left[\left(\frac{q}{p}\right)^L \mathbb{1}_A + \mathbb{1}_{A^c}\right]$ ou encore $\mathbb{P}(A)\left(\frac{q}{p}\right)^L + (1-\mathbb{P}(A)) = \left(\frac{q}{p}\right)^x$, et finalement :

$$\boxed{\mathbb{P}(\text{"le joueur fait fortune"}) = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^L - 1}}$$

IV. Convergence de martingales

On veut des hypothèses faibles qui assurent la convergence de martingales presque sûre ou dans L^1.

IV. A. Exemples

EXEMPLE 9.32. [MARCHE ALÉATOIRE SIMPLE]

Regardons d'abord la marche aléatoire simple sur $\mathbb{Z}$: on rappelle que l'on définit S^x par

$$\begin{cases} S_0^x = x \\ S_n^x = x + \sum_{i=1}^n X_i \quad \text{pour } n \in \mathbb{N}^* \end{cases}$$

où $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires i.i.d.de loi de RADEMACHER.

On a $\forall n \in \mathbb{N}, |S_{n+1}^x - S_n^x| = 1$ p.s. donc $(S_n^x)_{n\in\mathbb{N}}$ ne converge pas presque sûrement.
De plus, par le théorème central limite, on a $\frac{S_n^x}{\sqrt{n}} \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1)$, donc $\mathbb{E}[|S_n^x|] = O(\sqrt{n}) \underset{n\to+\infty}{\longrightarrow} +\infty$.

- Soit $T_{\{0,L\}} = \inf\{n \in \mathbb{N} \mid S_n \in \{0, L\}\}$ où x, L sont des entiers tels que $0 < x < L$.
  On a vu que $(S^x_{n\wedge T_{\{0,L\}}})_{n\in\mathbb{N}}$ est une martingale et $T_{\{0,L\}} < +\infty$ p.s..
  Donc $S^x_{n\wedge T_{\{0,L\}}} \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} S^x_{T_{\{0,L\}}}$.
  On a même convergence dans L^1, par convergence dominée puisque $\forall n \in \mathbb{N}, \left|S^x_{n\wedge T_{\{0,L\}}}\right| \leq L$.
- Considérons $(S^1_{n\wedge T_0})_{n\in\mathbb{N}}$ où $T_0 = \inf\{n \in \mathbb{N} \mid S_n^1 = 0\}$. On sait que $T_0 < +\infty$ p.s..
  Donc $S^1_{n\wedge T_0} \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} S^1_{T_0}$.
  Cependant, on n'a pas convergence dans L^1 puisque

  $$\mathbb{E}[S^1_{n\wedge T_0}] = S^1_{0\wedge T_0} = 1 \neq 0 = \mathbb{E}[S^1_{T_0}]$$

  On peut donc avoir $\mathbb{E}[M_T] \neq \mathbb{E}[M_0]$ pour T temps d'arrêt fini presque sûrement.

EXEMPLE 9.33.

– Revenons au casino et considérons la suite $H = (H_n)_{n\in\mathbb{N}^*}$ définie par :

$$\begin{cases} H_1 = 1 \\ \forall n \geq 2, H_n = \begin{cases} 2H_{n-1} & \text{si } X_{n-1} = -1 \\ 1 & \text{si } X_{n-1} = 1 \end{cases} \end{cases}$$

En gros, on part d'une mise initiale unitaire puis on double la mise lorsque l'on vient de perdre, et on remise 1 quand on vient de gagner.

Soit, pour $n \in \mathbb{N}$, $M_n = (H.S)_n = \sum_{i=1}^n H_i X_i$. Alors $(M_n)_{n\in\mathbb{N}}$ est une martingale.
Posons $T_1 = \inf(\{n \in \mathbb{N}^* \mid X_n = 1\})$. Alors :

$$(H.S)_{T_1} = \sum_{\ell=0}^{T_1-1} -2^\ell + 2^{T_1-1} = 1$$

On définit alors par récurrence, pour $i \in \mathbb{N}^*$, $T_{i+1} = \inf(\{n > T_i \mid X_n = 1\})$ le temps de la i-ème victoire. On en déduit $(H.S)_{T_i} = i$ p.s.. Ainsi :

$$(H.S)_{T_i} \underset{i\to+\infty}{\longrightarrow} +\infty \text{ p.s.}$$

Est-ce la bonne manière de parier ? Il semble en effet que l'on puisse gagner un gain infini. Cependant, on remarque que pour $i \in \mathbb{N}^*$, $T_{i+1} - T_i \sim \mathcal{G}\left(\frac{1}{2}\right)$ (en posant par convention $T_0 = 0$). Ce n'est donc pas la bonne stratégie de pari, car on peut avoir des temps d'attente très long sans gain, et donc être ruiné avant d'empocher un gain important.

– Soit $(Y_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes d'espérance nulle. Alors on vérifie que $(\sum_{i=1}^n Y_i)_{n\in\mathbb{N}}$ est une martingale.
Soit $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires i.i.d. de loi de RADEMACHER. Appliquant ce résultat à la suite définie par $M_n = \frac{1}{2} + \sum_{i=1}^n \frac{X_i}{2^i}$ pour $n \in \mathbb{N}$, on a que $(M_n)_{n\in\mathbb{N}}$ est une martingale et de plus $M_n \underset{n\to+\infty}{\overset{L^1|\text{ p.s.}}{\longrightarrow}} M_\infty$ où $M_\infty = \frac{1}{2} + \sum_{i=1}^{+\infty} \frac{X_i}{2^i}$.
Comme les $(\frac{X_i+1}{2})_{i\in\mathbb{N}^*}$ suivent une loi $\mathcal{B}(\frac{1}{2})$, on a que $M_\infty \sim \mathcal{U}([0,1])$.

– Soit $M = (M_n)_{n\in\mathbb{N}}$ une martingale telle que $M_n \underset{n\to+\infty}{\overset{L^1}{\longrightarrow}} M_\infty$.
On a, pour tout $k \in \mathbb{N}$, $\mathbb{E}[M_{n+k} \mid \mathcal{F}_n] = M_n$, donc $\mathbb{E}[M_\infty \mid \mathcal{F}_n] = M_n$ p.s. puisque $\mathbb{E}[. \mid \mathcal{B}]$ est linéaire et contractante pour L_1 (c'est-à-dire $|\mathbb{E}[. \mid \mathcal{B}]| \leq \mathbb{E}[|.| \mid \mathcal{B}]$) donc est continue pour L^1.

PROPOSITION 9.34. *Soit $Z \in L^1$ et $\mathcal{F}$ une filtration.*

Alors la suite $(M_n = \mathbb{E}[Z \mid \mathcal{F}_n])_{n\in\mathbb{N}}$ est une martingale, appelée martingale fermée.

REMARQUE 9.35. On vient de voir que si $M_n \underset{n\to+\infty}{\overset{L^1}{\longrightarrow}} M_\infty$, alors $(M_n)_{n\in\mathbb{N}}$ est fermée.

PREUVE On a $\forall n \in \mathbb{N}, \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[M_\infty \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[M_\infty \mid \mathcal{F}_n] = M_n$ p.s.. □

EXEMPLE 9.36. Soient $U \sim \mathcal{U}([0,1])$ et pour $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma\left(\left\{\frac{k}{2^n} \leq U < \frac{k+1}{2^n} \mid k \in \mathbb{Z}\right\}\right)$.
Alors $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}^*}$ est une filtration. Posons, pour $n \in \mathbb{N}$, $M_n = \mathbb{E}[U \mid \mathcal{F}_n]$. C'est une martingale

fermée.
Considérons le développement dyadique de U :

$$U = \sum_{i=1}^{\infty} \frac{U_i}{2^i} \text{ p.s.}$$

On sait que les $(U_i)_{i\in\mathbb{N}^*}$ sont des $\mathcal{B}(1/2)$ indépendantes.
On a, pour $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma(\{U_i \,|\, i \in [\![1, n]\!]\})$ et

$$\begin{aligned} M_n = \mathbb{E}[U \,|\, \mathcal{F}_n] &= \mathbb{E}\Big[\sum_{i=1}^{n} \frac{U_i}{2^i} \,|\, U_1, \dots, U_n\Big] + \mathbb{E}\Big[\sum_{i=n+1}^{\infty} \frac{U_i}{2^i} \,|\, U_1, \dots, U_n\Big] \\ &= \sum_{i=1}^{n} \frac{U_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \sum_{i=1}^{n} \frac{2U_i - 1}{2^{i+1}} + \sum_{i=1}^{\infty} \frac{1}{2^{i+1}} = \frac{1}{2} + \sum_{i=1}^{n} \frac{2U_i - 1}{2^{i+1}} \end{aligned}$$

On retrouve le même exemple que précédemment.

IV. B. Convergence presque sûre de martingales

LEMME 9.37.

- *Soient M une martingale et ϕ une fonction convexe. Alors $(\phi(M_n))_{n\in\mathbb{N}}$ est une sous-martingale.*
- *Soient M une sous-martingale et ϕ une fonction convexe et croissante. Alors $(\phi(M_n))_{n\in\mathbb{N}}$ est une sous-martingale.*

PREUVE

- Par l'inégalité de JENSEN, on a $\mathbb{E}[\phi(M_{n+1}) \,|\, \mathcal{F}_n] \geq \phi(\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n]) = \phi(M_n)$.
- De même, l'égalité étant remplacée par $\geq$.

□

Il y a en gros deux obstacles pour qu'une suite $(u_n)_{n\in\mathbb{N}}$ ne converge pas :

- soit $(|u_n|)_{n\in\mathbb{N}}$ diverge,
- soit $(u_n)_{n\in\mathbb{N}}$ oscille.

Soient $u = (u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ et $a < b$.
On pose $S^1_{a,b}(u) = \inf\{n \in \mathbb{N} \,|\, u_n \leq a\}$ et $T^1_{a,b}(u) = \inf\{n > S^1_{a,b}(u) \,|\, u_n \geq b\}$ puis par récurrence :

$$\forall i \in \mathbb{N}^*, \begin{cases} S^{i+1}_{a,b}(u) = \inf\{n > T^i_{a,b}(u) \,|\, u_n \leq a\} \\ T^{i+1}_{a,b}(u) = \inf\{n > S^{i+1}_{a,b}(u) \,|\, u_n \geq b\} \end{cases}$$

puis on considère pour $n \in \mathbb{N}$, $N^n_{a,b}(u) = \sup\{i \in \mathbb{N}^* \,|\, S^i_{a,b}(u) \leq n\}$ le nombre de montées avant n (où l'on définit une montée comme étant l'un des intervalles $([S^i, T^i])_{i\in\mathbb{N}^*}$).
On controle le nombre d'oscillations entre a à b en regardant $N^{\infty}_{a,b}(u) = \lim_{n\to+\infty} \uparrow N^n_{a,b}(u)$.

NOTE 9.38. Dans la suite, lorsque a et b seront fixés, on pourra omettre les indices a, b aux quantités définies ci-dessus.

LEMME 9.39. $\left(\forall a<b \text{ rationnels}, N_{a,b}^{\infty}(u)<+\infty\right) \iff \left(\exists \ell \in \mathbb{R}\cup\{\pm\infty\} \mid u_n \underset{n\to+\infty}{\longrightarrow} \ell\right)$

PREUVE

$\Longleftarrow$ Si $u_n \underset{n\to+\infty}{\longrightarrow} \ell$, on a $\ell \neq a$ ou $\ell \neq b$. Supposons par exemple $\ell \neq b$ et $b > \ell$.
Pour $n \in \mathbb{N}$ assez grand, $u_n < b$ donc pour certain $i_0 \in \mathbb{N}^*$, on a $\forall i \geq i_0, S_{a,b}^i(u) = +\infty$.
On en déduit $N_{a,b}^{\infty}(u) \leq i_0 < +\infty$.

$\Longrightarrow$ Par contraposée, si $u_n \underset{n\to+\infty}{\not\to} \ell$, alors u a au moins 2 points d'accumulation distincts ℓ_1 et ℓ_2.
Prenons $a, b \in \mathbb{Q}$ tels que $\ell_1 < a < b < \ell_2$. Alors $N_{a,b}^{\infty}(u) = +\infty$.

□

LEMME 9.40. [INÉGALITÉ DES MONTÉES DE DOOB]

Soient X une sous-martingale et $a < b$ fixés. Alors :

$$\forall n \in \mathbb{N}, \mathbb{E}[N_{a,b}^n(X)] \leq \frac{1}{b-a}\mathbb{E}[(X_n-a)_+ - (X_0-a)_+]$$

PREUVE Regardons la suite $(Y_n = (X_n - a)_+)_{n\in\mathbb{N}}$. C'est une sous-martingale par le lemme 9.37 puisque $x \longmapsto (x-a)_+$ est convexe et croissante.
Considérons alors, pour $n \in \mathbb{N}^*$, $H_n = \mathbb{1}_{\text{"}n \text{ dans une montée"}} = \sum_{i=1}^n \mathbb{1}_{S_{a,b}^i(X) < n \leq T_{a,b}^i(X)}$.
Par construction, la suite $Z = (Z_n)_{n\in\mathbb{N}}$ définie par

$$\forall i \in \mathbb{N}^*, \begin{cases} Z_{2i-1} = S_{a,b}^i(X) \\ Z_{2i} = T_{a,b}^i(X) \end{cases}$$

est une suite croissante de temps d'arrêt qui est même strictement croissante jusqu'à ce que l'un des termes ne vale éventuellement $+\infty$.

On a que $(H_n)_{n\in\mathbb{N}}$ est positif, borné (puisque pour tout $n \in \mathbb{N}$, $H_n \in \{0,1\}$) et prévisible. On calcule alors, à $n \in \mathbb{N}$ fixé :

$$\begin{aligned}(H.Y)_n &= \sum_{k=1}^{S^1} \underbrace{H_k}_{=0}(Y_k - Y_{k-1}) + \sum_{i=1}^{N^n-1}\left(\sum_{k=S^i+1}^{T^i} \underbrace{H_k}_{=1}(Y_k - Y_{k-1}) + \sum_{k=T^i+1}^{S^{i+1}} \underbrace{H_k}_{=0}(Y_k - Yk-1)\right) \\ &\qquad + \sum_{k=S^{N^n}+1}^{T^{N^n}} \underbrace{H_k}_{=1}(Y_k - Y_{k-1}) \quad + R_n \\ &= \sum_{i=1}^{N^n} \underbrace{Y_{T^i}}_{\geq (b-a)_+} - \underbrace{Y_{S^i}}_{0} \quad + R_n \\ &\geq N^n(b-a) + R_n\end{aligned}$$

où

$$R_n = \sum_{k=T^{N^n}+1}^{n} H_k(Y_k - Y_{k-1}) = \begin{cases} 0 & \text{si } S_{N^n+1} \geq n \\ \sum_{k=S^{N^n}+1}^{n} (Y_k - Y_{k-1}) = Y_n - Y_{S^{N^n}} \geq 0 & \text{si } S_{N^n+1} \leq n \end{cases}$$

puisque $T^{N^n+1} > n$. Ainsi $R_n \geq 0$.
Finalement, on a pour tout $n \in \mathbb{N}^*$, $(H.Y)_n \geq N^n(b-a)$.

De plus, $(1-H_n)_{n\in\mathbb{N}^*}$ est borné, positif et prévisible, donc $((1-H).Y)$ est une sous-martingale.
Ainsi, pour $n \in \mathbb{N}^*$, $\mathbb{E}[((1-H).Y)_n] \geq \mathbb{E}[((1-H).Y)_0] = 0$, puis $\mathbb{E}[(H.Y)_n] \geq (b-a)\mathbb{E}[N^n] = 0$.
Donc par bilinéarité de $X, Y \longmapsto (X.Y)$, on a $\mathbb{E}[(1.Y)_n] \geq (b-a)\mathbb{E}[N^n] = 0$.
Ce qui conclut puisque $\mathbb{E}[(1.Y)_n] = \mathbb{E}[Y_n - Y_0] = \mathbb{E}[(X_n - a)_+ - (X_0 - a)_+]$. □

THÉORÈME 9.41. *Soit X une martingale vérifiant $\sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+] < +\infty$.*

Alors il existe une variable aléatoire $X_\infty \in L^1$ telle que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$.

REMARQUE 9.42.

- Attention, on n'a pas forcément convergence L^1 !
- Une suite croissante bornée est convergente.
  Pour une sous-martingale, on a $\forall n \in \mathbb{N}, \mathbb{E}[M_{n+1} \mid \mathcal{F}_n] \geq M_n$, il suffit donc de la majorer pour la rendre convergente.
- $\forall k \in \mathbb{N}, \mathbb{E}[(X_k)_-] = \mathbb{E}[(X_k)_+] - \mathbb{E}[X_k] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+] - \mathbb{E}[X_0] \leq +\infty$.
  Donc $\forall k \in \mathbb{N}, \mathbb{E}[|X_k|] \leq 2\sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+] - \mathbb{E}[X_0]$.
  La condition du théorème est donc équivalente à $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < +\infty$.
- On a également le résultat : si X est une sur-martingale positive alors il existe une variable aléatoire $X_\infty \in L^1$ telle que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$.
  (X positive implique $\forall n \in \mathbb{N}, \mathbb{E}[(X_n)_-] = 0$).
- On en déduit plein d'autres critères de convergence pour les martingales : si M est une martingale *bornée dans L^1* ou *minorée* ou *positive* ou …, alors $M_n \xrightarrow[n\to+\infty]{\text{p.s.}} M_\infty \in L^1$.

PREUVE Soit X une martingale telle que $\sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+] < +\infty$.
Soient $a < b$ deux rationnels et $C = \sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+]$.
D'après l'inégalité des montées, on a pour $n \in \mathbb{N}$:

$$\mathbb{E}[N^n_{a,b}(X)] \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)_+ \underbrace{-(X_0 - a)_+}_{\leq 0}] \leq \frac{1}{b-a}\mathbb{E}[(X_n)_+] + |a|)$$

et on obtient par convergence monotone :

$$\mathbb{E}[N^\infty_{a,b}(X)] \leq \frac{C + |a|}{b-a}$$

Ainsi $N^\infty_{a,b}(X) < +\infty$ p.s., puis $\left(\forall a, b \in \mathbb{Q}, N^\infty_{a,b}(X) < +\infty\right)$ p.s. (en tant qu'union dénombrable d'événements presque sûrs).
Ainsi, il existe $X_\infty \in \mathbb{R} \cup \{\pm\infty\}$ telle que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$ et on a par le lemme de FATOU :

$$\mathbb{E}[|X_\infty|] = \mathbb{E}\Big[\liminf_{n\to+\infty} |X_n|\Big] \leq \liminf_{n\to+\infty} \mathbb{E}[|X_n|] < +\infty$$

Donc $X_\infty \in L^1$. En particulier, $X_\infty < +\infty$ p.s.. □

EXEMPLE 9.43. [PROCESSUS DE GALTON-WATSON]

En supposant les $(X_j^i)_{i,j\in\mathbb{N}^*}$ dans L^1 et en notant $m = \mathbb{E}[X_1] = \sum_{k\in\mathbb{N}} k\mu(\{k\}) < +\infty$, on a vu que $\left(Y_n = \frac{Z_n}{m^n}\right)_{n\in\mathbb{N}}$ est une martingale.
La propriété précédente donne qu'il existe $Y_\infty \in L^1$ telle que $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y_\infty$.

- Si $m < 1$: on a $m^{-n}Z_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y_\infty$ donc $Z_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0$.
  Puisque $(Z_n)_{n\in\mathbb{N}}$ est une suite d'entiers, on en déduit que $Z_n = 0$ p.s. pour n assez grand. Et donc on a extinction presque sûrement.
- Si $m = 1$: On a $\forall n \in \mathbb{N}, Y_n = Z_n$ et donc Z_n est stationnaire presque sûrement, ce qui n'est possible que si $Z_n = 0$ p.s. pour n assez grand ou $\mu = \delta_1$.
- Si $m > 1$: on a $m^{-n}Z_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y_\infty$.
  Ainsi, si $\mathbb{P}(Y_\infty > 0) > 0$, alors lorsque $(Y_\infty > 0)$, on a $Z_n \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty$.
  Ainsi il y a survie avec probabilité strictement positive.

  REMARQUE 9.44. On retrouve la propriété 8.33. Cependant, ici, rien ne permet de montrer que $\mathbb{P}(Y_\infty > 0) > 0$.

