

# Probabilités, Modèles & Applications

Adrien Hardy\*

28 novembre 2019

*Ce document, en construction, contient les notes du cours “Probabilités, Modèles et Applications” donné au semestre 1 du Master MAS et CHPS. Il vise à fournir les bases de la théorie des probabilités nécessaires au futur ingénieur ou chercheur en statistiques appliquées et ses ramifications. Il est calibré pour un volume horaire de  $12 \times 1h40$  de cours et de  $12 \times 2h30$  de TD. Toute remarque et aide au débusquage de coquilles est bienvenu.*

La théorie des probabilités donne un cadre rigoureux pour manipuler quantitativement la notion de hasard. Les applications pratiques liées à l'aléa reposent sur ce socle, ce qui inclut les modèles du *big-data*, du *machine learning* ou de l'*intelligence artificielle*, des modèles qui modifient actuellement nos sociétés en profondeur. Le principe général derrière ces applications est le suivant : on récolte des données souvent complexes ♠ et on imagine que ces données sont des réalisations de variables aléatoires. En utilisant des résultats théoriques au niveau des modèles choisis pour ces variables aléatoires, on peut alors extraire des informations clés ♡ depuis les données avec des *garanties quantitatives*, c'est-à-dire un contrôle de l'erreur d'avoir extrait une mauvaise information. Ces garanties sont donc précieuses pour l'aide à la décision et l'estimation du risque. Plus la partie théorique des modèles est développée, plus les garanties sont solides.

Le but de ce cours est de fournir le langage nécessaire à la définition et à la compréhension des modèles aléatoires ainsi que les outils utiles à l'obtention de garanties quantitatives ; ces aspects sont développés en parallèle dans les cours “Statistique Mathématique” et les modules informatiques (TISD, TIAD) du Master MAS et CHPS.

## Table des matières

<b>1</b>	<b>Théorie de la mesure : Kit de survie</b>	<b>3</b>
1.1	Ce qui ne peut être mesuré . . . . .	3
1.2	Mesures et ensembles mesurables . . . . .	4
1.3	Fonctions mesurables et intégration . . . . .	5
1.4	Ensembles négligeables et densités . . . . .	8

---

\*Laboratoire Paul Painlevé, Université de Lille, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France. Email: [adrien.hardy@univ-lille.fr](mailto:adrien.hardy@univ-lille.fr)

♠. par exemple, l'évolution du cours d'une action pendant une semaine, un ensemble d'images de Google image, les résultats médicaux de patients testant un nouveau médicament, les données météorologiques d'une région, etc

♡. par exemple, la valeur future du cours d'une action (prédiction), identifier si une image représente un chat ou non (classification), décider si un médicament fait mieux que l'effet placebo (test d'hypothèse), ajuster les paramètres d'une équation d'évolution météorologique (régression paramétrique).

1.5	Integration et limites de fonctions . . . . .	9
1.6	Mesures produit . . . . .	10
1.7	Changement de variables . . . . .	10
1.8	EXERCICES – Théorie de la mesure . . . . .	12
<b>2</b>	<b>Probabilités : Boîte à outils</b>	<b>15</b>
2.1	Variables aléatoires . . . . .	15
2.2	Variables aléatoires réelles . . . . .	16
2.3	Vecteurs aléatoires . . . . .	17
2.4	Quelques inégalités importantes . . . . .	20
2.5	Convergence de variables aléatoires . . . . .	23
2.6	Loi des grands nombres . . . . .	25
2.7	Théorème central limite . . . . .	26
2.8	EXERCICES – Probabilités . . . . .	28
<b>3</b>	<b>Espérance conditionnelle</b>	<b>32</b>
3.1	Motivations et exemple . . . . .	32
3.2	Espérance conditionnelle par rapport à une sous-tribu . . . . .	34
3.3	Propriétés de base . . . . .	35
3.4	Inégalités conditionnelles . . . . .	36
3.5	Théorèmes de convergence conditionnelle . . . . .	37
3.6	Calcul pratique d'espérances conditionnelles . . . . .	37
3.7	EXERCICES – Espérance conditionnelle . . . . .	40
<b>4</b>	<b>Chaines de Markov</b>	<b>41</b>
4.1	Processus aléatoires (généralités) . . . . .	41
4.2	Chaînes de Markov . . . . .	42
4.3	Propriété de Markov (forte) . . . . .	44
4.4	Mesures invariantes . . . . .	45
4.5	Récurrence . . . . .	47
4.6	Application : L'algorithme PageRank . . . . .	56
4.7	EXERCICES – chaînes de Markov . . . . .	58

# 1 Théorie de la mesure : Kit de survie

Comprendre les applications des statistiques et leurs ramifications requiert donc une compréhension solide de la théorie des probabilités <sup>♣</sup>. Depuis que A. Kolmogorov <sup>♡</sup> en a posé les fondations, ce cadre théorique utilise comme matière première le concept de *mesure d'un ensemble*. Définir rigoureusement ce qu'est une mesure comme on peut l'imaginer intuitivement n'est en fait pas évident ; c'est l'objet de la théorie de la mesure que nous allons survoler.

## 1.1 Ce qui ne peut être mesuré

Tout étudiant en mathématiques a été confronté au problème de l'infini, cette notion qui permet de créer tous ces contre-exemples à des propriétés qu'on pensait intuitives. Rappelez-vous de la preuve qui montre que  $\mathbb{N}$  et  $\mathbb{R}$  n'ont pas le même infini pour cardinal. Manipuler l'infini peut être assez traumatisant en théorie des ensembles, qui est à la base des fondations des mathématiques. Il faut garder en tête que les mathématiques reposent sur une série d'axiomes (par exemple, que le principe de preuve par récurrence marche) et nous prenons ici les axiomes usuels de la théorie des ensembles de Zermelo-Fraenkel avec l'Axiome du choix, utilisés par la grande majorité de la communauté mathématique. Ce dernier axiome est nécessaire à la preuve d'importants théorèmes des probabilités et d'analyse fonctionnelle, mais il va compliquer notre intuition de ce qu'est une *mesure*.