Le théorème précédent nous dit qu'il est très simple pour une martingale de converger presque sûrement vers une limite dans L^1 : il suffit de garantir la non-explosion :

- pour une sous-martingale ("croissant en moyenne"), il suffit qu'elle soit majorée ($\forall n, M_n \leq C$ p.s.) ou que $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty$ ou que $\sup_{n\in\mathbb{N}} \mathbb{E}[(M_n)_+] < +\infty$,
- pour une sur-martingale ("décroissant en moyenne"), il suffit qu'elle soit minorée ou qu'elle soit positive ou qu'elle soit bornée dans L^1,
- pour une martingale, n'importe laquelle de ces conditions suffit.

IV. C. Convergence L^1

Soit $M = (M_n)_{n\in\mathbb{N}}$ une martingale. On sait que si M converge dans L^1, alors M converge presque sûrement, mais la réciproque est fausse :

EXEMPLE 9.45. Considérons une marche aléatoire $(S_n^1)_{n\in\mathbb{N}}$ issue de 1.
Soit $T = \inf\{n \in \mathbb{N} \mid S_n^1 = 0\}$. On sait que $T < +\infty$ p.s.. Donc $S^1_{n\wedge T} \xrightarrow[n\to+\infty]{\text{p.s.}} S_T^1 = 0$.
Cependant, on n'a pas convergence dans L^1 puisque

$$\mathbb{E}[S^1_{n\wedge T_0}] = S^1_{0\wedge T_0} = 1 \underset{n\to+\infty}{\not\to} 0 = \mathbb{E}[S^1_{T_0}]$$

PREUVE [M CONVERGE DANS $L^1 \Longrightarrow M$ CONVERGE PRESQUE SÛREMENT]

On a $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] \leq \sup_{n\in\mathbb{N}}(\mathbb{E}[|M_n - M_\infty|] + \mathbb{E}[|M_\infty|]) < +\infty$.
Donc il existe Z_∞ telle que $M_n \xrightarrow[n\to+\infty]{\text{p.s.}} Z_\infty$. Ainsi, $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} Z_\infty$. Mais puisque $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} M_\infty$, on obtient par unicité de la limite que $Z_\infty = M_\infty$ p.s.. □

La notion fondamentale pour obtenir la convergence L^1 est l'*uniforme intégrabilité*.

RAPPEL 9.46. Si $X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X_\infty$:

$$(X_n)_{n\in\mathbb{N}} \text{ est U.I. } \iff X_n \xrightarrow[n\to+\infty]{L^1} X_\infty$$

LEMME 9.47. *Si $X \in L^1$ et $(\mathcal{B}_i)_{i\in I}$ est une famille de sous-tribus de $\mathcal{A}$, alors $(\mathbb{E}[X \mid \mathcal{B}_i])_{i\in I}$ est uniformément intégrable.*

PREUVE On sait que X est uniformément intégrable.
Si on fixe $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$(\star) \quad \forall A, \mathbb{P}(A) \leq \delta \Longrightarrow \mathbb{E}[|X|\,\mathbb{1}_A] \leq \varepsilon$$

Fixons ce δ. On a, par l'inégalité de MARKOV :

$$\mathbb{P}(|\mathbb{E}[X \mid \mathcal{B}_i]| > L) \leq \frac{\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}_i]|]}{L} \leq \frac{\mathbb{E}[\mathbb{E}[|X| \mid \mathcal{B}_i]]}{L} \leq \frac{\mathbb{E}[|X|]}{L}$$

Soit $L_0 = \frac{\mathbb{E}[|X|]}{\delta}$. Alors, pour tout $i \in I$ et $L \geq L_0$, on a $\mathbb{P}(|\mathbb{E}[X \mid \mathcal{B}_i]| > L) \leq \delta$. Calculons :

$$\begin{aligned}
\mathbb{E}[|\mathbb{E}[X \mid \mathcal{B}_i]|\,\mathbb{1}_{|\mathbb{E}[X \mid \mathcal{B}_i]|>L}] &\leq \mathbb{E}[\mathbb{E}[|X| \mid \mathcal{B}_i] \underbrace{\mathbb{1}_{|\mathbb{E}[X \mid \mathcal{B}_i]|>L}}_{\mathcal{B}_i\text{-mesurable borné}}] \\
&= \mathbb{E}[|X|\,\mathbb{1}_{|\mathbb{E}[X \mid \mathcal{B}_i]|>L}] && \text{par la propriété caractéristique de } \mathbb{E}[. \mid \mathcal{B}_i] \\
&\leq \varepsilon && \text{par } (\star)
\end{aligned}$$

□

THÉORÈME 9.48. *Soit $M = (M_n)_{n\in\mathbb{N}}$ une martingale. On a équivalence entre :*

(i) $(M_n)_{n\in\mathbb{N}}$ est uniformément intégrable.

(ii) il existe une variable aléatoire $M_\infty \in L^1$ telle que $M_n \xrightarrow[n\to+\infty]{\text{p.s. } | \, L^1} M_\infty$.

(iii) M est une martingale fermée : il existe $Z \in L^1$ telle que $\forall n \in \mathbb{N}, M_n = \mathbb{E}[Z \mid \mathcal{F}_n]$.

De plus, on a un lien entre Z et M_∞ :

- *Si $M_n \xrightarrow[n\to+\infty]{\text{p.s.}|L^1} M_\infty$ alors $\forall n \in \mathbb{N}, M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n]$: on peut donc choisir $Z = M_\infty$.*
- *Si $\forall n \in \mathbb{N}, M_n = \mathbb{E}[Z \mid \mathcal{F}_n]$, alors $M_n \xrightarrow[n\to+\infty]{\text{p.s.}|L^1} \mathbb{E}[Z \mid \mathcal{F}_\infty]$ avec $\mathcal{F}_\infty = \sigma(\bigcup_{n\in\mathbb{N}} \uparrow \mathcal{F}_n)$.*

PREUVE

(iii) $\Longrightarrow$ (ii) On applique le lemme précédent.

(i) $\Longrightarrow$ (ii) Si $(M_n)_{n\in\mathbb{N}}$ est uniformément intégrable, alors $\sup_{n\in\mathbb{N}}\mathbb{E}[|M_n|] < +\infty$, donc (par le théorème de convergence presque sûre des sous et sur-martingales) il existe $M_\infty \in L^1$ tel que $M_n \xrightarrow[n\to+\infty]{\text{p.s.}} M_\infty$. Et donc, en utilisant le rappel, on a $M_n \xrightarrow[n\to+\infty]{L^1} M_\infty$.

(ii) $\Longrightarrow$ (iii) Déjà vu : par continuité de $\mathbb{E}[.\mid\mathcal{F}_n]$ sur L^1, on a pour $n\in\mathbb{N}$:

$$M_n = \mathbb{E}[M_{n+k}\mid\mathcal{F}_n] \xrightarrow[k\to+\infty]{L^1} \mathbb{E}[M_\infty\mid\mathcal{F}_n]$$

Donc $M_n = \mathbb{E}[M_\infty\mid\mathcal{F}_n]$.

Reste à montrer le tout dernier point.
Si $\forall n\in\mathbb{N}, M_n = \mathbb{E}[Z\mid\mathcal{F}_n]$, alors $M_n \xrightarrow[n\to+\infty]{\text{p.s.}\mid L^1} M_\infty$ puisque (iii) $\Longrightarrow$ (ii). Montrons que $M_\infty = \mathbb{E}[Z\mid\mathcal{F}_\infty]$ p.s..
Soit $A\in\mathcal{F}_n$. Pour $k\geq n$:

$$\mathbb{E}[Z\underbrace{\mathbb{1}_A}_{\in\mathcal{F}_k}] = \mathbb{E}[\mathbb{E}[Z\mid\mathcal{F}_k]\mathbb{1}_A] = \mathbb{E}[M_k\mathbb{1}_A] \xrightarrow[k\to+\infty]{} \mathbb{E}[M_\infty\mathbb{1}_A]$$

puisque $|\mathbb{E}[M_\infty\mathbb{1}_A] - \mathbb{E}[M_k\mathbb{1}_A]| \leq \mathbb{E}[|M_\infty - M_k|] \xrightarrow[k\to+\infty]{} 0$ puisque $M_k \xrightarrow[k\to+\infty]{L^1} M_\infty$.
Donc pour tout $A\in\bigcup_{n\in\mathbb{N}}\mathcal{F}_n$, on a $\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[M_\infty\mathbb{1}_A]$.
Posons $\mathcal{C} = \bigcup_{n\in\mathbb{N}}\mathcal{F}_n$, qui est stable par intersection finie et $\mathcal{M} = \{A\mid\mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[M_\infty\mathbb{1}_A]\}$.
On vérifie que $\mathcal{M}$ est une classe monotone et on a vu que $\mathcal{C}\subset\mathcal{M}$, donc par le lemme des classes monotones : $\mathcal{F}_\infty = \mathcal{M}$. Ainsi :

$$\forall A\in\mathcal{F}_\infty, \mathbb{E}[Z\mathbb{1}_A] = \mathbb{E}[M_\infty\mathbb{1}_A]$$

Ainsi $M_\infty = \lim_{n\to+\infty} M_n$ p.s. est $\mathcal{F}_\infty$-mesurable, donc $\mathbb{E}[Z\mid\mathcal{F}_\infty] = M_\infty$. □

REMARQUE 9.49. M est bornée dans L^1 implique convergence presque sûre, mais pas forcément convergence L^1 (il faut l'uniforme intégrabilité en plus).
Mais pour $p>1$, M est bornée dans L^p implique convergence L^p.

a. Application : les urnes de POLYA

On considère deux populations de bactéries, des rouges et des bleues. Initialement, on a $R_0 = r$ bactéries rouges et $B_0 = b$ bactéries bleues où $r,b\in\mathbb{N}^*$. A l'étape $n\in\mathbb{N}$, une bactérie prise au hasard se divise en 2, disparaît et est remplacée par 2 bactéries de sa couleur.

On se donne $(X_n)_{n\in\mathbb{N}}$ variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{U}([0,1])$.
Soit $n\in\mathbb{N}$. R_n et B_n étant fixés, la probabilité que ce soit une bactérie rouge qui se divise à l'étape $n+1$ est $\frac{R_n}{R_n+B_n}$. On a $\mathbb{1}_{X_{n+1}\leq\frac{R_n}{R_n+B_n}} \sim \mathcal{B}(\frac{R_n}{R_n+B_n})$, qui vaut donc 1 avec probabilité $\frac{R_n}{R_n+B_n}$.
On pose ainsi $(R_0,B_0) = (r,b)$ et par récurrence :

$$\forall n\in\mathbb{N}, \begin{cases} R_{n+1} = R_n + \mathbb{1}_{X_{n+1}\leq\frac{R_n}{R_n+B_n}} \\ B_{n+1} = B_n + \mathbb{1}_{X_{n+1}>\frac{R_n}{R_n+B_n}} \end{cases}$$

Si $\mathcal{F}_n = \sigma(\{X_i \mid i \in [\![1,n]\!]\})$, alors R et B sont des processus adaptés à $\mathcal{F}$.
L'objectif est de comprendre l'évolution de la proportion $\left(Y_n = \frac{R_n}{R_n+B_n}\right)_{n\in\mathbb{N}}$ de bactéries rouges.

Soit $n \in \mathbb{N}$. Tout d'abord, on remarque que $R_{n+1} + B_{n+1} = R_n + B_n + 1$ (en sommant les deux égalités ci-dessus), et donc par récurrence :

$$R_n + B_n + 1 = r + b + n$$

Donc $Y_n = \frac{R_n}{r+b+n}$. Calculons :

$$\begin{aligned}
\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] &= \mathbb{E}\left[\frac{R_n + \mathbb{1}_{X_{n+1}\leq Y_n}}{r+b+n+1} \mid \mathcal{F}_n\right] = \frac{1}{r+b+n+1}(R_n + \mathbb{E}[\mathbb{1}_{X_{n+1}\leq Y_n} \mid \mathcal{F}_n]) \\
&= \frac{1}{r+b+n+1}\left(R_n + \int_0^1 \mathbb{1}_{x\leq Y_n}\, dx\right) = \frac{1}{r+b+n+1}(R_n + Y_n) \\
&= \frac{1}{r+b+n+1}\left(1 + \frac{1}{r+b+n}\right) R_n \\
&= Y_n
\end{aligned}$$

Donc $(Y_n)_{n\in\mathbb{N}}$ est une martingale.

RAPPEL 9.50. On a utilisé que si X est indépendante de $\mathcal{B}$ et Y est $\mathcal{B}$-mesurable, alors on a :

$$\mathbb{E}[\phi(X,Y) \mid \mathcal{B}] = \int \phi(x,Y)\, d\mu_X(x)$$

De plus, on a $0 \leq Y_n \leq 1$ p.s., donc $(Y_n)_{n\in\mathbb{N}}$ est uniformément intégrable. Ainsi $Y_n \overset{\text{p.s. | } L^1}{\underset{n\to+\infty}{\longrightarrow}} Y_\infty$.

- Regardons le cas particulier où $r = b = 1$ (on a initialement une bactérie de chaque couleur).
  On montre par récurrence que $\forall n \in \mathbb{N}, R_n \sim \mathcal{U}([\![1, n+1]\!])$:
  - C'est vrai pour $R_0 \sim S_1$.
  - On calcule :

$$\begin{aligned}
\mathbb{P}(R_n = k) &= \mathbb{P}(R_{n-1} = k-1 \cap \text{ "un rouge se divise"}) \\
&\quad + \mathbb{P}(R_{n-1} = k \cap \text{ "un bleu se divise"}) \\
&= \frac{1}{n}\frac{k-1}{1+n} + \frac{1}{n}\left(1 - \frac{k}{1+n}\right) \qquad \text{par hypothèse de récurrence} \\
&= \frac{1}{n+1}
\end{aligned}$$

  (le même calcul reste vrai pour $k = 1$ et $k = n+1$ car le terme absent est alors nul)

  Ainsi, pour $0 \leq t \leq 1$:

$$\mathbb{P}(Y_n \leq t) = \mathbb{P}(R_n \leq t(n+2)) = \frac{\lfloor t(n+2)\rfloor}{n+1} \underset{n\to+\infty}{\longrightarrow} t$$

Donc $F_{Y_\infty}(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t \geq 1 \end{cases}$. Ainsi $\boxed{Y_\infty \sim \mathcal{U}([0,1])}$.

– Cas général : si r, b sont des entiers non nuls quelconques, on peut montrer que :

$$\boxed{Y_n \underset{n\to+\infty}{\overset{\text{p.s.}\,|\,L^1}{\longrightarrow}} Y_\infty \sim c_{r,b}x^{r-1}(1-x)^{b-1}\mathbb{1}_{0\leq x\leq 1}\,dx}$$

où $c_{r,b}$ est une constante de normalisation.

REMARQUE 9.51. On a une convergence presque sûre vers une variable aléatoire de loi qui n'est pas un dirac.

Si $T : \Omega \longrightarrow \mathbb{N}\cup\{+\infty\}$ est un temps d'arrêt (avec éventuellement $\mathbb{P}(T=+\infty)>0$) et X est un processus tel que $X_n \underset{n\to+\infty}{\overset{\text{p.s.}}{\longrightarrow}} X_\infty$, on définit $X_T = \sum_{n\in\mathbb{N}} X_n\mathbb{1}_{T=n} + X_\infty\mathbb{1}_{T=+\infty}$.

THÉORÈME 9.52. *Si M est une martingale uniformément intégrable et T est un temps d'arrêt, on sait que $M_n \underset{n\to+\infty}{\overset{\text{p.s.}\,|\,L^1}{\longrightarrow}} M_\infty$ et $M_n = \mathbb{E}[M_\infty \,|\, \mathcal{F}_n]$. Alors on a $\boxed{M_T = \mathbb{E}[M_\infty \,|\, \mathcal{F}_T]}$. En particulier :*

$$\forall n\in\mathbb{N}, \mathbb{E}[M_T] = \mathbb{E}[M_n]$$

REMARQUE 9.53. Pour les problèmes de ruine du joueur, $(S^x_{n\wedge T_{\{0,L\}}})_{n\in\mathbb{N}}$ est bornée (car dans $[0,L]$) donc uniformément intégrable. On peut donc appliquer ce résultat et on obtient :

$$\mathbb{E}[S^x_{T_{\{0,L\}}}] = \mathbb{E}[S^x_0]$$

PREUVE Montrons que M_T est L^1 :

$$\begin{aligned}
\mathbb{E}[|M_T|] &= \sum_{n\in\mathbb{N}} \mathbb{E}[|M_n|\,\mathbb{1}_{T=n}] + \mathbb{E}[|M_\infty|\,\mathbb{1}_{T=\infty}] \\
&\leq \sum_{n\in\mathbb{N}} \mathbb{E}[\mathbb{E}[|M_\infty|\,|\,\mathcal{F}_n]\ \underbrace{\mathbb{1}_{T=n}}_{\mathcal{F}_n\text{-mesurable}}\ ] + \mathbb{E}[|M_\infty|\,\mathbb{1}_{T=\infty}] \quad \text{puisque } |M_n| = |\mathbb{E}[M_\infty\,|\,\mathcal{F}_n]| \leq \mathbb{E}[|M_\infty|\,|\,\mathcal{F}_n] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[|M_\infty|\,\mathbb{1}_{T=n}] + \mathbb{E}[|M_\infty|\,\mathbb{1}_{T=\infty}] \\
&= \mathbb{E}[|M_\infty|] < +\infty
\end{aligned}$$

Donc $M_T \in L^1$. Il vient alors, pour $A\in\mathcal{F}_T$:

$$\begin{aligned}
\mathbb{E}[M_T\mathbb{1}_A] &= \sum_{n\in\mathbb{N}} \mathbb{E}[\underbrace{M_T}_{=M_n}\mathbb{1}_A\mathbb{1}_{T=n}] + \mathbb{E}[M_T\mathbb{1}_A\mathbb{1}_{T=+\infty}] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[\mathbb{E}[M_\infty\,|\,\mathcal{F}_n]\ \underbrace{\mathbb{1}_{A\cap(T=n)}}_{\mathcal{F}_n\text{-mesurable}}] + \mathbb{E}[M_\infty\mathbb{1}_A\mathbb{1}_{T=+\infty}] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[M_\infty\mathbb{1}_A\mathbb{1}_{T=n}] + \mathbb{E}[M_\infty\mathbb{1}_A\mathbb{1}_{T=+\infty}] \\
&= \mathbb{E}[M_\infty\mathbb{1}_A]
\end{aligned}$$

M_T est $\mathcal{F}_T$-mesurable, donc $M_T = \mathbb{E}[M_\infty\,|\,\mathcal{F}_T]$ et ainsi :

$$\forall n\in\mathbb{N}, \mathbb{E}[M_T] = \mathbb{E}[M_\infty] = \mathbb{E}[M_0] = \mathbb{E}[M_n]$$

□

IV. D. Convergence L^p

On souhaite contrôler la trajectoire des martingales à partir de la loi en un temps donné. C'est le point délicat de ce paragraphe.
Dans L^1, on a vu l'inégalité des montées. Ici, on va utiliser l'inégalité maximale :

PROPOSITION 9.54. [INÉGALITÉ MAXIMALE]

Soit X une sous-martingale. On a pour $t > 0$:

$$t\mathbb{P}\Big(\sup_{0\leq k\leq n} X_k > t\Big) \leq \mathbb{E}[X_n \mathbb{1}_{\sup_{0\leq k\leq n} X_k > t}] \leq \mathbb{E}[(X_n)_+]$$

PREUVE

- Montrons que s'il existe $C < +\infty$ et deux temps d'arrêt S et T tels que $S \leq T < C$ p.s., alors $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.
  Posons, pour $n \in \mathbb{N}$, $H_n = \mathbb{1}_{S<n\leq T}$. H est alors prévisible, borné et positif, et donc $(H.X)$ est une sous-martingale. Ainsi $(\mathbb{E}[(H.X)_n])_{n\in\mathbb{N}}$ est croissante d'où :

$$0 = \mathbb{E}[(H.X)_0] \leq \mathbb{E}[(H.X)_C] = \mathbb{E}\Big[\sum_{n=S+1}^{T} (X_n - X_{n-1})\Big] = \mathbb{E}[X_T - X_S]$$

- Posons $T = \inf(\{n \in \mathbb{N} \mid X_n > t\})$.
  $T \wedge n \leq n$ et n sont 2 temps d'arrêt bornés, donc $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n]$, ou encore :

$$\mathbb{E}[X_T \mathbb{1}_{T\leq n}] + \mathbb{E}[X_n \mathbb{1}_{T>n}] \leq \mathbb{E}[X_n]$$

  Et puisque $\{T \leq n\} = \left\{\sup_{0\leq k\leq n} X_k > t\right\}$, on a :

$$t\mathbb{P}\Big(\sup_{0\leq k\leq n} X_k > t\Big) \leq \mathbb{E}[X_T \mathbb{1}_{T\leq n}] \leq \mathbb{E}[X_n] - \mathbb{E}[X_n \mathbb{1}_{T>n}] \leq \mathbb{E}[X_n \mathbb{1}_{T\leq n}]$$

□

PROPOSITION 9.55. *Soient $p > 1$ et X une sous-martingale positive bornée dans L^p :*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] < +\infty$$

Posons $\tilde{X}_n = \sup_{0\leq k\leq n} |X_k|$. Alors :

$$\forall n \in \mathbb{N}, \mathbb{E}[|\tilde{X}_n|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p]$$

PREUVE Fixons $n \in \mathbb{N}$. On sait que $(\star)$ $\quad t\mathbb{P}(\tilde{X}_n > t) \leq \mathbb{E}[X_n \mathbb{1}_{\tilde{X}_n > t}]$.
On a $\mathbb{E}[X^p] = p\int_0^{+\infty} t^{p-1}\mathbb{P}(X > t)\, dt$ (déjà vu, utiliser FUBINI).
En multipliant $(\star)$ par t^{p-2} et en intégrant sur $\mathbb{R}^+$, il vient :

$$\int_0^{+\infty} t^{p-1}\mathbb{P}(\tilde{X}_n > t)\, dt \leq \int_0^{+\infty} \mathbb{E}[X_n t^{p-2} \mathbb{1}_{\tilde{X}_n > t}]\, dt$$

Or, d'une part $\int_0^{+\infty} t^{p-1}\mathbb{P}(\tilde{X}_n > t)\,dt = \frac{\mathbb{E}[\tilde{X}_n^p]}{p}$ et d'autre part :

$$\int_0^{+\infty} \mathbb{E}[X_n t^{p-2}\mathbb{1}_{\tilde{X}_n>t}]\,dt = \mathbb{E}\Big[X_n \int_0^{+\infty} t^{p-2}\mathbb{1}_{\tilde{X}_n>t}\Big] = \mathbb{E}\Big[X_n \frac{(\tilde{X}_n)^{p-1}}{p-1}\Big]$$
$$\leq \frac{1}{p-1}\mathbb{E}[(X_n)^p]^{1/p}\mathbb{E}[(\tilde{X}_n)^{(p-1)\frac{p}{p-1}}]^{\frac{p-1}{p}} \qquad \text{par l'inégalité de HÖLDER}$$

Ainsi $\mathbb{E}[\tilde{X}_n^p] \leq \frac{p}{p-1}\mathbb{E}[(X_n)^p]^{1/p}\mathbb{E}[(\tilde{X}_n)^p]^{1-1/p}$, ou encore :

$$\mathbb{E}[\tilde{X}_n^p]^{1/p} \leq \frac{p}{p-1}\mathbb{E}[(X_n)^p]^{1/p}$$

□

THÉORÈME 9.56. *Soient $p > 1$ et X est une martingale bornée dans L^p ($\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] < +\infty$). Posons $\tilde{X}_n = \sup_{0\leq k\leq n} |X_k|$ et $\tilde{X}_\infty = \sup_{k\in\mathbb{N}} |X_k|$.*

Alors $\tilde{X}_\infty \in L^p$ et il existe X_∞ tel que $X_n \xrightarrow[n\to+\infty]{\text{p.s. } | L^p} X_\infty$ et

$$\begin{cases} \forall n \in \mathbb{N}, \mathbb{E}[|\tilde{X}_n|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p] \\ \mathbb{E}[|\tilde{X}_\infty|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_\infty|^p] \end{cases}$$

PREUVE

- $(X_n)_{n\in\mathbb{N}}$ est bornée dans L^1 donc il existe $X_\infty \in L^1$ tel que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$. On a par le lemme de FATOU :
$$\mathbb{E}[|X_\infty|^p] = \mathbb{E}\Big[\liminf_{n\to+\infty} |X_\infty|^p\Big] \leq \liminf_{n\to+\infty} \mathbb{E}[|X_\infty|^p] < +\infty$$
Donc $X_\infty \in L^p$.
- $x \longmapsto |x|$ est convexe donc $(|X_n|)_{n\in\mathbb{N}}$ est une sous-martingale positive, donc le résultat précédent s'applique :
$$\forall n \in \mathbb{N}, \mathbb{E}[|\tilde{X}_n|^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|X_n|^p]$$
On a $\tilde{X}_\infty = \lim_{n\to+\infty} \uparrow \tilde{X}_n = \lim_{n\to+\infty} \uparrow \sup_{0\leq k\leq n} |X_k|$, donc par convergence monotone :
$$\mathbb{E}[|\tilde{X}_\infty|^p] \leq \Big(\frac{p}{p-1}\Big)^p \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] < +\infty$$
Ainsi $\tilde{X}_\infty \in L^p$.
On a $(\forall n \in \mathbb{N}, \tilde{X}_\infty = \sup_{k\in\mathbb{N}} |X_k| \geq X_n)$ p.s. et donc $\tilde{X}_\infty \geq X_\infty$ p.s..
Ainsi, par convergence dominée : $\mathbb{E}[|X_n - X_\infty|^p] \xrightarrow[n\to+\infty]{} 0$ puisque $|X_n - X_\infty| \leq 2\tilde{X}_\infty \in L^p$ et $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$.
Puis $x \longmapsto |x|^p$ est convexe donc $(|X_n|^p)_{n\in\mathbb{N}}$ est une sous-martingale positive, donc $(\mathbb{E}[|X_n|^p])_{n\in\mathbb{N}}$ est croissante, d'où :
$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] = \lim_{n\in\mathbb{N}} \mathbb{E}[|X_n|^p] = \mathbb{E}[|X_\infty|^p]$$
puisque $X_n \xrightarrow[n\to+\infty]{L^p} X_\infty$. Et donc $\mathbb{E}[|\tilde{X}_\infty|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_\infty|^p]$.

□

a. Application au processus de GALTON-WATSON

Supposons $m = \mathbb{E}[X_1^1] = \sum_{k\in\mathbb{N}} k\mu(\{k\}) < +\infty$. Dans l'exemple 9.43, on a vu que $(Y_n = \frac{Z_n}{m^n})_{n\in\mathbb{N}}$ est une martingale qui tend presque sûrement vers une variable aléatoire Y_∞. On en avait déduit que si $m \leq 1$ et $\mu \neq \delta_1$, alors on a extinction presque sûrement, c'est-à-dire ($Z_n = 0$ pour n assez grand) p.s..
Si $m > 1$, on aimerait montrer que l'on a survie avec probabilité strictement positive. On a vu qu'il faudrait pour cela savoir que $\mathbb{P}(Y_\infty > 0) > 0$. Mais cela n'est pas toujours vrai.
Faisons l'hypothèse que $\underline{\sum_{k\in\mathbb{N}} k^2\mu(\{k\}) < +\infty}$ (ou de manière équivalente $\mathbb{E}[(X_1^1)^2] < +\infty$).

Posons $\sigma^2 = \mathrm{Var}(X_1^1)$. On calcule, pour $n \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{E}[Z_n^2] &= \mathbb{E}\Big[\sum_{k=0}^{+\infty} (X_1^n + \cdots + X_k^n)^2 \mathbb{1}_{Z_{n-1}=k}\Big] \\
&= \sum_{k=0}^{+\infty} \mathbb{E}\Big[\sum_{j=1}^{k} (X_j^n)^2 + \sum_{1\leq j\neq \ell\leq k} X_j^n X_\ell^n\Big]\mathbb{P}(Z_{n-1} = k) \\
&= \sum_{k=0}^{+\infty} (k\mathbb{E}[(X_i^n)^2] + k(k-1)m^2)\mathbb{P}(Z_{n-1} = k) \\
&= \sum_{k=0}^{+\infty} k\sigma^2\mathbb{P}(Z_{n-1} = k) + \sum_{k=0}^{+\infty} k^2m^2\mathbb{P}(Z_{n-1} = k) \\
&= \mathbb{E}[Z_{n-1}]\sigma^2 + m^2\mathbb{E}[(Z_{n-1})^2] \\
&= m^{n-1}\sigma^2 + m^2\mathbb{E}[(Z_{n-1})^2]
\end{aligned}$$

Par récurrence, on obtient que $\forall n \in \mathbb{N}, Z_n \in L^2$. Puis :

$$\begin{aligned}
\mathbb{E}[Y_n^2] &= \frac{1}{m^{2n}}\mathbb{E}[Z_n^2] = \frac{\sigma^2}{m^{n+1}} + \mathbb{E}[(Y_{n-1})^2] \\
&= \sigma^2\left(\frac{1}{m^{n+1}} + \frac{1}{m^n} + \cdots + \frac{1}{m^2}\right) = \sigma^2\frac{\frac{1}{m^2} - \frac{1}{m^{n+2}}}{1 - \frac{1}{m}} = \sigma^2\frac{1 - \frac{1}{m^n}}{m-1} \\
&\xrightarrow[n\to+\infty]{} \sigma^2\frac{1}{m-1}
\end{aligned}$$

Donc $(Y_n)_{n\in\mathbb{N}}$ est bornée dans L^2, donc $Y_n \xrightarrow[n\to+\infty]{L^2} Y_\infty$, donc $Y_n \xrightarrow[n\to+\infty]{L^1} Y_\infty$.
Ainsi $\mathbb{E}[Y_\infty] = \mathbb{E}[Y_0] = 1$ donc $\boxed{\mathbb{P}(Y_\infty > 0) > 0}$.
Sur cet événement, on a $Z_n \sim m^n Y_\infty$.
Donc avec probabilité strictement positive, on a explosion exponentielle de la population à vitesse m^n.

REMARQUE 9.57. On peut montrer qu'il suffit que $\sum_{k\in\mathbb{N}} k\ln(k)\mu(\{k\}) < +\infty$ pour avoir $\mathbb{P}(Y_\infty > 0) > 0$.

CHAPITRE 10

Chaînes de MARKOV

Une filtration canonique $(\mathcal{F}_n = \sigma(\{X_i \mid 0 \leq i \leq n\}))_{n\in\mathbb{N}}$ structure le temps : à un instant "présent" n, $\mathcal{F}_n$ est l'ensemble des événements du passé et X_{n+1} est dans le futur. L'idée des chaînes de MARKOV est de considérer des processus pour lesquels le futur ne dépend du passé qu'au travers du présent.

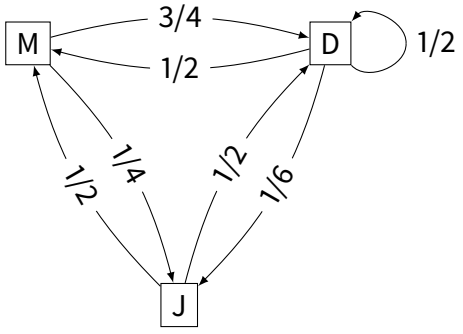
I. Exemples et définition

EXEMPLE 10.1. [LE HAMSTER]

On veut modéliser la vie d'un hamster. Chaque heure, il est dans l'un des trois états : il mange (M), il dort (D) ou il joue (J) (il fait de la roue).
Selon l'activité à une heure n fixée, le hamster change d'activité à l'heure $n+1$ selon les probabilités suivantes, et ce **de manière indépendantes des activités précédentes** :

- après avoir mangé, il dort avec probabilité 3/4 et joue avec probabilité 1/4.
- après avoir dormi, il mange avec probabilité 1/2, continue de dormir avec probabilité 1/3 et joue avec probabilité 1/6.
- après avoir joué, il mange avec probabilité 1/2 et dort avec probabilité 1/2.

On note X_n l'état du hamster à la n-ème heure. On a $X_n \in \{\mathsf{M}, \mathsf{D}, \mathsf{J}\}$.
Alors on verra que $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de MARKOV.
On peut représenter par un *graphe orienté* les différentes probabilités de transition entre les états :



De manière équivalente, toutes les données sont codées dans la matrice $Q \in \mathcal{M}_3([0,1])$ de coefficient général Q_{ij} la probabilité de passer d'un état i à un état j, appelée *noyau ou matrice de transition*. Si M est l'état 1, D l'état 2 et J l'état 3, on a :

$$Q = \begin{pmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

On peut alors se poser plusieurs questions qui vont se résoudre en partie par du calcul matriciel. Par exemple, on peut essayer de trouver la proportion de temps passé à manger, ce qui revient à calculer $\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i=\mathsf{M}}$.

DÉFINITION 10.2. [NOYAU DE TRANSITION]

Soit E un ensemble fini ou infini dénombrable qu'on appelle *espace d'état*. Q est un *noyau de transition* sur E (ou une *matrice de transition*) si $Q : E \times E \longrightarrow [0,1]$ est tel que

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

REMARQUE 10.3.

- Si E est fini, on identifie E à $[\![1,n]\!]$, et alors un noyau de transition sur E s'identifie à la matrice $(Q(i,j))_{1\leq i,j\leq n}$. C'est une matrice *stochastique* : ces coefficients sont dans $[0,1]$ et chaque ligne a pour somme 1.
- $Q(x,.)$ est une suite de nombres positifs de somme 1 qui s'identifie à une mesure de probabilité. On veut que ce soit la loi de X_{n+1} lorsque $X_n = x$.

DÉFINITION 10.4. [CHAÎNE DE MARKOV]

Soient Q un noyau de transition sur l'espace d'état E (fini ou infini dénombrable) et μ_0 une mesure de proba sur E.
On dit que $X = (X_n)_{n\in\mathbb{N}} \in E^{\mathbb{N}}$ est la *chaîne de* MARKOV de mesure initiale μ_0 et de noyau de transition Q si $X_0 \sim \mu_0$ et pour tout $n \in \mathbb{N}$ puis pour tout $(x_i)_{0\leq i\leq n}$ et y éléments de E tels que $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \neq 0$, on a :

$$\boxed{\mathbb{P}(X_{n+1} = y \,|\, X_0 = x_0, \ldots, X_n = x_n) = Q(x_n, y)}$$

NOTE 10.5. Afin de simplifier les écritures, on notera $x_{0:n} = (x_0, \ldots, x_n)$. La condition s'écrit donc

$$\mathbb{P}(X_{n+1} = y \,|\, X_{0:n} = x_{0:n}) = Q(x_n, y)$$

REMARQUE 10.6. Les martingales sont des processus à valeurs dans $\mathbb{R}$ alors que les chaînes de MARKOV sont à valeurs dans un espace discret (graphe).

NOTE 10.7. On a en fait défini ici les chaîne de MARKOV *homogènes*. En toute généralité, on demande $\mathbb{P}(X_{n+1} = y \,|\, X_{0:n} = x_{0:n}) = Q_n(x_n, y)$ où $(Q_n)_{n\in\mathbb{N}}$ est une famille de noyaux de transitions.

On supposera toujours nos chaînes homogènes.

EXEMPLE 10.8. Soit $(S_n^x)_{n\in\mathbb{N}}$ une marche aléatoire issue de x dans $\mathbb{R}^d$. Alors pour tout $n \in \mathbb{N}$ puis pour tout $(x_i)_{0\leq i\leq n}$ et y éléments de $\mathbb{Z}^d$ tels que $\mathbb{P}(X_{0:n} = x_{0:n}) \neq 0$, on a :

$$\begin{aligned}\mathbb{P}(S_{n+1}^x = y \mid S_{0:n}^x = x_{0:n}) &= \mathbb{P}(S_n^x + X_{n+1} = y \mid S_{0:n}^x = x_{0:n}) \\ &= \mathbb{P}(x_n + X_{n+1} = y \mid S_{0:n}^x = x_{0:n}) \\ &= \mathbb{P}(X_{n+1} = y - x_n) \qquad \text{par indépendance} \\ &= \mu^{\perp}(\{y - x_n\})\end{aligned}$$

Donc S^x est une chaîne de MARKOV de mesure initiale δ_x et de noyau de transition $Q(x, y) = \mu(\{y - x\})$.

Il faut bien faire attention à la définition. On peut avoir l'impression que certains processus sont des chaînes de MARKOV alors que ce n'est pas le cas :

EXEMPLE 10.9. Regardons $(S_n^0)_{n\in\mathbb{N}}$ la marche aléatoire simple issue de 0 (on rappelle que les $(X_n)_{n\in\mathbb{N}^*}$ suivent des lois de RADEMACHER et sont indépendants).

- On se donne Z indépendant de S de loi de RADEMACHER et on pose pour $n \in \mathbb{N}$, $Y_n = Z\,|S_n^0|$. $(Y_n)_{n\in\mathbb{N}}$ est positive avec probabilité 1/2, négative avec probabilité 1/2, et vérifie :

  $$\forall x \in \mathbb{Z}, \mathbb{P}(Y_{n+1} = x \pm 1 \mid Y_n = x) = 1/2$$

  Cependant, ce n'est pas une chaîne de MARKOV, puisque :

  $$\begin{cases} \mathbb{P}(Y_3 = 1 \mid (Y_0, Y_1, Y_2) = (0, 1, 0)) = 1 \\ \mathbb{P}(Y_3 = 1 \mid (Y_0, Y_1, Y_2) = (0, -1, 0)) = 0 \end{cases}$$

- Soit $V_n = \sup_{0\leq k\leq n} S_k^0$. On peut montrer que ce n'est pas une chaîne de MARKOV (exercice).

EXEMPLE 10.10. [PROCESSUS DE GALTON-WATSON]

En reprenant les notations des chapitres précédents, on a :

$$\begin{aligned}\mathbb{P}(Z_{n+1} = y \mid Z_{0:n} = x_{0:n}) &= \mathbb{P}(X_1^{n+1} + \cdots + X_{x_n}^{n+1} = y \mid Z_{0:n} = x_{0:n}) \\ &= \mathbb{P}(X_1^{n+1} + \cdots + X_{x_n}^{n+1} = y) \qquad \text{par indépendance} \\ &= \mu^{*x_n}(\{y\})\end{aligned}$$

EXEMPLE 10.11. [MARCHE ALÉATOIRE SUR UN GRAPHE]

Soit $G = (S, A)$ un graphe, où S est un ensemble non vide de sommets et A un ensemble d'arrêtes orientées $(x, y, p) \in S \times S \times [0, 1]$ tel que, à x fixé, $\sum_{y\,|\,(x,y,p)\in A} p = 1$, alors la marche aléatoire sur ce graphe est la chaîne de MARKOV de noyau de transition

$$Q(x, y) = \begin{cases} p & \text{si } (x, y, p) \in A \\ 0 & \text{sinon} \end{cases}$$

En fait, toute chaîne de MARKOV se représente par un graphe pour lequel S est l'espace d'état. C'est une propriété d'existence que l'on montrera plus tard.

II. Propriétés des chaînes de Markov

On veut une propriété qui permette de remplacer "ne dépend du passé qu'au travers du présent".

Proposition 10.12. *Soit E un espace d'état et Q un noyau de transition. On a équivalence entre :*

(i) X *est une chaîne de* Markov *de noyau de transition* Q.

(ii) X *vérifie :*

$$\forall n \in \mathbb{N}, \forall p \in \mathbb{N}^*, \forall (y_i)_{1\leq i\leq p} \in E^p, \forall x_n \in E, \forall A \in \mathcal{F}_n^X, \mathbb{P}(X_n = x_n, A) \neq 0$$
$$\implies \mathbb{P}(X_{n+1=n+p} = y_{1:p} \,|\, X_n = x_n, A) = Q(x_n, y_1)Q(y_1, y_2)\dots Q(y_{p-1}, y_p)$$

(iii) $\forall n \in \mathbb{N}, \forall (x_i)_{0\leq i\leq n} \in E^{n+1}, \mathbb{P}(X_{0:n} = x_{0:n}) = \mathbb{P}(X_0 = x_0)Q(x_0, x_1)\dots Q(x_{n-1}, x_n)$

Preuve

(i) $\implies$ (ii) On procède par récurrence sur p.
Si $p = 1$, on a $A \in \sigma(\{X_i \,|\, 0 \leq i \leq n\})$, donc il existe $\Gamma \subset E^{n+1}$ tel que $A = \{X_{0:n} \in \Gamma\}$. On a donc, lorsque $\mathbb{P}(X_n = x_n, A) \neq 0$:

$$\begin{aligned}
\mathbb{P}(X_{n+1} = y_1, X_n = x_n, A) &= \sum_{\substack{z_0,\dots,z_n \\ z_{0:n}\in\Gamma}} \mathbb{P}(X_{n+1} = y_1, X_n = x_n, X_{0:n} = z_{0:n}) \\
&= \sum_{\substack{z_0,\dots,z_{n-1} \\ (z_{0:n-1},x_n)\in\Gamma}} \mathbb{P}(X_{n+1} = y_1 \,|\, X_n = x_n, X_{0:n-1} = z_{0:n-1})\mathbb{P}(X_n = x_n, X_{0:n-1} = z_{0:n-1}) \\
&= Q(x_n, y) \sum_{\substack{z_0,\dots,z_{n-1} \\ (z_{0:n-1},x_n)\in\Gamma}} \mathbb{P}(X_n = x_n, (X_0, \dots, X_{n-1}) = (z_0, \dots, z_{n-1})) \\
&= Q(x_n, y)\mathbb{P}(X_n = x_n, A)
\end{aligned}$$

Et donc on obtient

$$\mathbb{P}(X_{n+1} = y_1 \,|\, X_n = x_n, A) = \frac{1}{\mathbb{P}(X_n = x_n, A)}\mathbb{P}(X_{n+1} = y_1, X_n = x_n, A) = Q(x_n, y)$$

La propriété est donc vraie au rang 1.