Idéalement, une *mesure*  $\mu$  définie sur un ensemble  $E$  devrait assigner une mesure numérique (comme une “longueur”, un “volume” ou une “masse”) à chaque sous-ensemble de  $E$ , c'est-à-dire définir une application  $\mu : \mathcal{P}(E) \rightarrow [0, +\infty]$  où  $\mathcal{P}(E)$  désigne l'ensemble des sous-ensembles de  $E$ . On aimerait que  $\mu$  satisfasse des propriétés raisonnables, en l'occurrence que  $\mu(\emptyset) = 0$  et que, si  $A, B \in \mathcal{P}(E)$  sont disjoints, alors  $\mu(A \cup B) = \mu(A) + \mu(B)$ . Par itération, cette dernière condition implique que la mesure d'une réunion finie d'ensembles disjoints est la somme de leur mesure ; on dit que  $\mu$  est *additive*.

Pour les applications pratiques ce n'est pas suffisant car, dès que l'on veut prendre des limites “ $n \rightarrow \infty$ ”, il va falloir considérer des réunions infinies dénombrables d'ensembles. On fait alors l'hypothèse plus forte que la propriété précédente marche aussi pour les réunions infinies dénombrables ; on dit que  $\mu$  est  *$\sigma$ -additive*. Le problème est que, même pour un ensemble  $E$  aussi peu exotique que  $\mathbb{R}$ , n'est pas mesure qui veut : il est par exemple impossible (cf. Exercice 7) de définir une application  $\mu : \mathcal{P}(\mathbb{R}) \rightarrow [0, +\infty]$  qui est  $\sigma$ -additive et telle que  $\mu([a, b]) = b - a$  pour tout  $a < b$  dans  $\mathbb{R}$ , qui correspondrait bien à la “mesure intuitive” d'un intervalle de  $\mathbb{R}$ . Cette obstruction nécessite cependant l'utilisation l'Axiome du choix.

La conclusion est que si l'on ne veut pas se passer de la  $\sigma$ -additivité ou de l'Axiome du choix, il va falloir accepter qu'une mesure ne soit pas définie sur *tout*  $\mathcal{P}(E)$  et de restreindre son ensemble de définition. Cet ensemble de définition doit quand même satisfaire quelques propriétés de stabilité : c'est là qu'entre en jeu la notion d'ensemble mesurable et de tribu.

---

♣. mais ça ne suffit pas. Des bases solides en informatique et une expérience pratique de l'exploitation des données est tout aussi nécessaire.

♡. Andreï Kolmogorov, 1903–1987. Vous remarquerez que la théorie des probabilités, considéré comme une branche des mathématiques, est relativement récente.

## 1.2 Mesures et ensembles mesurables

Soit  $E$  un ensemble. Une tribu sur  $E$  est un sous-ensemble de  $\mathcal{P}(E)$  stable par complémentaire et réunion dénombrable, et qui contient l'ensemble vide  $\emptyset$  :

**Définition 1.1.** Une tribu (aussi appelée  $\sigma$ -algèbre)  $\mathcal{T}$  de  $E$  est une collection de sous-ensembles de  $E$  qui satisfait :

- (a)  $\emptyset \in \mathcal{T}$
- (b)  $A \in \mathcal{T} \Rightarrow A^c := E \setminus A \in \mathcal{T}$
- (c) Si  $A_n \in \mathcal{T}$  pour tout  $n \geq 1$ , alors  $\bigcup_{n \geq 1} A_n \in \mathcal{T}$ .

On dit que  $(E, \mathcal{T})$  est un espace mesurable et  $A \in \mathcal{T}$  est un ensemble mesurable.

Notez que si  $A, B \in \mathcal{T}$  alors  $A \cap B \in \mathcal{T}$  (prouvez-le).

**Exemple :**  $\{\emptyset, E\}$  et  $\mathcal{P}(E)$  sont des tribus, respectivement la plus petite et la plus grande des tribus possibles sur  $E$ .

Comme  $\{\emptyset, E\}$  ne contient pas assez d'ensembles mesurables et, comme expliqué en introduction,  $\mathcal{P}(E)$  en contient souvent trop pour définir des mesures raisonnables, il va falloir faire un compromis et considérer des tribus intermédiaires. Une notion clef est alors celle de tribu engendrée par une sous-partie de  $\mathcal{P}(E)$ .

**Définition 1.2.** Si  $M \subset \mathcal{P}(E)$ , la tribu engendrée par  $M$ , que l'on note  $\sigma(M)$ , est définie comme la plus petite tribu de  $\mathcal{P}(E)$  qui contient  $M$ . Plus formellement,

$$\sigma(M) := \bigcap_{\substack{\mathcal{T} \subset \mathcal{P}(E) \text{ tribu} \\ M \subset \mathcal{T}}} \mathcal{T}.$$

Pour justifier que  $\sigma(M)$  existe, notez que l'intersection (quelconque) de tribus est une tribu (vérifiez-le) et qu'il existe au moins une tribu qui contient  $M$  (laquelle?).

Si on prend par exemple  $E = \mathbb{R}$ , une tribu fréquentable contiendrait tous les ensembles de la forme  $]a, b[$  ou  $[a, b[$  ou  $]a, b]$  ou  $[a, b]$  avec  $-\infty \leq a \leq b \leq +\infty$ . Par définition d'une tribu, il suffit qu'elle contienne les ensembles ouverts  $]a, b[$  et on peut alors considérer la tribu engendrée par ces ouverts. Comme les réunions dénombrables d'ensembles de la forme  $]a, b[$  engendrent tous les ouverts de  $\mathbb{R}$ , cela revient à prendre la tribu engendrée par les ouverts de  $\mathbb{R}$ . C'est ce qu'on fait souvent dans le cadre plus général d'un espace topologique.

**Définition 1.3.** Si  $E$  est un espace topologique<sup>?</sup>, sa tribu borélienne  $\mathcal{B}(E)$  est la tribu engendrée par les ouverts de  $E$ .

Revenons maintenant à la définition de mesure discutée en introduction.

**Définition 1.4.** Une application  $\mu : \mathcal{T} \rightarrow [0, +\infty]$  est une mesure  $\mu$  sur un espace mesurable  $(E, \mathcal{T})$  si elle satisfait :

- (a)  $\mu(\emptyset) = 0$

---

?. Rappelons qu'un *espace topologique* est un ensemble  $E$  muni d'une collection  $\mathcal{O}$  d'*ouverts*, qui satisfait les axiomes :  $\emptyset, E \in \mathcal{O}$  et  $O_i \in \mathcal{O}$  pour tout  $i \in I \Rightarrow \bigcup_{i \in I} O_i \in \mathcal{O}$ , où  $I$  n'est pas forcément dénombrable.

(b)  $\mu$  est  $\sigma$ -additive : si  $A_n \in \mathcal{T}$  pour tout  $n \geq 1$  et  $A_n \cap A_m = \emptyset$  pour tout  $n \neq m$ ,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

On dit que  $(E, \mathcal{T}, \mu)$  est un **espace mesuré**.

Le résultat (non-trivial) suivant explique que l'on peut finalement définir une mesure "naturelle" sur l'espace mesurable  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**Théorème 1.5** (Existence de la mesure de Lebesgue). *Il existe une unique mesure notée **Leb** sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  telle que  $\text{Leb}([a, b]) = b - a$  pour tout  $-\infty < a \leq b < +\infty$ . On l'appelle la **mesure de Lebesgue** (de  $\mathbb{R}$ ). De plus, elle est invariante par translation : pour tout  $A \in \mathcal{B}(\mathbb{R})$  et  $x \in \mathbb{R}$  on a  $\text{Leb}(A + x) = \text{Leb}(A)$ .*

On a utilisé la notation  $A + x := \{a + x : a \in A\}$ .

Un autre exemple important de mesure est la **mesure de Dirac**  $\delta_a$ , qui est définie sur tout  $\mathcal{P}(E)$  par

$$\delta_a(A) := \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

Il est facile de vérifier que la somme de deux mesures  $\mu$  et  $\nu$  sur un espace arbitraire  $(E, \mathcal{T})$ , définie par  $(\mu + \nu)(A) := \mu(A) + \nu(A)$  pour tout  $A \in \mathcal{T}$ , est également une mesure. De même, la somme dénombrable de mesures est une mesure. Si l'espace  $E$  est discret (fini ou dénombrable), on l'équipera le plus souvent de sa **mesure de comptage**  $\mu$  définie sur tout  $\mathcal{P}(E)$  par

$$\mu := \sum_{x \in E} \delta_x.$$

On a donc  $\mu(A) = \#(E \cap A)$  pour tout  $A \in \mathcal{P}(E)$ .

### 1.3 Fonctions mesurables et intégration

Etant donné un espace mesuré  $(E, \mathcal{T}, \mu)$ , on veut donner un sens à l'intégrale  $\int f d\mu$  d'une fonction  $f$  par rapport à une mesure  $\mu$ . On commence par identifier une classe de fonctions pour lesquelles cela va être possible.

**Définition 1.6.** *Etant donné deux espaces mesurables  $(E, \mathcal{T})$  et  $(E', \mathcal{T}')$ , une fonction  $f : E \rightarrow E'$  est **mesurable** si :*

$$A \in \mathcal{T}' \quad \Rightarrow \quad f^{-1}(A) := \{x \in E : f(x) \in A\} \in \mathcal{T}. \quad (1.1)$$

Deux propriétés utiles :

- Si  $E' = \mathbb{R}$  et  $f, g$  mesurables, alors  $f + g, fg, \min(f, g), \max(f, g)$  sont mesurables.
- Si  $\mathcal{T}' = \sigma(M)$  pour un  $M \subset \mathcal{P}(E)$ , il suffit de vérifier (1.1) pour tout  $A \in M$ .

Passons maintenant à la construction de l'intégrale d'une fonction mesurable par rapport à une mesure  $\mu$ .

**Étape 1.** Pour définir l'intégrale d'une fonction on commence par décider que, si l'on note la fonction caractéristique d'un ensemble  $A$  par

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon,} \end{cases}$$

alors on définit  $\int \mathbf{1}_A d\mu := \mu(A)$  dès que  $A \in \mathcal{T}$ . Ensuite, on force l'additivité de l'intégrale (parce qu'on veut avoir la propriété  $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ ) en décidant que l'intégrale de la combinaison linéaire de fonction caractéristiques est la combinaison linéaire de leur intégrales. Plus précisément, on dit que  $f$  est une *fonction étagée* si elle est de la forme

$$f(x) = \sum_{k=1}^m v_k \mathbf{1}_{A_k}(x), \quad (1.2)$$

avec  $A_1, \dots, A_m \in \mathcal{T}$  disjoints et  $v_k \neq v_\ell$  si  $k \neq \ell$ . Remarquons que  $A_k = f^{-1}(\{v_k\})$ . Pour une telle fonction on définit :

$$\int f d\mu := \sum_{k=1}^m v_k \mu(A_k).$$

**Étape 2.** Si  $f : E \rightarrow \mathbb{R}$  est mesurable et **positive**, on définit ensuite

$$\int f d\mu := \sup \left\{ \int g d\mu \mid g : X \rightarrow \mathbb{R}_+ \text{ est étagée et } 0 \leq g \leq f \right\}. \quad (1.3)$$