Si par hypothèse de récurrence, la propriété est vraie au rang p, on a lorsque $\mathbb{P}(X_n = x_n, A) \neq 0$:

$$\begin{aligned}
\mathbb{P}(X_{n+1:n+p+1} = y_{1:p+1} \,|\, X_n = x_n, A) = \mathbb{P}(X_{n+p+1} = y_{p+1} \,|\, X_{n+1:n+p} = y_{1:p}, X_n = x_n, A) \\
\times \mathbb{P}(X_{n+1:n+p} = y_{1:p} \,|\, X_n = x_n, A)
\end{aligned}$$

(sauf si $(X_{n+1:n+p} = y_{1:p})$ est incompatible avec $(X_n = x_n, A)$, auquel cas l'égalité ci-dessus est triviale puisque les deux membres sont nuls).
Or, on a (la première égalité provenant du résultat au rang 1 et la seconde de l'hypothèse de récurrence) :

$$\begin{cases}
\mathbb{P}(X_{n+p+1} = y_{p+1} \,|\, X_{n+1:n+p} = y_{1:p}, X_n = x_n, A) = Q(y_p, y_{p+1}) \\
\mathbb{P}(X_{n+1:n+p} = y_{1:p} \,|\, X_n = x_n, A) = Q(x_n, y_1)Q(y_1, y_2)\dots Q(y_{p-1}, y_p)
\end{cases}$$

On obtient donc finalement :

$$\mathbb{P}(X_{n+1:n+p+1} = y_{1:p+1} \mid X_n = x_n, A) = Q(x_n, y_1)Q(y_1, y_2)\dots Q(y_p, y_{p+1})$$

Et donc le résultat est vrai au rang $p+1$.

D'où le résultat par récurrence.

(ii) $\Longrightarrow$ (iii) On applique en prenant $A = (X_0 = x_0)$.

(iii) $\Longrightarrow$ (i) On a lorsque $\mathbb{P}(X_{0:n} = x_{0:n}) \neq 0$:

$$\begin{aligned}\mathbb{P}(X_{n+1} = y \mid X_{0:n} = x_{0:n}) &= \frac{\mathbb{P}(X_{0:n+1} = (x_{0:n}, y))}{\mathbb{P}(X_{0:n} = x_{0:n})} \\ &= \frac{\mathbb{P}(X_0 = x_0)Q(x_0, X_1)\dots Q(x_{n-1}, x_n)Q(x_n, y)}{\mathbb{P}(X_0 = x_0)Q(x_0, X_1)\dots Q(x_{n-1}, x_n)} \\ &= Q(x_n, y)\end{aligned}$$

□

REMARQUE 10.13. La loi d'une chaîne de MARKOV X est déterminée par la mesure initiale μ_0 et la matrice de transition Q. La loi de X est la probabilité sur la tribu cylindrique $\mathcal{C}$. Elle est déterminée par ses valeurs $\mathbb{P}((X_0, \dots, X_n) \in A)$ avec $A \subset E^{n+1}$. On a :

$$\mathbb{P}(X_{0:n} \in A) = \sum_{x_{0:n} \in A} \mu_0(\{x_0\})Q(x_0, x_1)\dots Q(x_{n-1}, x_n)$$

On interprète donc Q comme un ensemble de probabilités conditionnelles :

$$\boxed{\forall n \in \mathbb{N}, \forall x, y \in E, Q(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x)}$$

On se demande comment réaliser plusieurs pas d'un coup ?

DÉFINITION 10.14. Si P et Q sont deux noyaux de transition sur E, on définit le noyau :

$$\begin{array}{rccc} P.Q : & E \times E & \longrightarrow & [0,1] \\ & x, z & \longmapsto & \sum_{y \in E} P(x, y)Q(y, z) \end{array}$$

C'est un noyau de transition.

En effet, on a :

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

et il en découle que $(P.Q) \in [0, 1]$.

NOTE 10.15. $P, Q \longmapsto P.Q$ est associative.

REMARQUE 10.16. Si E est fini, on identifie la matrice de transition avec de vraies matrices. Cette opération correspond alors au produit matriciel.

DÉFINITION 10.17. Si Q est un noyau de transition, on note par récurrence :

$$\begin{cases} Q^0(x,y) = \mathbb{1}_{x=y} \\ Q^1 = Q \\ \forall n \in \mathbb{N}^*, Q^{n+1} = Q.Q^n \end{cases}$$

PROPOSITION 10.18. *Si X est une chaîne de* MARKOV *de noyau de transition Q, on a pour tout $n, p \in \mathbb{N}$ et tout $A \in \mathcal{F}_n^X$:*

$$\mathbb{P}(X_{n+p} = y \,|\, X_n = x_n, A) = Q^p(x_n, y)$$

Une conséquence de cette proposition est que $(X_{pn})_{n\in\mathbb{N}}$ est une chaîne de MARKOV *de noyau de transition Q^p.* On sait donc comment réaliser p pas d'un coup : Q^p est le noyau de transition de p pas de la chaîne.

PREUVE On a :

$$\begin{aligned} \mathbb{P}(X_{n+p} = y \,|\, X_n = x_n, A) &= \sum_{y_1,\dots,y_{p-1}\in E} \mathbb{P}(X_{n+1:n+p} = (y_{1:p-1}, y) \,|\, X_n = x_n, A) \\ &= \sum_{y_1,\dots,y_{p-1}\in E} Q(x,y_1)Q(y_1,y_2)\dots Q(y_{p-1}) = Q^p(x_n, y) \end{aligned}$$

□

EXEMPLE 10.19. [RETOUR DANS LA CAGE DU HAMSTER]

Appliquons ce résultat pour trouver la probabilité qu'au bout de sept heures, le hamster dorme (2), en supposant qu'initialement, le hamster mange (1). On a :

$$\mathbb{P}(X_7 = 2 \,|\, X_0 = 1) = (Q^7)_{12}$$

Q est explicite donc on se ramène à du calcul matriciel.

REMARQUE 10.20. Pour de grandes puissances on aura intérêt à diagonaliser la matrice.

On peut aussi chercher la probabilité que jusqu'à la 7-ème heure, le hamster ne dorme pas (toujours en supposant $X_0 = \text{M}$) ?

$$\mathbb{P}(\forall 1 \le n \le 7, X_n \neq 2 | X_0 = 1) = \mathbb{P}((X_{1:7}) = (3,1,3,1,3,1,3) \,|\, X_0 = 1) = Q(1,2)^4 Q(2,1)^3 = \frac{1}{2048}$$

II. A. Résolution canonique des chaînes de MARKOV

On se pose la question de l'existence de chaînes de MARKOV *à μ_0 et Q fixés.*

THÉORÈME 10.21. [RÉSOLUTION CANONIQUE DES CHAÎNES DE MARKOV]

Soit Q un noyau de transition sur E. Soient $\Omega_0 = E^{\mathbb{N}}$ et $\mathcal{C}$ la tribu cylindrique.

Il existe une famille de probabilités $(\mathbb{P}_x)_{x\in E}$ telle que si $X : \Omega_0 \longrightarrow \Omega_0$ est l'identité, alors la loi de X sur $(\Omega, \mathcal{C}, \mathbb{P}_x)$ est celle d'une chaîne de MARKOV *de transition Q et de mesure initiale δ_x.*

PREUVE Soient $(Z_y^n)_{n\in\mathbb{N},y\in E}$ des variables aléatoires indépendantes telles que pour tout entier $n \in \mathbb{N}$, on a $\forall y \in E, Z_y^n = Q(y,.)$ (et donc $Z_y^n \in E$ p.s.).

On définit $X_0 = x$ et pour $n \geq 1$, $X_n = Z_{X_{n-1}}^n$. Alors $X_0 \sim \delta_x$ et on a :

$$\begin{aligned}\mathbb{P}(X_{n+1} = x_{n+1} \mid X_{0:n} = x_{0:n}) &= \mathbb{P}(Z_{X_n}^{n+1} = x_{n+1} \mid X_{0:n} = x_{0:n}) \\ &= \mathbb{P}(Z_{x_n}^{n+1} = x_{n+1} \mid X_{0:n} = x_{0:n}) \\ &= Q(x_n, x_{n+1}) \qquad \text{par indépendance}\end{aligned}$$

Définissons $\mathbb{P}_x$ la loi de X. Alors l'identité sur $(\Omega, \mathcal{C}, \mu_X = \mathbb{P}_x)$ a même loi que X. D'où le résultat. □

Pour résumer, si on se donne une transition Q, on a construit pour $x \in E$ une probabilité $\mathbb{P}_x$ sur $(\Omega, \mathcal{A}) = (E^{\mathbb{N}}, \mathcal{C})$ telle que si X est l'identité sur Ω, alors X est la chaîne de MARKOV de transition Q et telle que $X_0 \sim x$. On peut reformuler cela en disant que $X \sim \mathbb{P}_x$ équivaut à X est une chaîne de MARKOV de transition Q et telle que $X_0 = x$ p.s.. Changer le point de départ équivaut à changer la probabilité $\mathbb{P}_x$ que l'on choisie.

II. B. Propriétés de MARKOV simple et forte

DÉFINITION 10.22. Si μ est une mesure de probabilité sur E, on définit :

$$\mathbb{P}_\mu = \sum_{x\in E} \mu(\{x\})\mathbb{P}_x$$

C'est une mesure de probabilité sur $E^{\mathbb{N}}$.

REMARQUE 10.23.

- Si X est l'identité sur $(E^{\mathbb{N}}, \mathcal{C}, \mathbb{P}_\mu)$, alors X est une chaîne de MARKOV de transition Q et telle que $X_0 \sim \mu$. En effet, on a :

$$\begin{aligned}\mathbb{P}_\mu(X_{0:n} = x_{0:n}) &= \sum_{x\in E} \mu(\{x\})\, \mathbb{P}_x(X_{0:n} = x_{0:n}) \\ &= \mu(\{x_0\})\, \mathbb{P}_{x_0}(X_{0:n} = x_{0:n}) \\ &= \mu(\{x_0\})Q(x_0, x_1)\dots Q(x_{n-1}, x_n)\end{aligned}$$

- Jusqu'à présent, on ne travaillait qu'avec une seule mesure de probabilité. Désormais, lorsque l'on travaille avec la réalisation canonique, on travaille avec un espace de probabilité fixé $(E^{\mathbb{N}}, \mathcal{C}, \mathbb{P}_\mu)$ où $\mathbb{P}_\mu$ est une famille de mesure de probabilité.
  On notera alors $\mathbb{E}_x$ l'espérance sous $\mathbb{P}_x$ et $\mathbb{E}_\mu$ l'espérance sous $\mathbb{P}_\mu$.
  Utiliser $\mathbb{P}_\mu$ permet de gagner en souplesse puisqu'on peut "changer le point de départ".
- Si on prouve pour un certain événement A que pour tout μ, on a $X \in A$ $\mathbb{P}_\mu$-p.s., alors on a prouvé que pour toute chaîne de MARKOV Y de transition Q, on a $Y \in A$ p.s. (découle de la propriété de résolution canonique).

L'intérêt de la réalisation canonique est de pouvoir renforcer la propriété que le futur ne dépend du passé qu'au travers du présent. On va voir deux énoncés importantes : les propriétés de MARKOV simple et forte.

DÉFINITION 10.24. [OPÉRATEUR "SHIFT"]

On définit l'opérateur "shift" :

$$\begin{array}{rccc} \Theta : & E^{\mathbb{N}} & \longrightarrow & E^{\mathbb{N}} \\ & (x_i)_{i\in\mathbb{N}} & \longmapsto & (x_{i+1})_{i\in\mathbb{N}} \end{array}$$

On a en particulier $\Theta^n(x_i)_{i\in\mathbb{N}} = (x_{i+n})_{i\in\mathbb{N}}$. Si $X = (X_i)_{i\in\mathbb{N}}$ est un processus, on voit $\Theta^n X$ comme le futur.

PROPOSITION 10.25. [PROPRIÉTÉ DE MARKOV SIMPLE]

Soient Q une transition sur E sous la réalisation canonique et $\mathcal{C}$ la tribu cylindrique. Alors pour tout $n \in \mathbb{N}$, $B \in \mathcal{C}$, $A \in \mathcal{F}_n^X$, $x \in E$ et toute mesure de probabilité μ sur E, les propriétés équivalentes suivantes sont vérifiées :

(i) $\mathbb{P}_\mu(X_n = x, A) \neq 0 \Longrightarrow \mathbb{P}_\mu(\Theta^n X \in B \,|\, X_n = x, A) = \mathbb{P}_x(X \in B)$,

(ii) $\mathbb{P}_\mu(\Theta^n X \in B, X_n = x, A) = \mathbb{P}_x(X \in B)\mathbb{P}_\mu(X_n = x, A)$,

(iii) Pour toute f mesurable bornée :

$$\mathbb{E}_\mu[f(\Theta^n X) \,|\, \mathcal{F}_n^X]\mathbb{1}_{X_n=x} = \mathbb{E}_x[f(X)]\mathbb{1}_{X_n=x} \quad \mathbb{P}_\mu\text{- p.s.}$$

(iv) Pour toute f mesurable bornée, on a en posant $\mathbb{E}_{X_n}[f(X)] = \sum_{x\in E} \mathbb{E}_x[f(X)]\mathbb{1}_{X_n=x}$:

$$\mathbb{E}_\mu[f(\Theta^n X) \,|\, \mathcal{F}_n^X] = \mathbb{E}_{X_n}[f(X)] \quad \mathbb{P}_\mu\text{- p.s.}$$

PREUVE

(i) On veut identifier la loi de $\Theta^n X : (E^{\mathbb{N}}, \mathcal{C}, \mathbb{P}_\mu(.\,|\,A, X_n = x)) \longrightarrow E^{\mathbb{N}}$ et de $X : (E^{\mathbb{N}}, \mathcal{C}, \mathbb{P}_x) \longrightarrow E^{\mathbb{N}}$.
Il suffit donc de les identifier sur une classe génératrice stable par intersection finie.

Pour $p \in \mathbb{N}$ et $b_0, \ldots, b_p \in E$, on pose $B_{b_{0:p}} = \{(x_i)_{i\in\mathbb{N}} \,|\, x_{0:p} = b_{0:p}\}$.
Alors $\left\{B_{b_{0:p}} \,|\, p \in \mathbb{N}, b_0, \ldots, b_p \in E\right\} \cup \{\varnothing\}$ est une classe stable par intersection finie qui engendre la tribu. On a :

$$\mathbb{P}_\mu(X_{n+1:n+p} = b_{0:p} \,|\, X_n = x, A) = \mathbb{1}_{b_0=x} Q(b_0, b_1) \ldots Q(b_{p-1}, b_p) = \mathbb{P}_x(X_{0:p} = b_{0:p})$$

(ii) $\Longrightarrow$ (iii) On sait que $\mathbb{E}_x[f(X)]$ est $\mathcal{F}_n$-mesurable.
Il suffit d'avoir le résultat pour les fonctions indicatrices (on conclut alors classiquement par linéarité pour les fonctions étagées puis par convergence monotone pour toute f).

Soit donc $f = \mathbb{1}_B$. Pour $A \in \mathcal{F}_n^X$:

$$\begin{aligned}\mathbb{E}_\mu[\mathbb{1}_{\theta^n X \in B}\mathbb{1}_A] &= \sum_{x\in E} \mathbb{P}_\mu(\theta^n X \in B, A, X_n = x)\\ &= \sum_{x\in E} \mathbb{P}_x(X \in B)\mathbb{P}_\mu(A, X_n = x) \qquad \text{par (ii)}\\ &= \mathbb{E}_\mu\Big[\underbrace{\sum_{x\in E} \mathbb{P}_x(X \in B)\mathbb{1}_{X_n=x}}_{\mathcal{F}_n\text{-mesurable}}\mathbb{1}_A\Big]\end{aligned}$$

Donc

$$\mathbb{E}_\mu[f(\theta^n X) \,|\, \mathcal{F}_n^X] = \sum_{x\in E} \mathbb{E}_x[f(X)]\mathbb{1}_{X_n=x} \quad \mathbb{P}_\mu\text{- p.s.}$$

En multipliant des deux côtés par $\mathbb{1}_{X_n=x}$, on obtient le résultat.

□

PROPOSITION 10.26. [PROPRIÉTÉ DE MARKOV FORTE]

Soient Q une transition sur E, T un temps d'arrêt (par rapport à $\mathcal{F}^X$), $A \in \mathcal{F}_T$ et B quelconque. Alors pour tout $x \in E$, les propriétés équivalentes suivantes sont vérifiées :

(i) $\mathbb{P}_\mu(T < +\infty, X_T = x, A) \neq 0 \Longrightarrow \mathbb{P}_\mu(\Theta^T X \in B \,|\, T < +\infty, X_T = x, A) = \mathbb{P}_x(X \in B)$,

(ii) $\mathbb{P}_\mu(\Theta^T X \in B, T < +\infty, X_T = x, A) = \mathbb{P}_x(X \in B)\mathbb{P}_\mu(T < +\infty, X_T = x, A)$,

(iii) Pour toute f mesurable bornée :

$$\mathbb{E}_\mu[f(\Theta^T X) \,|\, \mathcal{F}_T]\mathbb{1}_{T<+\infty, X_T=x} = \mathbb{E}_x[f(X)]\mathbb{1}_{T<+\infty, X_T=x} \quad \mathbb{P}_\mu\text{- p.s.}$$

REMARQUE 10.27. Dans la pratique, c'est la propriété (ii) qui est la plus utile.

PREUVE

(i) On a :

$$\begin{aligned}\mathbb{P}_\mu(\Theta^T X \in B, T < +\infty, X_T = x, A) &= \mathbb{P}_\mu(\Theta^T X \in B, \sqcup_{n\in\mathbb{N}^*}(T = n), X_T = x, A)\\ &= \sum_{n\in\mathbb{N}^*} \mathbb{P}_\mu(\Theta^T X \in B, T = n, X_T = x, A)\\ &= \sum_{n\in\mathbb{N}^*} \mathbb{P}_\mu(\Theta^n X \in B, X_n = x, \underbrace{(T = n) \cap A}_{\in\mathcal{F}_n^X})\\ &= \sum_{n\in\mathbb{N}^*} \mathbb{P}_x(X \in B)\mathbb{P}_\mu(X_n = x, T = n, A) \qquad \text{par MARKOV simple}\\ &= \mathbb{P}_x(X \in B)\mathbb{P}_\mu(X_T = x, T < +\infty, A)\end{aligned}$$

(i) $\Longrightarrow$ (iii) On calcule :

$$\begin{aligned}\mathbb{E}_\mu[f(\Theta^T X)\mathbb{1}_{T<+\infty, X_T=x}\mathbb{1}_A] &= \sum_{n\in\mathbb{N}^*}\mathbb{E}_\mu[f\theta^n X)\mathbb{1}_{T=n,A}\mathbb{1}_{X_n=x}]\\ &= \sum_{n\in\mathbb{N}^*}\mathbb{E}_\mu[\mathbb{E}_\mu[f\theta^n X)\,|\,\mathcal{F}_n]\mathbb{1}_{T=n,A}\mathbb{1}_{X_n=x}]\\ &= \mathbb{E}_x[f(X)]\sum_{n\in\mathbb{N}^*}\mathbb{E}_\mu[\mathbb{1}_{T=n,A}\mathbb{1}_{X_n=x}]\\ &= \mathbb{E}_x[f(X)]\mathbb{E}_\mu[\mathbb{1}_{T<+\infty}\mathbb{1}_A\mathbb{1}_{X_T=x}]\end{aligned}$$

Donc

$$\mathbb{E}_\mu[f(\Theta^T X)\underbrace{\mathbb{1}_{T<+\infty,X_T=x}}_{\mathcal{F}_T\text{-mesurable}}\,|\,\mathcal{F}_T] = \mathbb{E}_x[f(X)]\mathbb{1}_{T<+\infty}\mathbb{1}_{X_T=x}\quad \mathbb{P}_\mu\text{- p.s.}$$

□

II. C. Nombre de passages en un point

On veut savoir si une chaîne de MARKOV X *passe souvent en un point ou non.*

Notons $\tilde{T}_x$ le premier retour en x et N_x le nombre de passages en x, c'est-à-dire :

$$\tilde{T}_x = \inf(\{n\in\mathbb{N}^*\,|\,X_n=x\})\quad\text{et}\quad N_x = \mathrm{card}(\{n\in\mathbb{N}\,|\,X_n=x\}) = \sum_{n\in\mathbb{N}}\mathbb{1}_{X_n=x}$$

REMARQUE 10.28. $\tilde{T}_x$ est différent de l'indice de la première atteinte de x lorsque $X_0=x$!