Pour comprendre cette formule, imaginons que  $E = \mathbb{R}$ , choisissez votre fonction mesurable positive préférée, et dessinez son graphe. On se donne alors  $m \geq 1$  valeurs  $v_1, \dots, v_m$  strictement positives qu'on dispose sur l'axe des ordonnées et on trace les droites  $\mathcal{D}_k$  d'équation  $y = v_k$  qui intersectent le graphe de  $f$ . On prend  $A_k$  l'ensemble des points où le graphe de  $f$  passe entre  $\mathcal{D}_k$  et  $\mathcal{D}_{k+1}$ , c'est-à-dire  $A_k := \{x \in E : v_k \leq f(x) < v_{k+1}\}$  (avec  $v_0 := 0$  et  $v_{m+1} := +\infty$ ). On voit donc que la fonction étagée  $g := \sum_{k=1}^m v_k \mathbf{1}_{A_k}$  est positive et que  $g \leq f$ . En gros, on a découpé le graphe de  $f$  en tranches *horizontales* et on a pris la plus grande fonction étagée sous  $f$  qui vit sur ces droites horizontales. On prend alors pour  $\int f d\mu$  la plus grande valeur de l'intégrale  $\int g d\mu$  après avoir fait varier  $v_1, \dots, v_m > 0$  (les hauteurs des droites) et  $m \geq 1$  (le nombre de droites) de toutes les façons possibles.

**Étape 3.** Finalement, si  $f : E \rightarrow \mathbb{R}$  est mesurable mais pas forcément positive, on note sa partie positive  $f_+ := \max(f, 0)$  et sa partie négative  $f_- := \max(-f, 0)$ , de façon à avoir la décomposition  $f = f_+ - f_-$  avec  $f_+$  et  $f_-$  des fonctions positives. Si

$$\int f_+ d\mu < \infty, \quad \int f_- d\mu < \infty,$$

alors on dit que  $f$  est *absolument intégrable*, ce qu'on note  $f \in L^1(\mu)$ , et on définit

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu.$$

**Remarque 1.7.** Comme  $|f| = f_+ + f_-$ , on voit que  $f : E \rightarrow \mathbb{R}$  est absolument intégrable si et seulement si  $\int |f| d\mu < \infty$ .

**Notations :** On notera indifféremment  $\int f d\mu$ , ou  $\int f(x)d\mu(x)$ , ou  $\int f(x)\mu(dx)$ . Quand on intègre par rapport à la Lebesgue sur  $\mathbb{R}$ , on note simplement  $\int f dx$  au lieu de  $\int f d\text{Leb}$ .

Quelques propriétés élémentaires mais clefs de l'intégrale qu'on vient de construire :

**Proposition 1.8.** Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré.

— **(Linéarité de l'intégrale)** Si  $f, g \in L^1(\mu)$  et  $\alpha, \beta \in \mathbb{R}$ , alors  $\alpha f + \beta g \in L^1(\mu)$  et

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

— **(Positivité de l'intégrale)**  $f : E \rightarrow \mathbb{R}$  est mesurable et  $f \geq 0 \Rightarrow \int f d\mu \geq 0$ .

Cela entraîne que si  $f, g \in L^1(\mu)$  et  $f \geq g$ , alors  $\int f d\mu \geq \int g d\mu$ . En particulier, cela montre que si  $f$  est étagée alors le supremum dans (1.3) est atteint en  $f$  et donc les définitions de  $\int f d\mu$  de l'étape 1 et de l'étape 2 coïncident.

On a aussi une propriété évidente de "linéarité de l'intégrale par rapport aux mesures" : Si  $\nu$  est une autre mesure sur  $(E, \mathcal{T})$  et  $\alpha, \beta > 0$ , alors  $(\alpha\mu + \beta\nu)(A) := \alpha\mu(A) + \beta\nu(A)$  définit aussi une mesure sur  $E$  et, pour toute fonction  $f \in L^1(\mu) \cap L^1(\nu)$ , on a

$$\int f d(\alpha\mu + \beta\nu) = \alpha \int f d\mu + \beta \int f d\nu.$$

**Lien avec l'intégrale de Riemann.** Si  $f : [a, b] \rightarrow \mathbb{R}$  est intégrable au sens de Riemann, alors  $f$  est intégrable pour la mesure de Lebesgue  $\clubsuit$  et les deux intégrales coïncident. Rappelons qu'une fonction continue est Riemann-intégrable. En particulier on peut utiliser toute l'artillerie des résultats de l'intégration de Riemann, comme le théorème fondamental de l'analyse  $?$  ou l'intégration par parties.

**Fonctions à valeurs dans  $\mathbb{R}^d$  ou  $\mathbb{C}$ .** Si  $f : E \rightarrow \mathbb{R}^d$  s'écrit  $f(x) = {}^t(f_1(x), \dots, f_d(x))$  et que chaque entrée  $f_j$  est mesurable positive, ou intégrable, alors on étend la définition de l'intégrale en posant

$$\int f d\mu := {}^t \left( \int f_1 d\mu, \dots, \int f_d d\mu \right).$$

En identifiant  $\mathbb{C}$  avec  $\mathbb{R}^2$  via  $x + iy \leftrightarrow {}^t(x, y)$ , on définit ainsi l'intégrale d'une fonction  $f = \Re(f) + i\Im(f)$  à valeurs dans  $\mathbb{C}$  par

$$\int f d\mu = \int \Re(f) d\mu + i \int \Im(f) d\mu.$$

– Fin du cours 1 –

$\spadesuit$ . on peut même montrer que  $f$  est continu sauf éventuellement sur un ensemble dénombrable.

$?$ . dont la version simplifiée dit que, si  $f : [a, b] \rightarrow \mathbb{R}$  est continue, alors  $F(x) := \int_a^x f(x) dx$  est dérivable sur  $]a, b[$  de dérivée  $f$ , et que les primitives de  $f$  sont égales à  $F$  à une constante additive près.

## 1.4 Ensembles négligeables et densités

Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré.

**Définition 1.9.** Si  $A \in \mathcal{T}$  est tel que  $\mu(A) = 0$ , on dit que  $A$  est  $\mu$ -négligeable. Une propriété est vraie  $\mu$ -presque partout (abrégé  $\mu$ -p.p.) si elle est vraie à un ensemble  $\mu$ -négligeable près.

Par exemple, si  $f$  et  $g$  sont deux fonctions mesurables,  $f = g$   $\mu$ -p.p. quand

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0.$$

**Exemple 1.10.** Montrons que si  $f : E \rightarrow \mathbb{R}$  mesurable positive est telle que  $S := \{x \in E : f(x) \neq 0\}$  et  $\mu(S) = 0$ , alors  $\int f \, d\mu = 0$ . En effet, si  $f$  est de plus bornée, c'est à dire  $\|f\|_\infty := \sup_{x \in E} |f(x)| < \infty$ , alors

$$0 \leq \int f \, d\mu = \underbrace{\int \mathbf{1}_S f \, d\mu}_{\leq \|f\|_\infty \underbrace{\mu(S)}_{=0}} + \underbrace{\int \mathbf{1}_{S^c} f \, d\mu}_{=0} \leq 0.$$

Comme est une fonction étagée est mesurable et bornée, pour une fonction  $f$  mesurable quelconque, on a

$$\int f \, d\mu = \sup$$

**Mesures à densité.** Si  $f : E \rightarrow \mathbb{R}_+$  est mesurable, alors

$$\nu(A) := \int \mathbf{1}_A f \, d\mu$$

définit une mesure sur  $(E, \mathcal{T})$  et pour toute fonction mesurable  $h : E \rightarrow \mathbb{R}_+$ , on a

$$\int h \, d\nu = \int h f \, d\mu.$$

On note souvent  $d\nu = f \, d\mu$  et on dit que  $f$  est la *densité de  $\nu$  par rapport à  $\mu$* , que l'on note aussi  $f = \frac{d\nu}{d\mu}$ . Si  $E = \mathbb{R}$  et  $\mu$  est la mesure de Lebesgue, on écrit tout simplement  $d\nu = f \, dx$ . Notez que si  $\mu(A) = 0$  alors  $\mathbf{1}_A = 0$   $\mu$ -p.p., donc  $f \mathbf{1}_A = 0$   $\mu$ -p.p., et finalement on a montré que pour tout  $A \in \mathcal{T}$ ,

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0.$$

On dit alors que  $\nu$  est *absolument continue* par rapport à  $\mu$ . Il est remarquable que la réciproque soit vraie lorsque l'on suppose que les mesures  $\mu$  et  $\nu$  sont  $\sigma$ -finies.

**Définition 1.11.** Une mesure  $\mu$  est *finie* si  $\mu(E) < \infty$ . Elle est  *$\sigma$ -finie* si il existe une suite  $E_n \in \mathcal{T}$  telle que  $E = \cup_n E_n$  et  $\mu(E_n) < \infty$  pour tout  $n \geq 1$ .

Par exemple la mesure de Lebesgue sur  $\mathbb{R}$  n'est pas finie mais elle est  $\sigma$ -finie. De même pour la mesure de comptage de  $\mathbb{N}$ .



**Proposition 1.12** (Existence de densité). *Si  $\mu, \nu$  sont des mesures sur  $(E, \mathcal{F})$   $\sigma$ -finies et  $\nu$  est absolument continue par rapport à  $\mu$ , alors il existe une fonction mesurable  $f : E \rightarrow \mathbb{R}_+$ , unique à un ensemble  $\mu$ -négligeable près, telle que  $d\nu = f d\mu$ .*

La fonction  $f$  du précédent théorème est parfois appelée la *dérivée de Radon-Nikodym de  $\nu$  par rapport à  $\mu$* .

Voici un théorème de structure qui décrit toutes les mesures  $\sigma$ -finies sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  comme la somme d'une partie à densité et d'une partie singulière.

**Théorème 1.13** (Radon-Nikodym-Lebesgue ; cas particulier). *Si  $\mu$  est une mesure  $\sigma$ -finie sur  $\mathbb{R}$  alors il existe  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  mesurable et une mesure  $\eta$  sur  $\mathbb{R}$  telles que*

$$\mu = f dx + \eta$$

où la mesure  $\eta$  est singulière à la mesure de Lebesgue : Il existe  $S \in \mathcal{B}(\mathbb{R})$  de mesure de Lebesgue nulle tel que, pour tout  $A \in \mathcal{B}(\mathbb{R})$ , on a  $\eta(A) = \eta(S \cap A)$ .

Par exemple, on peut prendre pour  $\eta$  une mesure discrète, c'est à dire de la forme

$$\eta = \sum_{k \in \mathbb{N}} \alpha_k \delta_{x_k}$$

où  $\alpha_k > 0$  et  $x_k \in \mathbb{R}$  pour tout  $k \in \mathbb{N}$ . En effet, on voit que cette mesure est singulière en prenant  $S = \{x_k : k \in \mathbb{N}\}$  qui satisfait bien  $\text{Leb}(S) = 0$ . Il existe d'autres mesures singulières qui ne sont pas discrètes, comme la "mesure uniforme" sur un ensemble de Cantor, mais ces mesures n'apparaissent essentiellement jamais dans les applications.