RAPPEL 10.29. On a vu que la marche aléatoire simple sur Z passe une infinité de fois sur chaque site, c'est-à-dire :

$$\forall x,y\in\mathbb{Z}, N_y=+\infty \text{ et } \tilde{T}_y<+\infty\quad \mathbb{P}_x\text{- p.s.}$$

PROPOSITION 10.30. *Soient Q un noyau de transition sur E et $x\in E$. On a deux possibilités :*

- *soit $\tilde{T}_x<+\infty$ $\mathbb{P}_x$-p.s., alors $N_x=+\infty$ $\mathbb{P}_x$-p.s..*
  On finit toujours par revenir en x. On dit que x est récurrent.
- *soit $\mathbb{P}_x(\tilde{T}_x=+\infty)>0$, alors $N_x<+\infty$ $\mathbb{P}_x$-p.s..*
  On dit que x est transitoire *ou* transient.
  De plus, on a sous $\mathbb{P}_x$ que $N_x\sim\mathcal{G}(\mathbb{P}_x(\tilde{T}_x=+\infty))$. En particulier :

$$\mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(\tilde{T}_x=+\infty)} < +\infty$$

PREUVE Les deux conditions s'excluent. On remarque de plus que $X_0=x$ $\mathbb{P}_x$-p.s. donc $N_x\geq 1$ $\mathbb{P}_x$-p.s..

Soit $k\in\mathbb{N}$. On a :

$$\mathbb{P}_x(N_x\geq k+1) = \mathbb{P}_x(N_x\geq k+1,\tilde{T}_x<+\infty) + \underbrace{\mathbb{P}_x(N_x\geq k+1,\tilde{T}_x=+\infty)}_{=0}$$

On a $N_x = n_x(X)$ où $n_x((y_i)_{i\in\mathbb{N}}) = \sum_{n\in\mathbb{N}} \mathbb{1}_{y_n=x}$, d'où :

$$\begin{aligned}\mathbb{P}_x(N_x \geq k+1) &= \mathbb{P}_x(n_x(X) \geq k+1, \tilde{T}_x < +\infty, X_{\tilde{T}_x} = x)\\ &= \mathbb{P}_x(1 + n_x(\Theta^{\tilde{T}_x} X) \geq k+1, \tilde{T}_x < +\infty, X_{\tilde{T}_x} = x)\\ &= \mathbb{P}_x(n_x(\Theta^{\tilde{T}_x} X) \geq k, \tilde{T}_x < +\infty, X_{\tilde{T}_x} = x)\\ &= \mathbb{P}_x(n_x(X) \geq k)\mathbb{P}_x(\tilde{T}_x < +\infty, X_{\tilde{T}_x} = x) \qquad \text{par MARKOV fort}\end{aligned}$$

Donc $\mathbb{P}_x(N_x \geq k+1) = \mathbb{P}_x(N_x \geq k)\mathbb{P}_x(\tilde{X}_x < +\infty)$, et puisque $\mathbb{P}_x(N_x \geq 1) = 1$:

$$\mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(\tilde{X}_x < +\infty)^{k-1}$$

- Si $\tilde{T}_x < +\infty$ $\mathbb{P}_x$-p.s., alors $\mathbb{P}_x(N_x \geq k) = 1$ donc $N_x = +\infty$ $\mathbb{P}_x$-p.s..
- Si $\mathbb{P}_x(\tilde{T}_x < +\infty) < 1$:

$$\begin{aligned}\mathbb{P}_x(N_x = k) &= \mathbb{P}_x(N_x \geq k) - \mathbb{P}_x(N_x \geq k+1)\\ &= \mathbb{P}_x(\tilde{X}_x < +\infty)^{k-1}\mathbb{P}_x(\tilde{X}_x = +\infty)\\ &= (1 - \mathbb{P}_x(\tilde{X}_x = +\infty))^{k-1}\mathbb{P}_x(\tilde{X}_x = +\infty)\end{aligned}$$

□

REMARQUE 10.31. On a prouvé que

$$\forall k \in \mathbb{N}, \mathbb{P}_x(N_x = k) = \mathbb{P}_x(\tilde{X}_x < +\infty)^{k-1}\mathbb{P}_x(\tilde{X}_x = +\infty)$$

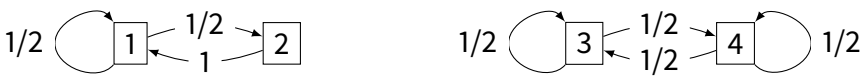
Cela signifie que la probabilité d'avoir k passages en x est la probabilité de revenir $k-1$ fois en x puis de s'échapper.

La propriété de MARKOV permet ainsi de découper dans le temps en événements indépendants !

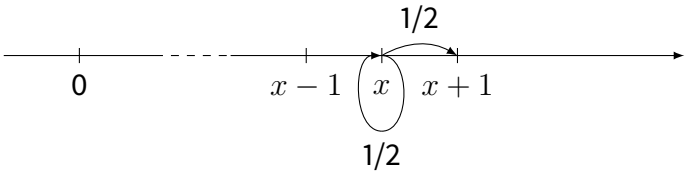
EXEMPLE 10.32. [DES CAS PEU INTÉRESSANTS]

- On considère une chaîne de MARKOV de noyau de transition :

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1/2 & 1/2\\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$



  L'étude se ramène à deux chaînes de MARKOV disjointes simples ...
- Sur $\mathbb{N}$, on passe de x à $x+1$ avec probabilité 1/2, sinon on reste en x.On ne va jamais vers le bas.



On veut une notion qui permette de déterminer si deux sites peuvent communiquer entre eux.

DÉFINITION 10.33. [POTENTIEL DE LA MARCHE]

Soit Q un noyau de transition sur E. On définit le *potentiel* de la marche $U : E \times E \longrightarrow E$ par :

$$\forall x, y \in E, U(x,y) = \mathbb{E}_x[N_y] = \mathbb{E}_x\Big[\sum_{n\in\mathbb{N}} \mathbb{1}_{X_n=y}\Big] = \sum_{n\in\mathbb{N}} \mathbb{P}_x(X_n = y) = \sum_{n\in\mathbb{N}} Q^n(x,y)$$

On a alors les équivalences, pour $x \neq y \in E$:

$$\mathbb{P}_x(\tilde{T}_y < +\infty) > 0 \iff \mathbb{P}_x(\exists n \in \mathbb{N} \,|\, X_n = y) > 0 \iff \exists n \in \mathbb{N}, \mathbb{P}_x(X_n = y) \neq 0$$
$$\iff \exists n \in \mathbb{N}, Q^n(x,y) > 0 \iff U(x,y) \neq 0$$

Lorsque ces propositions sont vraies, on note $x \rightsquigarrow y$ ("x communique avec y" : dans la représentation graphique, il existe un chemin orienté de x vers y). On note $\leftrightsquigarrow$ la relation d'équivalence engendrée. On dit que Q est *irréductible* s'il y a une seule classe pour cette relation.

PROPOSITION 10.34. *On a x récurrent $\iff U(x,x) = +\infty$.*

PREUVE

- Si x est récurrent, on a $N_x = +\infty$ $\mathbb{P}_x$-p.s., donc $U(x,x) = \mathbb{E}_x[N_x] = +\infty$.
- Si x est transitoire, on a $U(x,x) = \mathbb{E}_x[N_x] = \frac{1}{\mathbb{P}_x(\tilde{T}_x = +\infty)} < +\infty$.

□

EXEMPLE 10.35. [MARCHE ALÉATOIRE SIMPLE SUR $\mathbb{Z}^d$]

$\underline{d=1}$ Soit S^x la marche aléatoire simple issue de x. On a vu que $S^x = (S^x_n)_{n\in\mathbb{N}}$ est une chaîne de MARKOV de transition définie par :

$$\forall i, j \in \mathbb{Z}, \begin{cases} Q(i,i+1) = Q(i,i-1) = \frac{1}{2} \\ Q(i,j) = 0 \end{cases} \quad \text{si } j \notin \{i-1, i+1\}$$

On remarque que :

$$U(y,y) = \sum_{n\in\mathbb{N}} \mathbb{P}_y(S^y_n = y) = \sum_{n\in\mathbb{N}} \mathbb{P}\Big(\sum_{k=1}^{n} X_k = 0\Big) = \sum_{n\in\mathbb{N}} \mathbb{P}(S^0_{2n} = 0)$$

(si n est impair, on ne peut avoir $\sum_{k=1}^n X_k = 0$).
Or :

$$\mathbb{P}(S^0_{2n}) = \mathbb{P}(\mathrm{card}(\{1 \leq i \leq 2n \,|\, X_i = 1\}) = n) = \binom{2n}{n}\frac{1}{2^{2n}}$$
$$= \frac{(2n)!}{(n!)^2}\frac{1}{4^n} \underset{\text{STIRLING}}{\sim} \frac{1}{\sqrt{\pi n}}$$

Donc $U(y,y) = \sum_{n\in\mathbb{N}} \frac{1}{\sqrt{\pi n}}(1-o(1)) = +\infty$ donc **tous les sites sont récurrents.**

$\underline{d=2}$ Désormais, on prend $x \in \mathbb{Z}^2$ et on suppose les $(X_k)_{k\in\mathbb{N}^*}$ i.i.d. de loi $\mathcal{U}(\{(1,0),(0,1),(-1,0),(0,-1)\})$.

Si ϕ est la composition d'une rotation d'angle $\pi/4$ puis d'une dilatation par $\sqrt{2}$ (centrée en x), alors en posant pour $k\in\mathbb{N}$, $\tilde{X}_k = \phi(X_k)$ et $\tilde{S} = \phi(S)$, on a :

$$\tilde{X}_k \sim \mathcal{U}(\{(1,1),(1,-1),(-1,1),(-1,-1)\}) = \Big(\frac{1}{2}(\delta_1+\delta_{-1})\Big)\otimes\Big(\frac{1}{2}(\delta_1+\delta_{-1})\Big)$$

Décomposons alors $\tilde{S}^0 = (\tilde{S}^{0,1},\tilde{S}^{0,2})$ selon ses deux composantes.
On a $\tilde{S}^{0,\ell} = \sum_{k=1}^n \tilde{X}_k^\ell$ où $\tilde{X}_k = (\tilde{X}_k^1,\tilde{X}_k^2)$. Alors $\tilde{S}^{0,1}$ et $\tilde{S}^{0,2}$ sont indépendantes et leur loi est la marche aléatoire simple sur $\mathbb{Z}$. Il vient :

$$\begin{aligned} U(y,y) &= \sum_{n\in\mathbb{N}} \mathbb{P}(S_n^0=0) = \sum_{n\in\mathbb{N}} \mathbb{P}(S_{2n}^0=0) = \sum_{n\in\mathbb{N}} \mathbb{P}(\tilde{S}_{2n}^0) \\ &= \sum_{n\in\mathbb{N}} \mathbb{P}(\tilde{S}_{2n}^{0,1}=0)\mathbb{P}(\tilde{S}_{2n}^{0,2}=0) \\ &= \sum_{n\in\mathbb{N}} \Big(\frac{1}{\sqrt{\pi n}}(1+o(1))\Big)^2 \quad\Longleftrightarrow\quad \sum_{n\in\mathbb{N}}\frac{1}{n} = +\infty \end{aligned}$$

Donc tout site est récurrent pour la marche aléatoire simple dans $\mathbb{Z}^2$.

$\underline{d\geq 3}$ On suppose désormais $x\in\mathbb{R}^d$ et les $(\tilde{X}_k)_{k\in\mathbb{N}^*}$ i.i.d. de loi $\mathcal{U}(\{(x_1,\dots,x_d)\mid \forall i\in[\![1,d]\!], x_i\in\{-1,1\}\})$.
Posons pour $n\in\mathbb{N}$, $\tilde{S}_n^x = x + \sum_{k=1}^n \tilde{X}_k$. $\tilde{S}^x$ est la marche aléatoire le long des diagonales (attention, ce n'est pas la marche aléatoire simple).

Ecrivons $\tilde{S}^x = (\tilde{S}^{x,1},\dots,\tilde{S}^{x,d})$. Les coordonnées sont indépendantes et sont des marches aléatoires simples sur $\mathbb{Z}$. Alors :

$$U(y,y) = \sum_{n\in\mathbb{N}} \mathbb{P}(\tilde{S}_{2n}^0) = \sum_{n\in\mathbb{N}} \prod_{\ell=1}^d \mathbb{P}(\tilde{S}_{2n}^{0,\ell}) = \sum_{n\in\mathbb{N}} \Big(\frac{1}{\sqrt{\pi n}}(1+o(1))\Big)^d < +\infty$$

Donc tout site est transitoire.

REMARQUE 10.36. Attention, cela ne permet pas de savoir si c'est le cas pour la marche aléatoire simple sur $\mathbb{Z}^d$.

Revenons à la marche aléatoire simple sur $\mathbb{Z}^d$ avec d quelconque.

Posons $(e_i = (\delta_{i,j})_{1\leq j\leq d})_{1\leq i\leq d}$ la base canonique de $\mathbb{Z}^d$.
On considère S la marche aléatoire simple sur $\mathbb{Z}^d$ issue de 0 (c'est-à-dire les $(X_k)_{k\in\mathbb{N}}$ sont i.i.d. de loi $\mathcal{U}(\{e_1,-e_1,\dots,e_d,-e_d\})$. On a :

$$U(0,0) = \sum_{n\in\mathbb{N}} \mathbb{P}(S_n=0) = \sum_{n\in\mathbb{N}} \mathbb{P}(S_{2n}=0)$$

Calculons $\phi_{S_{2n}}$ pour $n\in\mathbb{N}$. On a :

$$\phi_{X_1}(\lambda) = \mathbb{E}[\mathrm{e}^{i\langle\lambda|X_1\rangle}] = \frac{1}{2d}\sum_{\ell=1}^d \mathrm{e}^{i\langle\lambda|e_\ell\rangle} + \mathrm{e}^{i\langle\lambda|-e_\ell\rangle} = \frac{1}{d}\sum_{\ell=1}^d \cos(\lambda_\ell)$$

Donc :

$$\phi_{S_{2n}}(\lambda) = \left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^{2n}$$

Or, on a $\phi_{S_{2n}}(\lambda) = \mathbb{E}[\mathrm{e}^{i\langle\lambda|S_{2n}\rangle}] = \sum_{k\in\mathbb{Z}^d}\mathrm{e}^{i\langle\lambda|k\rangle}\,\mathbb{P}(S_{2n}=k)$, d'où (la série est absolument convergente) :

$$\int_{\lambda\in[-\pi,\pi]^d}\phi_{S_{2n}}(\lambda)\,d\lambda = (2\pi)^d\mathbb{P}(S_{2n}=0)$$

Donc

$$\begin{aligned}U(0,0) &= \sum_{n\in\mathbb{N}}\mathbb{P}(S_{2n}=0) = \frac{1}{(2\pi)^d}\sum_{n\in\mathbb{N}}\int_{\lambda\in[-\pi,\pi]^d}\phi_{S_{2n}}(\lambda)\,d\lambda\\ &= \frac{1}{(2\pi)^d}\sum_{n\in\mathbb{N}}\int_{\lambda\in[-\pi,\pi]^d}\Bigg(\underbrace{\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)}_{\in[-1,1]\text{ et }\neq\pm1\text{ p.p.}}\Bigg)^{2n}d\lambda\\ &= \frac{1}{(2\pi)^d}\int_{\lambda\in[-\pi,\pi]^d}\frac{1}{1-\left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^2}\,d\lambda \qquad \text{par le théorème de FUBINI positif}\end{aligned}$$

Problème : le terme intégré tend vers $+\infty$ lorsque les λ_i tendent vers 0 ou vers π. Découpons alors :

$$U(0,0) = \frac{1}{(2\pi)^d}\left(\int_{]-\varepsilon,\varepsilon[^d}\frac{1}{1-\left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^2}\,d\lambda + \int_{(\pi+]-\varepsilon,\varepsilon[)^d}\frac{1}{1-\left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^2}\,d\lambda\right) + I$$

avec $I < +\infty$.
On a alors $U(0,0) < +\infty \iff \frac{1}{(2\pi)^d}\int_{]-\varepsilon,\varepsilon[^d}\frac{1}{1-\left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^2}\,d\lambda < +\infty$.

Or, en utilisant le développement limité $\cos(x) \underset{x\to0}{=} 1-\frac{x^2}{2}+o(x^2)$:

$$1-\left(\frac{1}{d}\sum_{\ell=1}^{d}\cos(\lambda_\ell)\right)^2 \underset{\varepsilon\to0}{=} 1-\left(1-\frac{1}{d}\sum_{\ell=1}^{d}\frac{\lambda_\ell^2}{2}+o(\varepsilon^2)\right)^2 \underset{\varepsilon\to0}{=} \frac{\lambda_1^2+\cdots+\lambda_d^2}{d}+o(\varepsilon^2)$$

Donc $U(0,0) < +\infty \iff \int_{]-\varepsilon,\varepsilon[^d}\frac{1}{\|\lambda\|_2^2}\,d\lambda < +\infty \iff \int_{\|\lambda\|<\varepsilon}\frac{1}{\|\lambda\|_2^2}\,d\lambda < +\infty$.

RAPPEL 10.37. Pour f mesurable positive, on a la formule de changement de coordonnées polaire dans $\mathbb{R}^d$:

$$\int_{\mathbb{R}^n} f(\|x\|)\,dx = \underbrace{n\,\mathrm{Vol}(B(0,1))}_{c_n}\int_{r\geq0} f(r)r^{n-1}\,dr$$

D'où

$$U(0,0) < +\infty \iff c_n\int_{r=0}^{\varepsilon}\frac{1}{r^2}r^{d-1}\,dr < +\infty \iff \int_0^{\varepsilon} r^{d-3}\,dr < +\infty$$

Et donc $\boxed{U(0,0) < +\infty \iff d \geq 3}$

PROPOSITION 10.38.

- *On a :*
$$\left.\begin{array}{l} x \text{ récurrent} \\ x \rightsquigarrow y \end{array}\right\} \implies \left\{\begin{array}{l} y \text{ récurrent} \\ y \rightsquigarrow x \end{array}\right. \quad \text{et} \quad N_y = +\infty \ \mathbb{P}_x\text{-p.s.}$$
- *De même :*
$$\left.\begin{array}{l} y \text{ transitoire} \\ x \rightsquigarrow y \end{array}\right\} \implies x \text{ transitoire}$$
et y transitoire $\implies \forall z, N_y < +\infty \ \mathbb{P}_z$-p.s..
- *Etre transitoire ou récurrent est une propriété de classe : si $x \leftrightsquigarrow y$ alors x récurrent $\iff y$ récurrent.*

PREUVE

- $x \rightsquigarrow y$ donc il existe n tel que $\mathbb{P}_x(X_n = y) > 0$. Comme $\{X_n = y\} \subset \left\{\tilde{T}_y < +\infty\right\}$, on a donc $0 < \mathbb{P}_x(\tilde{T}_y < +\infty)$. y est donc récurrent et $N_y = +\infty \ \mathbb{P}_x$-p.s..

$$\begin{aligned} 0 < \mathbb{P}_x(\tilde{T}_y < +\infty) &= \mathbb{P}_x(\tilde{T}_y < +\infty, N_x = +\infty) \\ &= \mathbb{P}_x(\tilde{T}_y < +\infty, n_x(X) = +\infty) \\ &= \mathbb{P}_x(\tilde{T}_y < +\infty, n_x(\Theta^{\tilde{T}_y} X) = +\infty, X_{\tilde{T}_y} = y) \\ &= \mathbb{P}_y(n_x(X) = +\infty)\mathbb{P}_x(\tilde{T}_y < +\infty, X_{\tilde{T}_y} = y) \qquad \text{par la propriété de MARKOV forte} \\ &= \mathbb{P}_y(N_x = +\infty) \underbrace{\mathbb{P}_x(\tilde{T}_y < +\infty)}_{\neq 0} \end{aligned}$$

Ainsi $N_x = +\infty \ \mathbb{P}_y$-p.s.. Donc $U(y,x) = +\infty$ ou encore $y \rightsquigarrow x$.