## 1.5 Intégration et limites de fonctions

**Théorème 1.14.** *Soit  $(E, \mathcal{F}, \mu)$  un espace mesuré et  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions mesurables  $E \rightarrow \mathbb{R}$ .*

- **(Convergence monotone)** *Si les fonctions  $f_n$  sont positives et la suite  $(f_n)_{n \in \mathbb{N}}$  est croissante  $\mu$ -p.p., c'est-à-dire  $0 \leq f_1(x) \leq f_2(x) \leq \dots$  pour  $\mu$ -presque tout  $x \in E$ , alors*

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n(x) d\mu.$$

- **(Convergence dominée)** *Si, pour  $\mu$ -presque tout  $x \in E$ ,  $f_n(x)$  a une limite  $f(x)$  quand  $n \rightarrow \infty$  et  $|f_n(x)| \leq g(x)$  pour une fonction  $g \in L^1(\mu)$ , alors  $f_n \in L^1(\mu)$  et*

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n(x) d\mu = \int f d\mu.$$

De plus, on a la convergence dans  $L^1(\mu)$ ,

$$\int |f_n - f| d\mu \xrightarrow{n \rightarrow \infty} 0.$$

## 1.6 Mesures produit

Soit  $(E_1, \mathcal{T}_1, \mu_1)$  et  $(E_2, \mathcal{T}_2, \mu_2)$  deux espaces mesurés. On veut définir une mesure naturelle sur le produit  $E_1 \times E_2$ . Pour se faire, on équipe  $E_1 \times E_2$  de la tribu engendrée par  $\mathcal{T}_1$  et  $\mathcal{T}_2$  que l'on note  $\mathcal{T}_1 \otimes \mathcal{T}_2$ . C'est à dire

$$\mathcal{T}_1 \otimes \mathcal{T}_2 := \sigma\left(\{A_1 \times A_2 : A_1 \in \mathcal{T}_1, A_2 \in \mathcal{T}_2\}\right).$$

**Proposition 1.15** (Existence de mesures produits). *Si  $\mu_1$  et  $\mu_2$  sont des mesures  $\sigma$ -finies, alors il existe une unique mesure sur  $(E_1 \times E_2, \mathcal{T}_1 \otimes \mathcal{T}_2)$ , noté  $\mu_1 \otimes \mu_2$ , telle que pour tout  $A_1 \in \mathcal{T}_1$  et  $A_2 \in \mathcal{T}_2$ ,*

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1) \mu_2(A_2).$$

Par exemple, en prenant  $E_1 = E_2 = \mathbb{R}$ ,  $\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{B}(\mathbb{R})$  et  $\mu_1 = \mu_2 = \text{Leb}$ , on obtient une mesure  $\text{Leb}^{\otimes 2}$  sur  $\mathbb{R}^2$  qui satisfait

$$\text{Leb}^{\otimes 2}([a_1, b_1] \times [a_2, b_2]) = (b_1 - a_1)(b_2 - a_2),$$

qui correspondant bien à la surface d'un rectangle. Par itération, on obtient une mesure  $\text{Leb}^{\otimes d}$  sur  $\mathbb{R}^d$  qui satisfait

$$\text{Leb}^{\otimes d}([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d).$$

On appelle  $\text{Leb}^{\otimes d}$  la *mesure de Lebesgue de  $\mathbb{R}^d$*  et on remarque que  $\mathcal{B}(\mathbb{R})^{\otimes d} = \mathcal{B}(\mathbb{R}^d)$ , puisque les ouverts de  $\mathbb{R}^d$  sont engendrés par les produits d'ouverts de  $\mathbb{R}$ .

**Théorème 1.16.** *Soit  $\mu_1, \mu_2$  des mesures  $\sigma$ -finies et  $f : E_1 \times E_2 \rightarrow \mathbb{R}$  mesurable.*

— **(Fubini-Tonelli)** *Si  $f$  est positive, alors*

$$\int f \, d\mu_1 \otimes \mu_2 = \int \left( \int f(x, y) \, d\mu_1(x) \right) d\mu_2(y) = \int \left( \int f(x, y) \, d\mu_2(y) \right) d\mu_1(x),$$

*et toutes les quantités présentes sont bien définies.*

— **(Fubini-Lebesgue)** *Si  $f \in L^1(\mu_1 \otimes \mu_2)$ , alors la même conclusion s'applique.*

Le théorème de Fubini-Tonelli nous donne que la condition  $f \in L^1(\mu_1 \otimes \mu_2)$  s'écrit

$$\int \left( \int |f(x, y)| \, d\mu_1(x) \right) d\mu_2(y) < \infty \quad \text{ou} \quad \int \left( \int |f(x, y)| \, d\mu_2(y) \right) d\mu_1(x) < \infty.$$

## 1.7 Changement de variables

Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré et  $(E', \mathcal{T}')$  un espace mesurable. Etant donné une application  $\varphi : E \rightarrow E'$  mesurable, la *mesure image*  $\varphi_*\mu$ , définie par  $\varphi_*\mu(A) := \mu(\varphi^{-1}(A))$  pour tout  $A \in \mathcal{T}'$ , est une mesure sur  $(E', \mathcal{T}')$  et on a pour toute fonction mesurable  $f : E' \rightarrow \mathbb{R}_+$  (cf. Exercice 5),

$$\int f \circ \varphi(y) \, d\mu(y) = \int f(x) \, d\varphi_*\mu(x).$$

Cette dernière formule est une formule de changement de variable généralisée. Malheureusement il n'existe pas de formule générale pour  $d\varphi_*\mu(x)$  et il faut travailler au cas par cas.

**Cas de la mesure de Lebesgue de  $\mathbb{R}^d$ .** Si  $\mu = \text{Leb}^{\otimes d}$  et si  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  est une application bijective dont la matrice jacobienne  $J_\varphi(x)$  est inversible pour tout  $x \in \mathbb{R}^d$ , c'est-à-dire si

$$\det J_\varphi(x) := \det [\partial_j \varphi_i(x)] \neq 0$$

pour tout  $x \in \mathbb{R}^d$  (on peut affaiblir toutes ces hypothèses), alors il s'avère que  $\varphi_*\mu$  a une densité par rapport à  $\text{Leb}^{\otimes d}$  qui est explicite (on a noté  $\varphi_i(x)$  la  $i$ -ième coordonnée de  $\varphi(x)$  et  $\partial_j$  la dérivée par rapport à la  $j$ -ième variable). En effet, dans ce cas on a pour tout  $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$  mesurable :

$$\int f \circ \varphi(y) dy = \int f(x) |\text{Jac}_{\varphi^{-1}}(x)| dx.$$

Attention, notez bien que ce n'est pas le Jacobien de  $\varphi$  mais bien celui de son inverse  $\varphi^{-1}$  qui apparaît ; on a donc  $d\varphi_*\text{Leb}(x) = |\text{Jac}_{\varphi^{-1}}(x)| dx$ . Notez aussi que l'existence de  $\text{Jac}_{\varphi^{-1}}$  est garantie par le théorème d'inversion locale. On peut utiliser cette dernière formule avec  $\varphi^{-1}$  à la place de  $\varphi$  pour obtenir :

$$\int f \circ \varphi^{-1}(y) dy = \int f(x) |\text{Jac}_\varphi(x)| dx. \quad (1.4)$$

Si on revient aux ensembles, en prenant  $f = \mathbf{1}_A$  où  $A \in \mathcal{B}(\mathbb{R}^d)$ , et que l'on suppose que  $\varphi$  est une application linéaire inversible (si bien que son Jacobien est constant et égal à  $\det \varphi \neq 0$ ), alors on voit que

$$\text{Leb}(\varphi(A)) = |\det \varphi| \text{Leb}(A).$$

Cette formule, qui décrit comment les volumes sont modifiés après une transformation linéaire, est l'essence même de la formule générale (1.4).

– *Fin du cours 2* –

## 1.8 EXERCICES – Théorie de la mesure

*Remarque préliminaire* :  $\mathbb{R}$  sera ici toujours équipé de sa tribu borélienne  $\mathcal{B}(\mathbb{R})$ .

**Exercice 1.** Soit  $(E, \mathcal{F}, \mu)$  un espace mesuré. Montrer que :

(a) Pour tout  $A, B \in \mathcal{F}$ , on a

$$\mu(A \cup B) \leq \mu(A) + \mu(B)$$

et, si on suppose de plus que  $\mu(E) < \infty$ ,

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B).$$

(b) Si  $A_n \in \mathcal{F}$  et  $A_n \subset A_{n+1}$  pour tout  $n \geq 1$ , alors

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu\left(\bigcup_{n \geq 1} A_n\right).$$

**Exercice 2.** On veut montrer que toute fonction étagée est mesurable. Soit  $(E, \mathcal{F})$  un espace mesurable et  $A \in \mathcal{P}(E)$ . Montrer que la fonction caractéristique  $\mathbf{1}_A : E \rightarrow \mathbb{R}$  est mesurable  $\Leftrightarrow A \in \mathcal{F}$ . Conclure.

**Exercice 3.** (a) Montrer que  $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$  et calculer sa mesure de Lebesgue.

(b) On considère la fonction  $\mathbf{1}_{\mathbb{Q} \cap [0,1]}$ . Quelle est son intégrale pour la mesure de Lebesgue ? Que peut-on dire de son intégrale de Riemann ?

**Exercice 4.** Soit  $E$  un ensemble et  $a \in E$ . La *masse de Dirac en  $a$*  est l'application définie sur  $\mathcal{P}(E)$  par

$$\delta_a(A) := \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

(a) Montrer que  $\delta_a$  est une mesure (on l'appelle aussi la *mesure de Dirac*).

(b) Montrer que pour toute fonction  $f : E \rightarrow \mathbb{R}_+$ , on a

$$\int f \delta_a = f(a).$$

(c) On équipe  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  de sa mesure de comptage  $\mu$  définie par

$$\mu(A) = \sum_{k=0}^{\infty} \delta_k(A).$$

Montrer que toute fonction  $f : \mathbb{N} \rightarrow \mathbb{R}_+$  est mesurable et que

$$\int f \, d\mu = \sum_{k=0}^{\infty} f(k).$$

**Exercice 5.** Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré et  $(E', \mathcal{T}')$  un espace mesurable. On se donne une application  $\varphi : E \rightarrow E'$  mesurable.

- (a) Montrer que la *mesure image*  $\varphi_*\mu$  définie par  $\varphi_*\mu(A) := \mu(\varphi^{-1}(A))$  pour tout  $A \in \mathcal{T}'$  est bien une mesure sur  $(E', \mathcal{T}')$ .  
 (b) Montrer que pour toute fonction mesurable  $f : E' \rightarrow \mathbb{R}_+$  on a :

$$\int_{E'} f(x) d\varphi_*\mu(x) = \int_E f \circ \varphi(y) d\mu(y).$$

- (c) Si  $(E, \mathcal{T}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$  avec  $\mu$  la mesure de Lebesgue et  $\varphi(x) = x^3$ , donner une forme explicite à  $\varphi_*\mu$ . Même question si  $(E, \mathcal{T}, \mu)$  est maintenant l'espace mesuré de l'exercice 4(c).

**Exercice 6.** Soit  $(E, \mathcal{T}, \mu)$  un espace mesuré et  $f : E \rightarrow \mathbb{R}_+$  un fonction mesurable. On considère la mesure  $\nu$  définie par

$$\nu(A) := \int \mathbf{1}_A f d\mu.$$

Montrer que  $\nu$  est une mesure sur  $(E, \mathcal{T}, \mu)$ .

**Exercice 7.** On considère la relation d'équivalence sur  $[0, 1]$  donnée par  $x \sim y \Leftrightarrow x - y \in \mathbb{Q}$ . On note  $[x]$  la classe d'équivalence associée à  $x \in [0, 1]$  pour cette relation et  $\mathcal{C}$  l'ensemble des classes d'équivalence.