Comme $x \leftrightsquigarrow y$, soient ℓ et p tels que $Q^\ell(x,y) > 0$ et $Q^p(x,y) > 0$.
Calculons :

$$U(y,y) = \sum_{n\geq 0} Q^n(y,y) \geq \sum_{n\geq 0} Q^{\ell+n+p}(y,y) \geq \underbrace{Q^\ell(y,x)}_{>0}\underbrace{U(x,x)}_{=+\infty}\underbrace{Q^p(x,y)}_{>0} = +\infty$$

Ainsi y est récurrent.

On a vu que x récurrent et $y \rightsquigarrow x$ implique $N_x = +\infty \ \mathbb{P}_y$-p.s..
Or on sait que y récurrent et $x \rightsquigarrow y$. D'où $N_y = +\infty \ \mathbb{P}_x$-p.s..

- Le système s'obtient par contraposée du premier point.
Calculons si y est transitoire et $z \neq y$:

$$U(z,y) = \mathbb{E}_z[N_y] = \mathbb{E}_z[\mathbb{1}_{\tilde{T}_y<+\infty} N_y] + \underbrace{\mathbb{E}_z[\mathbb{1}_{\tilde{T}_y=+\infty} N_y]}_{=0} = \mathbb{E}_z[n_y(\Theta^{\tilde{T}_y} X)\mathbb{1}_{\tilde{T}_y<+\infty, X_{\tilde{T}_y}=y}]$$

$$\begin{aligned} &= \mathbb{E}_z[\mathbb{E}_z[n_y(\Theta^{\tilde{T}_y} X) \mid \mathcal{F}_{\tilde{T}_y}]\mathbb{1}_{\tilde{T}_y<+\infty, X_{\tilde{T}_y}=y}] = \mathbb{E}_z[\mathbb{E}_y[n_y(X)]\mathbb{1}_{\tilde{T}_y<+\infty, X_{\tilde{T}_y}=y}] \qquad \text{par MARKOV fort} \\ &= \mathbb{E}_y[N_y]\mathbb{P}_z(\tilde{T}_y < +\infty) \end{aligned}$$

Ainsi $U(z,y) = U(y,y)\mathbb{P}_z(\tilde{T}_y < +\infty) < +\infty$ où $U(y,y) < +\infty$ car y est transitoire.
Donc $\mathbb{E}_z[N_y] < +\infty$ donc $N_y < +\infty \ \mathbb{P}_z$-p.s..

– Ok.

□

DÉFINITION 10.39. Soient Q et E fixés.

- On dit que Q est *irréductible* s'il n'y a qu'une seule classe dans E.
- Si Q est irréductible, on dira que Q est *récurrent* lorsque tous les points sont récurrents, que Q est *transitoire* lorsque tous les points sont transitoires.

REMARQUE 10.40.

- Si Q est irréductible, il est soit récurrent, soit transitoire par la propriété précédente.
- On attribue aussi les attributs irréductible, transitoire, récurrent au processus X.

EXEMPLE 10.41. Les marches aléatoires simples sur $\mathbb{Z}^d$ sont irréductibles.

Si $d = 1$ ou $d = 2$, elles sont récurrentes et alors pour tout x, y, on a $N_y = +\infty$ $\mathbb{P}_x$-p.s..

Si $d \geq 3$, elles sont transitoires et alors $N_y < +\infty$ $\mathbb{P}_x$-p.s.. Ainsi :

$$\mathrm{card}(n \in \mathbb{N} \,|\, \|X_n\| \leq R) = \sum_{\|y\| \leq R} N_y < +\infty \ \mathbb{P}_x\text{-p.s.}$$

On a donc $\liminf_{n\to+\infty} \|X_n\| \geq R$ $\mathbb{P}_x$-p.s. pour tout R. Donc :

$$\|X_n\| \underset{n\to+\infty}{\longrightarrow} +\infty \ \mathbb{P}_x\text{-p.s.}$$

THÉORÈME 10.42. [DÉCOMPOSITION D'UNE CHAÎNE DE MARKOV]

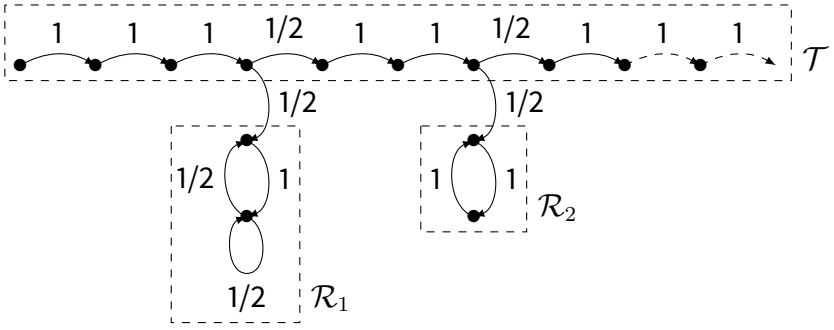
Soient Q et E fixés. Soient $\mathcal{T} = \{x \in E \,|\, x \text{ transitoire}\}$ et $\mathcal{R} = \{x \in E \,|\, x \text{ récurrent}\} = \bigcup_{i\in I} \mathcal{R}_i$ où les $(\mathcal{R}_i)_{i\in I}$ sont les classes d'équivalence pour $\leftrightsquigarrow$ (I est fini ou dénombrable).

Alors pour tout x, on a $\mathbb{P}_x$- p.s. :

- *ou bien* $\quad \forall n \in \mathbb{N}, X_n \in \mathcal{T} \quad$ *et* $\quad \forall y \in E, N_y < +\infty \quad$ *et* $\quad \forall y \in \mathcal{R}, N_y = 0$
- *ou bien $T = \inf(\{n \,|\, X_n \in \mathcal{R}\}) < +\infty$ et si $i \in I$ est tel que $X_T \in \mathcal{R}_i$, alors*

$$\forall y \in \mathcal{T}, N_y < +\infty \quad \text{et} \quad \forall y \in \mathcal{R} \setminus \mathcal{R}_i, N_y = 0 \quad \text{et} \quad \forall y \in \mathcal{R}_i, N_y = +\infty$$

EXEMPLE 10.43.



PREUVE

- Si $x \in \mathcal{R}_i$:
  - $\rightarrow$ si $y \notin \mathcal{R}_i$, on n'a pas $x \rightsquigarrow y$ donc $N_y = 0$ $\mathbb{P}_x$-p.s..
  - $\rightarrow$ si $y \in \mathcal{R}_i$, alors $x \rightsquigarrow y$, donc $N_y = +\infty$ $\mathbb{P}_x$-p.s..

  Soit $A_i = \{\forall y \in \mathcal{R}_i, N_y = +\infty, \forall y \notin \mathcal{R}_i, N_y = 0, \forall n \in \mathbb{N}, X_n \in \mathcal{R}_i\}$.
  On a pour tout $x \in \mathcal{R}_i$, $\mathbb{P}_x(A_i) = 1$.
- Si $x \in \mathcal{T}$: soit $T_i = \inf(\{n \in \mathbb{N} \mid X_n \in \mathcal{R}_i\}$ pour $i \in I$ et $T = \inf_{i\in I} T_i$. Les $(T_i)_{i\in I}$ finis sont distincts.
  Si $T = +\infty$, ok.
  Sinon, il existe $i \in I$ tel que $T = T_i$. Alors :

$$\begin{aligned}
&\mathbb{P}_x(T < +\infty, \forall n > T, X_n \in \mathcal{R}_i, \forall y \in \mathcal{R}_i, N_y = +\infty, \forall y \in \mathcal{R} \setminus \mathcal{R}_i, N_y = 0) \\
&\qquad = \mathbb{P}_x(T < +\infty, \sqcup_{y\in\mathcal{R}_i} X_T = y, \Theta^T X \in A_i) \\
&\qquad = \sum_{y\in\mathcal{R}_i} \mathbb{P}_x(\Theta^T X \in A, T < +\infty, X_T = y) \\
&\qquad = \sum_{y\in\mathcal{R}_i} \mathbb{P}_y(A_i)\mathbb{P}(T < +\infty, X_T = y) \qquad \text{par MARKOV fort} \\
&\qquad = \mathbb{P}(T = T_i < +\infty)
\end{aligned}$$

□

REMARQUE 10.44. Si E est fini et Q est irréductible, alors Q est récurrent puisque

$$\sum_{y\in E} N_y = \sum_{y\in E}\sum_{n\geq 0} \mathbb{1}_{X_n=y} = \sum_{n\geq 0} \underbrace{\sum_{y\in E} \mathbb{1}_{X_n=y}}_{=1} = +\infty$$

La première somme étant à support fini, si Q était transitoire la somme serait finie $\mathbb{P}_x$- p.s..

III. Comportement asymptotique des chaînes de MARKOV

Que devient X_n lorsque n tend vers $+\infty$?

- le comportement presque sûr de X_n est peu intéressant. E est discret donc $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X_\infty$ implique X_n constant à partir d'un certain rang, donc X_n s'arrête sur un site x tel que $Q(x, y) = \mathbb{1}_{x=y}$: un point *absorbant*.
  Par exemple, pour le processus de GALTON-WATSON, 0 est absorbant et si $\mathbb{E}[X_j^n] \leq 1$ et de loi distincte de δ_1, alors $Z_n \xrightarrow[n\to+\infty]{\text{p.s.}} 0$.
- pour la convergence en loi : μ_{X_n} la loi sur E a-t-elle une limite μ_∞ ?
- temps d'occupation : proportion du temps passé en un site : $\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{X_n=y} \xrightarrow[n\to+\infty]{} \mu_\infty(y)$ $\mathbb{P}_x$-p.s.?

Désormais, si ν est une mesure sur E, on notera $\nu(x)$ pour $\nu(\{x\})$.

Calculons $\mu_{X_{n+1}}$ à partir de μ_{X_n} :

$$\forall y \in E, \mu_{X_{n+1}}(y) = \mathbb{P}(X_{n+1} = y) = \sum_{x\in E} \mathbb{P}(X_n = x, X_{n+1} = y)$$
$$= \sum_{x\in E} \mathbb{P}(X_{n+1} = y \mid X_n = x)\mathbb{P}(X_n = x) = \sum_{x\in E} Q(x, y)\mu_{X_n}(x)$$

DÉFINITION 10.45.

- Si E est fini, énumérons ses éléments : $E = \{e_1, \dots, e_n\}$.
  On identifie une mesure ν sur E et le vecteur ligne $(\nu(e_1), \dots, \nu(e_n))$.
- Pour E quelconque, Q noyau de transition, ν mesure sur E, on définit la mesure νQ sur E par :
$$\forall y \in E, (\nu Q)(y) = \sum_{x\in E} \nu(x)Q(x, y)$$

Ceci coïncide avec le calcul matriciel.

REMARQUE 10.46. On définit cela pour des mesures ν pas forcément de probabilité mais telles que $\forall y, \nu(y) < +\infty$ (cela n'exclue pas que $\nu(E) = +\infty$).

On a donc $\forall n \in \mathbb{N}, \mu_{X_{n+1}} = \mu_{X_n} Q$, et donc par une récurrence immédiate $\mu_{X_n} = \mu_{X_0} Q^n$.

On s'attend à ce que μ_{X_n} converge vers un point fixe de $\mu \longmapsto \mu Q$.

DÉFINITION 10.47. Soit μ une mesure sur E telle que $\forall y \in E, \mu(y) < +\infty$ et $\mu(E) \neq 0$.
μ est dite *invariante* si $\mu Q = \mu$.

Les mesures invariantes sont les bonnes candidates pour les mesures limites :

PROPOSITION 10.48. *Soit Q fixé et X la réalisation canonique. Alors on a équivalence entre :*

(i) *μ est invariante,*

(ii) *il existe μ_0 telle que sous $\mathbb{P}_{\mu_0}$, on a $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z$ où $Z \sim \mu$.*

(iii) *Si $X_n \sim \mu$, alors $\forall k \geq n, X_k \sim \mu$.*

PREUVE

(i) $\Longrightarrow$ (ii) Prenons $\mu_0 = \mu$. Alors $\mu_{X_n} = \mu_{X_0} Q^n = \mu Q^n = \mu$ donc $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z$.

(iii) $\Longrightarrow$ (ii) Prenons $\mu_0 = \mu$. Alors $\forall n \in \mathbb{N}, \mu_{X_n} = \mu$.

(i) $\Longrightarrow$ (iii) Si $X_n \sim \mu$, on a pour $k \geq n$: $\mu_{X_k} = \mu_{X_n} Q^{k-n} = \mu$.

(ii) $\Longrightarrow$ (i) Si $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z$ alors $X_{n+1} \xrightarrow[n\to+\infty]{\mathcal{L}} Z$. On a $\forall y \in E, \mu_{X_{n+1}}(y) \xrightarrow[n\to+\infty]{} \mu(y)$.
Et $\forall y \in E, \mu_{X_{n+1}}(y) = (\mu_{X_n} Q)(y) = \sum_{x\in E} \mu_{X_n}(x)Q(x, y) \xrightarrow[n\to+\infty]{} \mu(y)$.
Donc par le lemme de FATOU :

$$\sum_{x\in E} \mu(x)Q(x, y) = \sum_{x\in E} \liminf_{n\to+\infty} \mu_{X_n}(x)Q(x, y) \leq \liminf_{n\to+\infty} \sum_{x\in E} \mu_{X_n}(x)Q(x, y) = \mu(y)$$

Donc $\sum_{x\in E}\mu(x)Q(x,y)\leq\mu(y)$. En sommant sur $y\in E$:

$$1=\sum_{x\in E}\mu(x)=\sum_{y\in E}\sum_{x\in E}\mu(x)Q(x,y)\leq\sum_{y\in E}\mu(y)=1$$

On a donc en fait que des égalités. Donc μ est invariante.

□

Désormais, *on cherche les mesures invariantes*, c'est-à-dire on cherche à résoudre $\mu Q=\mu$.
On veut connaître les conditions d'*existence* et d'*unicité* de cette équation.

EXEMPLE 10.49. [QU'EN PENSE NOTRE HAMSTER ?]

On se replace dans la situation du début du chapitre et on rappelle que

$$Q=\begin{pmatrix}0&3/4&1/4\\1/2&1/3&1/6\\1/2&1/2&0\end{pmatrix}$$

Soit $\mu_0=(\mu_0(1),\mu_0(2),\mu_0(3))$ la mesure initiale.

La somme sur chaque ligne de Q vaut 1, donc $Q(1,1,1)^\top=(1,1,1)^\top$. Ainsi 1 est valeur propre de Q donc de ${}^tQ$. Il existe μ vecteur ligne tel que ${}^tQ\,{}^t\mu={}^t\mu$, ou encore $\mu Q=\mu$. C'est donc une mesure invariante si ces coefficients sont positifs.

Pour trouver μ, réduisons Q. Q a pour valeurs propres 1, $x=\frac{-1+i\sqrt{2}}{3}$ et $y=\frac{-1-i\sqrt{2}}{3}$ (on a $|x|,|y|<1$).
Un vecteur propre associé à 1 pour ${}^tQ$ est $(1/3,1/2,1/6)^\top$. Donc $\boxed{\mu=(1/3,1/2,1/6)}$ est mesure invariante :

$${}^tQ=P\begin{pmatrix}1&0&0\\0&x&0\\0&0&y\end{pmatrix}P^{-1}\quad\text{avec}\quad P=\begin{pmatrix}1/3&&\\1/2&\star&\star\\1/6&&\end{pmatrix}$$

$$({}^tQ)^n=P\begin{pmatrix}1&0&0\\0&x^n&0\\0&0&y^n\end{pmatrix}P^{-1}\underset{n\to+\infty}{\longrightarrow}P\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}P^{-1}=\begin{pmatrix}1/3&0&0\\1/2&0&0\\1/6&0&0\end{pmatrix}P^{-1}$$

Donc ${}^t(\mu_{X_n})=({}^tQ)^n\,{}^t\mu_0\underset{n\to+\infty}{\longrightarrow}\begin{pmatrix}1/3&0&0\\1/2&0&0\\1/6&0&0\end{pmatrix}P^{-1}\,{}^t\mu_0=\alpha\mu^+$ pour un certain α.
Ainsi $1=\sum_{i=1}^3\mu_{X_n}(i)\underset{n\to+\infty}{\longrightarrow}\alpha\sum_{i=1}^3\mu(i)=\alpha$, donc $\alpha=1$ et alors $\mu_{X_n}(i)\underset{n\to+\infty}{\longrightarrow}\mu(i)$.

Soit Q un noyau de transition sur E. Soit X une chaîne de MARKOV de transition Q.
On sait que si $X_n\underset{n\to+\infty}{\xrightarrow{\mathcal{L}}}Z\sim\mu$ alors μ est une mesure de probabilité invariante ($\mu Q=\mu$). On a deux objectifs :

- trouver des conditions d'existence et d'unicité des mesures invariantes,
- comprendre comment trouver des mesures invariantes.

DÉFINITION 10.50. [MESURE SYMÉTRIQUE]

Soient E et Q fixés. μ mesure sur E est dite *symétrique* (ou *réversible*) si :

$$\forall x, y \in E, \mu(x)Q(x,y) = \mu(y)Q(y,x)$$

PROPOSITION 10.51. *Si μ est réversible, alors μ est invariante.*

REMARQUE 10.52. Il est plus facile de trouver les mesures réversibles (même s'il y a pleins de cas sans mesures réversibles).
Le nom vient du fait que si μ est une mesure réversible alors sous $\mathbb{P}_\mu$:

$$(X_0, \dots, X_n) \overset{\mathcal{L}}{=} (X_n, \dots, X_0)$$

PREUVE Si μ est réversible, on a :

$$\forall y \in E, (\mu Q)(y) = \sum_{z \in E} \mu(z)Q(z,y) = \sum_{z \in E} \mu(z)Q(y,z) = \mu(y) \sum_{z \in E} Q(y,z) = \mu(y)$$

□

EXEMPLE 10.53. Si S est la marche aléatoire simple sur $\mathbb{Z}^d$. On a :

$$Q(x,y) = \begin{cases} \frac{1}{2^d} & \text{si } \|x-y\|_1 = 1 \\ 0 & \text{sinon} \end{cases}$$

Cherchons les mesures symétriques : on cherche μ telle que $\forall x, y \in E, \mu(x)Q(x,y) = \mu(y)Q(y,x)$.
Cela donne alors pour tout $x, y \in E$ tels que $\|x-y\|_1 = 1$:

$$\mu(x)\frac{1}{2^d} = \mu(y)\frac{1}{2^d}$$

Et donc, par connexité, $\forall x \in E, \mu(x) = \mu(0)$.
Ainsi μ est un multiple de la *mesure de comptage*. Ce ne peut donc pas être une mesure de probabilité.

EXEMPLE 10.54. [URNES D'EHRENFEST]

On considère N boules bleues ou rouges et on note X_n le nombre de boules rouges au bout de n étapes. A chaque étape, on choisit une boule uniformément parmi les N et on change sa couleur.

X est alors une chaîne de MARKOV sur $E = [\![0, N]\!]$ de noyau de transition défini par :

$$\forall i, j \in E, \begin{cases} Q(i, i+1) = \frac{N-i}{N} \\ Q(i, i-1) = \frac{i}{N} \\ Q(i,j) = 0 \quad \text{si } j \notin \{i-1, i+1\} \end{cases}$$

Cherchons μ réversible. On doit avoir $\forall i, j \in E, \mu(i)Q(i,j) = \mu(j)Q(j,i)$. A $i \in E$ fixé, cela donne $0 = 0$ si $j \notin \{i-1, i+1\}$. La relation est donc équivalente à

$$\forall i \in E, \mu(i-1)\underbrace{Q(i-1,i)}_{=\frac{N-i+1}{N}} = \mu(i)\underbrace{Q(i,i-1)}_{=\frac{i}{N}}$$

D'où

$$\forall i \in E, \mu(i) = \mu(i-1)\frac{N+1-i}{i} = \mu(0)\frac{N+1-1}{1}\frac{N+1-2}{2}\dots\frac{N+1-i}{i} = \mu(0)\frac{N!}{(N-i)!i!}$$

Les mesures symétriques sont les mesures μ telles que pour une certain α, on a pour tout $i \in E$, $\mu(i) = \alpha\binom{N}{i}$. *Il y a donc une mesure de probabilité réversible*, pour $\alpha = \frac{1}{2^N}$: c'est une loi $\boxed{\mathcal{B}(N, 1/2)}$.

III. A. Existence de mesures invariantes pour E fini

Ici, chercher μ invariante, c'est résoudre $\mu Q = \mu$, ou encore chercher un vecteur propre à gauche pour la valeur propre 1.

PROPOSITION 10.55. *Si E est fini, alors il existe une mesure invariante pour Q.*

PREUVE La somme sur les lignes vaut 1. Donc $Q(1,\dots,1)^\top = (1,\dots,1)^\top$. Ainsi 1 est valeur propre de Q, donc valeur propre de ${}^tQ$. Ainsi il existe $v = (v_1,\dots,v_p)$ tel que $v\,{}^tQ = v$ (où $p = \mathrm{Card}(E)$).

Reste à montrer que pour tout $i \in [\![1,p]\!]$, on a $v_i \geq 0$. On pourra alors identifier v à une mesure.
Posons $w = |v| = (|v_1|,\dots,|v_p|)$. w a des coefficients positifs.
Notant $A_i = \sum_{j=1}^p w_j Q_{ji} - w_i$ pour $i \in [\![1,p]\!]$, il vient :

$$A_i = \sum_{j=1}^p |v_j|\,Q_{ji} - |v_i| \geq \left|\sum_{j=1}^p v_j Q_{ji}\right| - |v_i| = |v_i| - |v_i| = 0$$

Ainsi $A_i \geq 0$ et de plus :

$$\sum_{i=1}^p A_i = \sum_{1\leq i,j\leq p} w_j Q_{ji} - \sum_{i=1}^p w_i = \sum_{j=1}^p w_j - \sum_{i=1}^p w_i = 0$$

Donc les A_i sont nuls, et alors $wQ - w = 0$. w convient puisqu'il s'identifie à une mesure invariante.
□

REMARQUE 10.56. Automatiquement, sur un espace fini, si μ est une mesure invariante alors $\frac{\mu}{\mu(E)}$ est une mesure de probabilité invariante.

THÉORÈME 10.57. *Soient E et Q fixés. Si $x \in E$ est un point récurrent, on définit $\nu_x(y)$ le nombre moyen de passages en $y \in E$ lors d'une boucle de x à x :*

$$\nu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{\tilde{T}_x - 1} \mathbb{1}_{X_n = y}\right]$$

C'est une mesure invariante et on a $\forall y \in E, \nu_x(y) > 0 \iff y \rightsquigarrow x$.

REMARQUE 10.58. On a $\nu_x(x) = 1$ et $\nu_x(y) = \mathbb{E}_x\left[\sum_{n=1}^{\tilde{T}_x} \mathbb{1}_{X_n = y}\right]$.

PREUVE Calculons $\nu_x Q$. Soit $y \in E$. On a :

$$\begin{aligned}
\nu_x Q(y) &= \sum_{z\in E} \nu_x(z)Q(z,y) = \sum_{z\in E} \mathbb{E}_x\Big[\sum_{n=0}^{\tilde{T}_x-1} \mathbb{1}_{X_n=z}\Big]Q(z,y) \\
&= \sum_{z\in E}\sum_{n=0}^{+\infty} \mathbb{P}_x(\underbrace{\tilde{T}_x > n}_{\in\mathcal{F}_n^X}, X_n = z)Q(z,y) \\
&= \sum_{z\in E}\sum_{n=0}^{+\infty} \mathbb{P}_x(\tilde{T}_x > n, X_n = z, X_{n+1} = y) \\
&= \sum_{n=0}^{+\infty}\sum_{z\in E} \mathbb{P}_x(\tilde{T}_x > n, X_n = z, X_{n+1} = y) \qquad \text{par le théorème de FUBINI positif} \\
&= \sum_{n=0}^{+\infty} \mathbb{P}_x(\tilde{T}_x > n, X_{n+1} = y) \\
&= \mathbb{E}_x\Big[\sum_{m=1}^{+\infty} \mathbb{1}_{\tilde{T}_x\geq m}\mathbb{1}_{X_m=y}\Big] = \mathbb{E}_x\Big[\sum_{m=1}^{\tilde{T}_x} \mathbb{1}_{X_m=y}\Big] = \nu_x(y)
\end{aligned}$$

Donc $\nu_x Q = \nu_x$.

On sait que $\nu_x(x) = 1 > 0$. Si on n'a pas $x \rightsquigarrow y$, alors $\nu_x(y) = 0$ (puisque $\forall n \in \mathbb{N}, X_n \neq y$ $\mathbb{P}_x$-p.s.). Si $x \rightsquigarrow y$, alors il existe $n \in \mathbb{N}$ tel que $Q^n(x,y) > 0$. Alors :

$$\nu_x(y) = (\nu_x Q^n)(y) = \sum_{z\in E} \nu_x(z)Q^n(z,y) \geq \nu_x(z)Q^n(x,y) > 0$$

□

REMARQUE 10.59. Si $Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, alors pour tout $\alpha, \beta \geq 0$ tels que $\alpha + \beta = 1$, la mesure de probabilité $\mu_{\alpha,\beta} = \alpha(1/2\delta_1 + 1/2\delta_2) + \beta\delta_3$ est une invariante.

THÉORÈME 10.60. *Si Q est irréductible et récurrent alors il existe une unique (à un facteur multiplicatif près) mesure invariante.*

PREUVE Soit μ une mesure invariante et soit $x \in E$ tel que $\mu(x) > 0$. On peut supposer, quitte à la multiplier, que $\mu(x) = 1$. Montrons alors que $\mu = \nu_x$.

On aura alors le résultat puisque si cela est vrai, on a par irréductibilité $x \leftrightsquigarrow y$ pour tout $y \in E$, donc $\forall y \in E, \mu(y) > 0$. Ainsi pour tout $y \in E$, on a $\mu = \mu(y)\nu_y$. En particulier les $(\nu_y)_{y\in E}$ sont tous proportionnels.

Montrons par récurrence sur $k \in \mathbb{N}$ que pour tout $y \in E$, on a $\mu(y) \geq \mathbb{E}_x\Big[\sum_{n=0}^{(\tilde{T}_x-1)\wedge k} \mathbb{1}_{X_n=y}\Big]$.

Lorsque $y = x$, c'est vrai pour tout k. Et pour $y \neq x$, il vient, si le résultat est vrai au rang k :

$$\begin{aligned}
(\mu Q)(y) &= \sum_{z\in E} \mu(z)Q(z,y) \\
&\geq \sum_{z\in E} \mathbb{E}_x\Big[\sum_{n=0}^{k} \mathbb{1}_{\tilde{T}_x>n}\mathbb{1}_{X_n=z}\Big]Q(z,y) && \text{par hypothèse de récurrence} \\
&= \sum_{z\in E}\sum_{n=0}^{k} \mathbb{P}_x(\tilde{T}_x > n, X_n = z)Q(z,y) \\
&= \sum_{n=0}^{k} \mathbb{P}_x(\tilde{T}_x > n, X_{n+1} = y) = \sum_{n=0}^{k} \mathbb{E}_x(\mathbb{1}_{\tilde{T}_x>n}\mathbb{1}_{X_{n+1}=y}) \\
&= \mathbb{E}_x\Big[\sum_{m=1}^{k+1} \mathbb{1}_{\tilde{T}_x\geq m}\mathbb{1}_{X_m=y}\Big] = \mathbb{E}_x\Big[\sum_{m=1}^{\tilde{T}_x\wedge(k+1)} \mathbb{1}_{X_m=y}\Big] \\
&= \mathbb{E}_x\Big[\sum_{m=0}^{\tilde{T}_x\wedge(k+1)} \mathbb{1}_{X_m=y}\Big] && \text{car } x \neq y
\end{aligned}$$

D'où le résultat par récurrence. En passant à la limite sur k, on obtient alors par convergence monotone :

$$\forall y \in E, \mu(y) \geq \nu_x(y)$$

Soit $y \in E$. On sait que $y \rightsquigarrow x$, donc il existe $n \in \mathbb{N}$ tel que $Q^n(y,x) > 0$. On a alors :

$$1 = \mu(x) = (\mu Q^n)(x) = \sum_{z\in E} \mu(z)Q^n(z,x) \geq \sum_{z\in E} \nu_x(z)Q^n(z,x) = (\nu_x Q^n)(x) = \nu_x(x) = 1$$

Donc si pour un $z \in E$ on avait $\mu_z Q^n(z,y) > \nu_x(z)Q^n(z,y)$, alors on aurait $1 > 1$, ce qui est absurde.
Ainsi il n'y a que des égalités, et donc $\mu = \nu_x$. □

EXEMPLE 10.61. [MARCHE ALÉATOIRE SIMPLE SUR $\mathbb{Z}$ ET $\mathbb{Z}^2$]

La mesure de comptage est invariante et la chaîne est irréductible récurrente donc il n'y a pas de probabilité invariante.

THÉORÈME 10.62. [CLASSIFICATION DES NOYAUX IRRÉDUCTIBLES]

Soit Q irréductible. On a trois possibilités :

1) *Q est irréductible transitoire : $\forall x, y \in E, N_y < +\infty$ $\mathbb{P}_x$-p.s..*
   Il n'y a pas de mesure de probabilité invariante.
2) *Q est irréductible récurrent : $\forall x, y \in E, N_y = +\infty$ $\mathbb{P}_x$-p.s..*
   Il existe une mesure invariante, unique à un facteur multiplicatif près. Alors :
   a) *soit il existe ν mesure de probabilité invariante : on dit que Q est* récurrent positif *et on a*
   $$\forall x \in E, \mathbb{E}_x[\tilde{T}_x] = \frac{1}{\nu(x)}$$
   b) *soit les mesures invariantes vérifient $\mu(E) = +\infty$. On dit que Q est* récurrent nul *et alors*
   $$\forall x \in E, \mathbb{E}_x[\tilde{T}_x] = +\infty$$

PREUVE

– Dans le cas transitoire, supposons qu'il existe μ une mesure de probabilité invariante.
Sous $\mathbb{P}_\mu$, X vérifie $\forall n \in \mathbb{N}, X_n \sim \mu$. On a par convergence dominée :

$$\mathbb{P}_\mu(X_n = y) = \sum_{x\in E} \mu(x)\mathbb{P}_x(X_n = y) = \sum_{x\in E} \mu(x)\mathbb{E}_x[\underbrace{\mathbb{1}_{X_n=y}}_{\longrightarrow 0 \ \mathbb{P}_x\text{-p.s.}}] \underset{n\to+\infty}{\longrightarrow} 0$$

La domination vient du fait que $\sum_{x\in E} \mu(x) = 1$.
Or, $\mathbb{P}_\mu(X_n = y) = \mu(y)$ donc $\forall y \in E, \mu(y) = 0$. Ce qui est absurde.

– Dans le cas récurrent, soit $x \in E$.
On sait que ν_x est invariante. On a :

$$\nu_x(E) = \sum_{y\in E} \nu_x(y) = \sum_{y\in E} \mathbb{E}_x\left[\sum_{n=0}^{\tilde{T}_x-1} \mathbb{1}_{X_n=y}\right] \underset{\text{FUBINI}}{=} \mathbb{E}_x\left[\sum_{n=0}^{\tilde{T}_x-1} 1\right] = \mathbb{E}_x[\tilde{T}_x]$$

Donc si $\nu_x(E) = +\infty$, on est dans le cas récurrent nul et $\mathbb{E}_x[\tilde{T}_x] = +\infty$.
Si $\nu_x(E) < +\infty$, on est dans le cas récurrent positif. Alors $\nu = \frac{\nu_x}{\nu_x(E)}$ est une mesure de probabilité invariante et :

$$\forall x \in E, \nu(x) = \frac{\nu_x(x)}{\nu_x(E)} = \frac{1}{\nu_x(E)} = \frac{1}{\mathbb{E}_x[\tilde{T}_x]}$$

□

EXEMPLE 10.63. [URNES D'EHRENFEST]

On peut considérer un modèle équivalent provenant de la physique : N particules sont contenues dans deux pièces. On note X_n le nombre de particules dans la pièce de gauche à l'étape n. A chaque étape, une particule prise au hasard change de pièce.

$\nu = \mathcal{B}(N, 1/2)$ est une probabilité invariante donc la chaîne est irréductible récurrente positive.

On retrouve alors l'idée que l'on ne verra jamais toutes les particules se rassembler dans la même pièce (par exemple celle de droite : $X_n = 0$).
En effet, le temps typique entre deux moments où la pièce de gauche est vide est :

$$\mathbb{E}_0[\tilde{T}_0] = \frac{1}{\nu(0)} = \frac{1}{\binom{N}{0}\frac{1}{2^N}} = 2^N$$

alors que le temps typique qui sépare deux instants à l'équilibre est :

$$\mathbb{E}_{N/2}[\tilde{T}_{N/2}] = \frac{1}{\nu(N/2)} = \frac{1}{\binom{N}{N/2}\frac{1}{2^N}} \underset{N\to+\infty}{\sim} \frac{\sqrt{\pi N}}{2} \ll 2^N$$

Exemple 10.64. On regarde la marche aléatoire sur $\mathbb{Z}$ où l'on se déplace vers la droite avec proba p et vers la gauche avec proba $1-p$. C'est une chaîne de Markov de noyau de transition défini par :

$$\forall i,j \in \mathbb{Z}, \begin{cases} Q(i,i+1) = p \\ Q(i,i-1) = 1-p \\ Q(i,j) = 0 \qquad \text{si } j \notin \{i-1, i+1\} \end{cases}$$

Si $0 < p < 1$, Q est irréductible. Cherchons alors μ mesure symétrique. La condition est équivalente à

$$\forall i \in \mathbb{Z}, \mu(i)Q(i,i+1) = \mu(i+1)Q(i+1,i)$$

D'où $\forall i \in \mathbb{Z}, \mu(i+1) = \mu(i)\frac{p}{1-p}$, et par une récurrence immédiate, on obtient

$$\forall i \in \mathbb{Z}, \mu(i) = \left(\frac{p}{1-p}\right)^i \mu(0)$$

Cela définit une famille de mesures invariantes proportionnelles (il suffit de fixer la valeur de $\mu(0)$).

Par ailleurs, le modèle étant invariant par translation, on soupçonne que la mesure de comptage est invariante. Vérifions : on doit avoir pour $i \in \mathbb{Z}$:

$$\mu_c(i) = \sum_{j\in\mathbb{Z}} \mu_c(j)Q(j,i) = \mu_c(i-1)p + \mu_c(i+1)(1-p)$$

Ce qui est bien le cas puisque cela équivaut à $1 = p + (1-p)$.

Pour $p \neq 1/2$, on a donc deux mesures invariantes (à facteur multiplicatif près) : μ et μ_c.
Donc la chaîne est transitoire lorsque $p \neq 1/2$. *On n'a pas toujours unicité des mesures invariantes.*

Exemple 10.65. On considère la chaîne de Markov sur $\mathbb{N}$ de noyau de transition :

$$\forall i,j \in \mathbb{N}, \begin{cases} Q(i,i+1) = p_i \\ Q(i,0) = 1-p_i \\ Q(i,j) = 0 \qquad \text{si } j \notin \{0, i+1\} \end{cases}$$

où les $(p_i)_{i\in\mathbb{N}}$ sont des éléments de $[0,1]$.

Q est irréductible si pour tout $i \in \mathbb{N}$, $p_i \in]0,1[$. Cherchons alors μ mesure invariante.
Pour $i \geq 1$, on doit avoir $\mu(i) = \sum_j \mu(j)Q(j,i) = \mu(i-1)p_{i-1}$, et donc par récurrence :

$$\forall i \in \mathbb{N}^*, \mu(i) = \mu(0)p_0 \dots p_{i-1}$$

Puis

$$\mu(0) = \sum_{j\in\mathbb{N}} \mu_j Q(j,0) = \sum_{j\in\mathbb{N}} \mu(0)p_0\dots p_{j-1}(1-p_j) = \lim_{N\to+\infty} \mu(0)\sum_{j=0}^{N}(p_0\dots p_{j-1} - p_0\dots p_j)$$
$$= \mu(0)\lim_{N\to+\infty}(1-p_0\dots p_N) = \mu(0)\Big(1-\prod_{i=0}^{+\infty} p_i\Big)$$

Si $\prod_{i=0}^{+\infty} p_i > 0$ (par exemple avec $p_i = \exp(-\frac{1}{(i+1)^2})$, on a $\prod_{i=0}^{+\infty} p_i = \mathrm{e}^{-\pi^2/6} > 0$), alors on a $\mu(0) = 0$ et donc $\mu = 0$. Ainsi, il n'y a pas de mesure invariante et la chaîne est transitoire.

Exemple 10.66. On s'intéresse à la marche aléatoire biaisée avec un mur à gauche : c'est la chaîne de Markov sur $\mathbb{N}$ définie par

$$\begin{cases} Q(0,1) = 1 & \\ Q(i,i+1) = p & \text{pour } i \in \mathbb{N}^* \\ Q(i,i-1) = 1-p & \text{pour } i \in \mathbb{N}^* \\ Q(i,j) = 0 & \text{sinon} \end{cases}$$

Q est irréductible si $0 < p < 1$. Les μ invariantes vérifient pour $i \geq 2$:

$$(p + (1-p))\mu(i) = \mu(i-1)p + \mu(i+1)(1-p)$$

D'où $p(\mu(i) - \mu(i-1)) = (1-p)(\mu(i+1) - \mu(i))$.
Donc $(\mu(i+1) - \mu(i))_{i\in\mathbb{N}^*}$ est une suite géométrique de raison $\frac{p}{1-p}$. D'où :

$$\forall i \in \mathbb{N}^*, \mu(i+1) - \mu(i) = \left(\frac{p}{1-p}\right)^{i-1}(\mu(2) - \mu(1))$$

En $i = 0$, on doit avoir $\mu(0) = (1-p)\mu(1)$ et en $i = 1$, il faut $\mu(1) = \mu(0) + (1-p)\mu(2)$.
Donc $(\mu(2) - \mu(1))(1-p) = \mu(1) - 2\mu(0)$.
Fixons $\mu(0) = 1$, $\mu(1) = \frac{1}{1-p}$ et alors pour tout $i \geq 2$:

$$\mu(i) = \mu(1) + \sum_{j=1}^{i-1} \mu(j+1) - \mu(j) = \mu(1) + \sum_{j=1}^{i-1}\left(\frac{p}{1-p}\right)^{j-1}(\mu(2) - \mu(1)) = \dots = \left(\frac{p}{1-p}\right)^{i}\mu(0)$$

- Si $p < 1/2$, on reconnaît une loi géométrique (décalée de 1) de paramètre $\frac{p}{1-p}$ et invariante, donc on est dans le cas récurrent positif (on a alors le temps de retour en 0).
- Si $p = 1/2$, la mesure de comptage est invariante, on est dans le cas récurrent nul car $(X_n)_{n\in\mathbb{N}} \overset{\mathcal{L}}{=} (|S_n|)_{n\in}$ où S est la marche aléatoire simple sur $\mathbb{Z}$.
- Si $p > 1/2$, on est dans le cas transitoire : tant que l'on ne touche pas 0, la loi est la même que la marche biaisée sur $\mathbb{Z}$ qui a une probabilité strictement positive de ne jamais revenir en 0 et de partir à l'infini. Donc X aussi.

III. B. Comportement de X_n lorsque n tend vers $+\infty$

THÉORÈME 10.67. [LIMITE DES TEMPS D'OCCUPATION]

Soit Q un noyau irréductible récurrent (positif ou nul) sur E. Soit μ une mesure invariante (unique à constante multiplicative près). Soit $f : E \longrightarrow \mathbb{R}$ une fonction de $L^1(\mu)$.

Alors si X est une chaîne de MARKOV *de transition Q :*

$$\frac{1}{n}\sum_{i=0}^{n} f(X_i) \underset{n\to+\infty}{\longrightarrow} \begin{cases} \int_E f(x)\,d\nu(x) & \text{si } Q \text{ est récurrent positif, de proba invariante } \nu \\ 0 & \text{si } Q \text{ est récurrent nul} \end{cases}$$

REMARQUE 10.68.

- Dans le cas particulier où l'on considère $f_y : x \longmapsto \mathbb{1}_{x=y}$ pour $y \in E$ fixé, alors ce résultat dit :

$$\underbrace{\frac{\mathrm{Card}(\{1 \le i \le n \,|\, X_i = y\})}{n}}_{\text{proportion du temps passé en } y} \underset{n\to+\infty}{\longrightarrow} \begin{cases} \nu(y) & \text{si } Q \text{ est récurrent positif} \\ 0 & \text{sinon} \end{cases}$$

- C'est une généralisation de la loi forte des grands nombres. Si μ est une mesure de probabilité discrète sur $\mathbb{R}$ telle que $\int |x|\ d\mu(x) < +\infty$ et $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ, alors $X = (X_i)_{i\in\mathbb{N}^*}$ est une chaîne de MARKOV de transition $Q(x, y) = \mu(\{y\})$:

$$\mathbb{P}(X_{1:n} = x_{1:n}) = \prod_{i=1}^{n} \mu(\{x_i\})$$

Le résultat sur cette chaîne de MARKOV pour $f = \mathrm{id}$ redonne la loi forte des grands nombres. On est dans le cas récurrent positif avec μ probabilité invariante.

PREUVE

REMARQUE 10.69.

- Il suffit de faire la preuve dans le cas $f \ge 0$ (quitte à décomposer $f = f^+ - f^-$).
- On va prouver ce résultat sous la réalisation canonique et $\mathbb{P}_x$-p.s..
  Si on a un événement A qui est $\mathbb{P}_x$ presque sûr pour tout $x \in E$, alors

$$\mathbb{P}_\mu(A) = \sum_{x\in E} \mu(x)\mathbb{P}_x(A) = \sum_{x\in E} \mu(x) = 1$$

  Donc on aura A $\mathbb{P}_\mu$-p.s. pour toute probabilité μ sur E.

On travaille sous $\mathbb{P}_x$. On note $\tilde{T}_x^p$ le temps de p-ième retour en x, c'est-à-dire :

$$\begin{cases} \tilde{T}_x^0 = 0 \\ \tilde{T}_x^1 = \tilde{T}_x = \inf\{n > 0 \,|\, X_n = x\} \\ \tilde{T}_x^p = \inf\{n > \tilde{T}_x^{p-1} \,|\, X_n = x\} \end{cases} \quad \text{pour } p \in \mathbb{N}^*$$

On a :

$$\frac{1}{n}\sum_{i=1}^{\tilde{T}_x^n} f(X_i) = \frac{1}{n}\sum_{k=1}^{n}\Big(\sum_{j=\tilde{T}_x^{k-1}+1}^{\tilde{T}_x^k} f(X_j)\Big) = \frac{1}{n}\sum_{k=1}^{n} Y_k$$

où $Y_k = \sum_{j=\tilde{T}_x^{k-1}}^{\tilde{T}_x^k+1} f(X_j)$ est la somme des valeurs de f lors de la k-ième exploration.

Cherchons la loi de chaque $(Y_k)_{k\in\mathbb{N}^*}$?
E est dénombrable donc $f(E)$ aussi. Les $(Y_k)_{k\in\mathbb{N}^*}$ sont donc des éléments de {somme finie d'éléments de $f(E)$ qui est dénombrable. On cherche $\mathbb{P}(Y_{1:n} = y_{1:n})$.
On a $\tilde{T}_x^1 = t(X)$ où $t((x_k)_{k\in\mathbb{N}}) = \inf\{n > 0 \,|\, x_n = x\}$ puis pour $p \in \mathbb{N}^*$:

$$\tilde{T}_x^p = \tilde{T}_x^{p-1} + \inf\{n > 0 \,|\, \theta^{\tilde{T}_x^{p-1}} X_n = x\} = \tilde{T}_x^{p-1} + t(\theta^{\tilde{T}_x^{p-1}} X)$$

alors $Y_1 = y(X)$ où $y(x) = \sum_{j=1}^{t(x)} f(x_j)$ puis pour $k \in \mathbb{N}$:

$$Y_k = \sum_{j=\tilde{T}_x^{k-1}+1}^{\tilde{T}_x^k} f(X_j) = \sum_{j=\tilde{T}_x^{k-1}+1}^{\tilde{T}_x^{k-1}+t(\theta^{\tilde{T}_x^{k-1}}X)} f(X_j) = \sum_{j'=1}^{t(\theta^{\tilde{T}_x^{k-1}}X)} f((\theta^{\tilde{T}_x^{k-1}}X)_j) = y(\theta^{\tilde{T}_x^{k-1}}X)$$

Alors :

$$\begin{aligned}
\mathbb{P}_x(Y_{1:n} = y_{1:n}) &= \mathbb{P}_x(\underbrace{Y_{1:n-1} = y_{1:n-1}}_{\in\mathcal{F}_{\tilde{T}_x^{n-1}}}, y(\theta^{\tilde{T}_x^{n-1}}X) = y_n, X_{\tilde{T}_x^{n-1}} = x, \tilde{T}_x^{n-1} < +\infty) \\
&= \mathbb{P}_x(Y_{1:n-1} = y_{1:n-1})\mathbb{P}_x(y(X) = y_n) && \text{par Markov fort} \\
&= \mathbb{P}_x(Y_{1:n-1} = y_{1:n-1})\mathbb{P}_x(Y_1 = y_n) \\
&= \prod_{i=1}^{n} \mathbb{P}_x(Y_1 = y_i) && \text{par récurrence}
\end{aligned}$$

Sous $\mathbb{P}_x$ les $(Y_k)_{k\in\mathbb{N}^*}$ sont donc des variables aléatoires indépendantes et identiquement distribuées. Elles sont positives (car f l'est) et on a :

$$\begin{aligned}
\mathbb{E}[Y_1] = \mathbb{E}\Big[\sum_{j=1}^{\tilde{T}_x} f(X_j)\Big] &= \mathbb{E}\Big[\sum_{\substack{y\in E \\ 1\le j\le \tilde{T}_x}} \mathbb{1}_{X_j=y} f(y)\Big] = \sum_{y\in E} f(y)\mathbb{E}\Big[\sum_{j=1}^{\tilde{T}_x} \mathbb{1}_{X_j=y}\Big] \\
&= \sum_{y\in E} f(y)\nu_x(y) = \int f\, d\nu_x < +\infty && \text{car } f \in L^1
\end{aligned}$$

Donc par la loi forte des grands nombres :

$$\frac{1}{n}\sum_{j=1}^{\tilde{T}_x^n} f(X_j) \underset{n\to+\infty}{\longrightarrow} \int f\, d\nu_x \ \mathbb{P}_x\text{-p.s.} \tag{10.1}$$

– Si on est dans le cas récurrent positif, on a en prenant $f = 1$:

$$\frac{\tilde{T}_x^n}{n} = \frac{1}{n}\sum_{j=1}^{\tilde{T}_x^n} 1 \underset{n\to+\infty}{\longrightarrow} \int f\, d\nu_x = \nu_x(E) \ \mathbb{P}_x\text{-p.s.}$$

Donc pour f quelconque, en divisant (10.1) par $\tilde{T}_x^n/n$ et en passant à la limite :

$$\frac{1}{\tilde{T}_x}\sum_{j=1}^{\tilde{T}_x^n} f(X_j) \underset{n\to+\infty}{\longrightarrow} \frac{1}{\nu_x(E)}\int f\,d\nu_x = \int f\,d\nu \ \mathbb{P}_x\text{-p.s.}$$

où ν est la proba invariante.

Puis pour tout $p \in \mathbb{N}^*$, il existe un unique $N_p \in \mathbb{N}^*$ tel que $\tilde{T}_x^{N_p-1} \leq p < \tilde{T}_x^{N_p}$. On a que $(N_p)_{p\in\mathbb{N}^*}$ diverge vers $+\infty$. Alors :

$$\underbrace{\frac{N_p}{p}}_{\substack{\geq \frac{N_p}{\tilde{T}_x^{N_p}} \\ \sim \frac{1}{\nu_x(E)}}} \underbrace{\frac{N_p-1}{N_p}}_{\longrightarrow 1} \underbrace{\frac{\sum_{j=1}^{\tilde{T}_x^{N_p-1}} f(X_j)}{N_p-1}}_{\longrightarrow \int f\,d\nu_x \ \mathbb{P}_x\text{-p.s.}} \leq \frac{1}{p}\sum_{j=1}^{p} f(X_j) \leq \underbrace{\frac{N_p-1}{p}}_{\substack{\leq \frac{N_p-1}{\tilde{T}_x^{N_p-1}} \\ \sim \frac{1}{\nu_x(E)}}} \underbrace{\frac{N_p}{N_p-1}}_{\longrightarrow 1} \underbrace{\frac{\sum_{j=1}^{\tilde{T}_x^{N_p}} f(X_j)}{N_p}}_{\longrightarrow \int f\,d\nu_x \ \mathbb{P}_x\text{-p.s.}}$$

D'où le résultat par encadrement.

– Dans le cas récurrent nul, on rappelle que E ne peut être fini. Enumérons $E = (e_i)_{i\in\mathbb{N}}$. Fixons $\ell \in \mathbb{N}$ et prenons $f : x \longmapsto \mathbb{1}_{x\in\{e_1,\dots,e_\ell\}}$. On a $f \in L^1(\mu)$ et :

$$\frac{1}{n}\sum_{j=1}^{\tilde{T}_x^n} \mathbb{1}_{X_j\in\{e_1,\dots,e_\ell\}} \underset{n\to+\infty}{\longrightarrow} \nu_x(\{e_1,\dots,e_\ell\}) \ \mathbb{P}_x\text{-p.s.}$$

Soit $g \in L^1(\mu)$ positive. Comme on a f est à valeurs dans $\{0,1\}$, on a $p \geq \sum_{j=1}^{p} f(X_j) \geq \sum_{j=1}^{\tilde{T}^{N_p-1}} f(X_j)$ et donc :

$$0 \leq \frac{1}{p}\sum_{j=1}^{p} g(X_j) \leq \underbrace{\frac{N_p-1}{p}}_{\substack{\leq \frac{N_p-1}{\sum_{j=1}^{\tilde{T}^{N_p-1}} f(X_j)} \\ \sim \frac{1}{\nu_x(\{e_1,\dots,e_\ell\})}}} \underbrace{\frac{N_p}{N_p-1}}_{\longrightarrow 1} \underbrace{\frac{\sum_{j=1}^{\tilde{T}_x^{N_p}} g(X_j)}{N_p}}_{\longrightarrow \int g\,d\nu_x}$$

Et donc puisque $\nu_x(\{e_1,\dots,e_\ell\}) \underset{\ell\to+\infty}{\longrightarrow} \nu_x(E) = +\infty$:

$$0 \leq \limsup_{p\to+\infty} \frac{1}{p}\sum_{j=1}^{p} g(X_j) \leq \underbrace{\frac{\int g\,d\nu_x}{\nu_x(\{e_1,\dots,e_\ell\})}}_{\underset{\ell\to+\infty}{\longrightarrow} 0} \ \mathbb{P}_x\text{-p.s.}$$

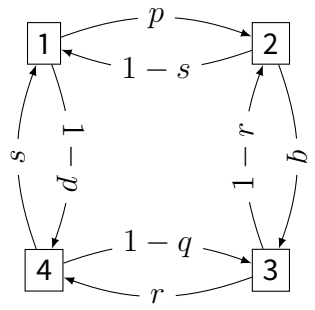
Ce qui conclut.

□

EXEMPLE 10.70. [LIMITE EN LOI DE X_n]

On considère le noyau :

$$Q = \begin{pmatrix} 0 & p & 0 & 1-p \\ 1-q & 0 & q & 0 \\ 0 & 1-r & 0 & r \\ s & 0 & r-s & 0 \end{pmatrix}$$



Si $X_0 = 1$ p.s., on a presque sûrement X_n pair pour n pair, impair pour n impair.

On n'a donc pas de convergence en loi puisque $\mathbb{P}(X_n \in \{1,3\}) = \begin{cases} 1 & \text{si } n \text{ pair} \\ 0 & \text{sinon} \end{cases}$.

DÉFINITION 10.71. Soient Q irréductible sur E et $x \in E$.

Les *périodes* de x sont $\mathcal{P}_x = \{n \in \mathbb{N}^* \mid Q^n(x,x) > 0\}$.

On a $Q^{n+m}(x,x) \geq Q^n(x,x)Q^m(x,x)$ donc $\mathcal{P}_x + \mathcal{P}_x \subset \mathcal{P}_x$. Donc $\mathcal{P}_x - \mathcal{P}_x$ est un sous-groupe de $\mathbb{Z}$: il existe $d_x \in \mathbb{N}^*$ tel que $\mathcal{P}_x - \mathcal{P}_x = d_x\mathbb{Z}$.
On appelle ce d_x la *période* de x.

PROPOSITION 10.72. *Soit Q irréductible.*

- $\forall x \in E, d_x = \mathrm{PGCD}(\mathcal{P}_x)$,
- d_x *ne dépend pas de* $x \in E$,
- *si* $d = d_x = 1$, *on dit que la chaîne est* apériodique *et alors*

$$\forall x, y \in E, \exists n_0 \in \mathbb{N}^* \mid \forall n \geq n_0, Q^n(x,y) > 0$$

PREUVE

- $d_x\mathbb{Z} = \mathcal{P}_x - \mathcal{P}_x$.
  Si $p \in \mathcal{P}_x$, $p + p \in \mathcal{P}_x$ donc $2p - p = p \in \mathcal{P}_x - \mathcal{P}_x$ donc $\mathcal{P}_x \subset d_x\mathbb{Z}$.
  Donc d_x divise tous les élément de $\mathcal{P}_x$.
  Si q divise tous les élément de $\mathcal{P}_x$, alors il divise tous les éléments de $\mathcal{P}_x - \mathcal{P}_x = d_x\mathbb{Z}$ donc $q \mid d_x$. Donc $d_x = \mathrm{PGCD}(\mathcal{P}_x)$.
- Pour tout $x, y \in E$, on a $x \leftrightsquigarrow y$, donc il existe p, q entiers tels que $Q^p(x,y) > 0$ et $Q^q(y,x) > 0$.
  Si $n \in \mathcal{P}_x$, on a $Q^n(x,x) > 0$, donc $Q^{n+p+q}(y,y) \geq Q^q(y,x)Q^n(x,x)Q^p(x,y) > 0$, d'où $n + p + q \in \mathcal{P}_y$. Donc $\mathcal{P}_x + p + q \subset \mathcal{P}_y$.
  Par différence $d_x\mathbb{Z} \subset d_y\mathbb{Z}$ donc $d_y \mid d_x$. Puis par symétrie on obtient $d_x = d_y$.
- Soient $x, y \in E$.
  Si $x = y$, écrivons $1 = d_x = p - q$ pour deux entiers $p, q \in \mathcal{P}_x$.
  Soit $n \geq q^2$. Posons $j = n - q^2$ et considérons sa division euclidienne par q : $j = aq + r$.
  On a $n = q^2 + aq + r = q^2 + aq + r(p - q)$.

Alors $Q^n(x,x) \geq (Q^q(x,x))^{q-r}(Q^q(x,x))^a(Q^p(x,x))^r > 0$. Donc $n_0 = q^2$ convient. Si $x \neq y$, on a pour tout $n \geq p$:

$$Q^n(x,y) \geq Q^p(x,y)Q^{n-p}(y,y)$$

Pour p bien choisi et $n \geq p$, on a $Q^p(x,y) > 0$. Puis $Q^{n-p}(y,y) > 0$ pour n assez grand par le point précédent.

□

THÉORÈME 10.73. *Soit Q irréductible récurrent positif, de probabilité invariante ν, et apériodique.*

Alors si X est une chaîne de MARKOV *de transition Q, on a* $\boxed{X_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z \sim \nu}$*, et même :*

$$\boxed{\sum_{y\in E} |\mathbb{P}(X_n = y) - \nu(y)| \xrightarrow[n\to+\infty]{} 0}$$

PREUVE On va définir Z et Y deux chaînes de MARKOV de transition Q telles que

- Z a même loi que X,
- pour tout $n \in \mathbb{N}$, on a $Y_n \sim \nu$,
- si $T = \inf\{n \mid Z_n = Y_n\}$, alors $(\forall k \geq T, Z_k = Y_k)$ p.s..

a) Supposons que deux telles chaînes de MARKOV existent.
Montrons que si de plus $T < +\infty$ p.s., alors on a le résultat. On a :

$$\begin{aligned}
\sum_{y\in E} |\mathbb{P}(X_n = y) - \nu(y)| &= \sum_{y\in E} |\mathbb{P}(Z_n = y) - \mathbb{P}(Y_n = y)| \\
&= \sum_{y\in E} |\mathbb{P}(Z_n = y, T \leq n) + \mathbb{P}(Z_n = y, T > n) \\
&\qquad - \mathbb{P}(Y_n = y, T \leq n) - \mathbb{P}(Y_n = y, T > n)| \\
&= \sum_{y\in E} |\mathbb{P}(Z_n = y, T > n) - \mathbb{P}(Y_n = y, T > n)| \\
&\leq \sum_{y\in E} \mathbb{P}(Z_n = y, T > n) + \mathbb{P}(Y_n = y, T > n) \\
&= 2\mathbb{P}(T > n) \xrightarrow[n\to+\infty]{} 0 \qquad \text{puisque } T < +\infty \text{ p.s.}
\end{aligned}$$

b) Construisons Z et Y.

Soient $(\zeta^Z_{n,y})_{n\geq 1, y\in E}$, $(\zeta^Y_{n,y})_{n\geq 1, y\in E}$, ζ^Z_0 et ζ^Y_0 des variables aléatoires indépendantes à valeurs dans E, telles que pour tout $n \geq 1$ et $y \in E$, on ait :

$$\zeta^Y_0 \sim \nu \qquad \zeta^Y_{n,y} \sim Q(y,.) \qquad \zeta^Z_0 \sim \mu_{X_0} \qquad \zeta^Y_{n,y} \sim Q(y,.)$$

Posant

$$\begin{cases} Y_0 = \zeta^Y_0 \\ Y_n = \zeta^Y_{n,Y_{n-1}} \end{cases} \text{ pour } n \geq 1 \qquad \text{puis} \qquad \begin{cases} Z_0 = \zeta^Y_0 Z \\ Z_n = \begin{cases} \zeta^Z_{n,Z_{n-1}} & \text{si } Z_{n-1} \neq Y_{n-1} \\ \zeta^Y_{n,Z_{n-1}} & \text{si } Z_{n-1} = Y_{n-1} \end{cases} \end{cases} \text{ pour } n \geq 1$$

(ceci force $Y_n = Z_n$ pour $n \geq T$), on a que Y est une chaîne de MARKOV de transition Q et de mesure initiale ν, puis on montre (exercice) que Z est une chaîne de MARKOV de transition Q qui a même loi que X.

c) Montrons que $T < +\infty$ p.s..

On doit montrer que 2 chaînes de MARKOV Y, Z de transition Q indépendantes se rencontrent presque sûrement. En fait, Y et Z sont indépendantes jusqu'à leur rencontre, mais faisons comme si elles étaient totalement indépendantes. On a (en notant $(a,b)_i = (a_i, b_i)$) :

$$\begin{aligned}\mathbb{P}((Z,Y)_{0:n} = (x,y)_{0:n}) &= \mathbb{P}(Z_{0:n} = x_{0:n})\mathbb{P}(Y_{0:n} = y_{0:n}) \\ &= \mathbb{P}((Z_0, Y_0) = (x_0, y_0)) \prod_{i=0}^{n-1} Q(x_i, x_{i+1})Q(y_i, y_{i+1})\end{aligned}$$

Donc $(Z_n, Y_n)_{n\in\mathbb{N}}$ est une chaîne de MARKOV sur $E \times E$ de transition définie par $\tilde{Q}((x,y),(x',y')) = Q(x,x')Q(y,y')$ pour $x, x', y, y' \in E$.

Par périodicité, on a pour n assez grand :

$$\tilde{Q}^n((x,y),(x',y')) = Q^n(x,x')Q^n(y,y') > 0$$

Donc $\tilde{Q}$ est irréductible et on a pour $x, y \in E$:

$$\begin{aligned}(\nu \otimes \nu.\tilde{Q})(x,y) &= \sum_{x',y'\in E} \nu(x')\nu(y')\tilde{Q}((x',y'),(x,y)) \\ &= \sum_{x'\in E} \nu(x')Q(x',x) \sum_{y'\in E} \nu(y')Q(y',y) \\ &= \nu(x)\nu(y) = (\nu\otimes\nu)(x,y)\end{aligned}$$

Donc $\nu \otimes \nu$ est une probabilité $\tilde{Q}$-invariante donc $\tilde{Q}$ est récurrent positif.
Ainsi pour $x \in E$, on a $T = \inf\{n \mid Y_n = Z_n\} \leq \inf\{n \mid (Y_n, Z_n) = (x,x)\} < +\infty$ p.s..

□