- (a) Montrer que les classes d'équivalences forment une partition de  $[0, 1]$ , c'est-à-dire que  $[x] \cap [y] \neq \emptyset \Leftrightarrow x \sim y$  et

$$\bigcup_{[x] \in \mathcal{C}} [x] = [0, 1].$$

Pour tout  $[x] \in \mathcal{C}$ , on choisit un élément  $p_{[x]} \in [x]$  de façon arbitraire et on considère l'ensemble  $V = \{p_{[x]} : [x] \in \mathcal{C}\}$  (le fait que  $V$  soit un ensemble bien défini requiert l'Axiome du choix).

- (b) Montrer qu'on a les inclusions d'ensembles

$$[0, 1] \subset \bigcup_{q \in [-1, 1] \cap \mathbb{Q}} V + q \subset [-1, 2].$$

- (c) En utilisant la propriété d'invariance par translation de la mesure de Lebesgue, déduire que  $V \notin \mathcal{B}(\mathbb{R})$ .

**Exercice 8.** Soit la mesure sur  $\mathbb{R}$  définie par  $\mu = \frac{1}{2}\mathbf{1}_{[0,1]}dx + \frac{1}{2}\delta_0$ . Montrer que c'est une mesure de probabilité et calculer  $\int x d\mu(x)$ .

**Exercice 9.** Donner un exemple de mesure qui n'est pas  $\sigma$ -finie.

**Exercice 10.** On considère la mesure  $\mu$  sur  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  définie par

$$\int f d\mu = \frac{1}{2\pi} \int_0^{2\pi} f(\cos \theta, \sin \theta) d\theta$$

pour toute fonction  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  mesurable. Est-ce que cette mesure est absolument continue par rapport à la mesure de Lebesgue de  $\mathbb{R}^2$  ?

**Exercice 11.** Soit  $(E, \mathcal{F}, \mu)$  un espace mesuré et  $f_n : E \rightarrow \mathbb{R}$  une suite de fonctions mesurables. Sous quelles conditions a-t-on

$$\sum_{n=1}^{\infty} \int f_n \, d\mu = \int \sum_{n=1}^{\infty} f_n \, d\mu \quad ?$$

**Exercice 12.** Démontrez le résultat suivant :

Soit  $I \subset \mathbb{R}$  un ouvert et  $f : E \times I \rightarrow \mathbb{R}$  telle que

1.  $x \mapsto f(x, t) \in L^1(\mu)$  pour tout  $t \in I$ ,
2.  $\partial_t f(x, t)$  existe pour tout  $t \in I$  et  $\mu$ -presque tout  $x \in E$ ,
3. il existe  $g \in L^1(\mu)$  tel que  $|\partial_t f(x, t)| \leq g(x)$  pour tout  $t \in I$ .

Alors, pour tout  $t \in I$ ,

$$\frac{d}{dt} \int f(x, t) \, \mu(dx) = \int \partial_t f(x, t) \, \mu(dx).$$

*Aide :* On pourra utiliser l'identité (qu'on démontrera)

$$\frac{f(x, t + \varepsilon) - f(x, t)}{\varepsilon} = \int_0^1 \partial_t f(x, t + u\varepsilon) \, du.$$

## 2 Probabilités : Boîte à outils

### 2.1 Variables aléatoires

Pour modéliser un évènement dont l'issue est incertaine, on s'appuiera sur un espace probabilisé que l'on a l'habitude de noter  $(\Omega, \mathcal{F}, \mathbb{P})$ . On peut le comprendre ainsi :

- $\Omega$  : l'espace de toutes les réalisations possibles (l'univers).
- $\mathcal{F}$  : l'ensemble de toutes les questions qui ont un sens.
- $\mathbb{P}$  : la mesure qui donne à chaque question une probabilité de réalisation.

Cet espace fera office d'outil pour faire marcher la théorie mais ne sera que rarement explicite. Quand une propriété est vraie pour  $\mathbb{P}$ -presque tout  $\omega \in \Omega$ , c'est-à-dire avec probabilité un, on dira plutôt qu'elle est vraie *presque sûrement* (abrégé *p.s*) ou *avec probabilité un*.

**Définition 2.1.** Une *variable aléatoire*  $X$  à valeurs dans un espace mesuré  $(E, \mathcal{T})$  est une application mesurable  $X : \Omega \rightarrow E$ .

Si  $A \in \mathcal{T}$ , on notera  $\mathbb{P}(X \in A)$  plutôt que  $\mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$ ; notez que ces quantités sont bien définies car  $X$  est mesurable par définition. On utilisera la notation, pour tout  $f : E \rightarrow \mathbb{R}$  mesurable positive ou absolument intégrable,

$$\mathbb{E}[f(X)] := \int f \circ X(\omega) d\mathbb{P}(\omega).$$

**Définition 2.2.** La loi  $\mu_X$  d'une variable aléatoire  $X$  est la mesure image  $X_*\mathbb{P}$ . En d'autres termes, pour tout  $A \in E$  on a  $\mathbb{P}(X \in A) = \mu_X(A)$  et, pour tout  $f : E \rightarrow \mathbb{R}_+$  mesurable,

$$\mathbb{E}[f(X)] = \int f \circ X(\omega) d\mathbb{P}(\omega) = \int f(x) d\mu_X(x).$$

Une façon pratique de caractériser la loi d'une variable *réelle*  $X$  est de considérer sa *fonction de répartition*,

$$F_X(t) := \mathbb{P}(X \leq t) = \mu_X([-\infty, t]), \quad t \in \mathbb{R},$$

ou encore sa transformée de Fourier (à un signe près), qu'on appelle en probabilités plutôt la *fonction caractéristique* de  $X$ ,

$$\varphi_X(t) := \mathbb{E}[e^{itX}] = \int e^{itx} d\mu_X(x).$$

En effet ces deux transformations de la mesure  $\mu_X$ , qui ne dépendent plus que d'un paramètre  $t$  et non de toute une classe de fonctions  $f$ , caractérisent les lois des variables, c'est-à-dire que si  $F_X(t) = F_Y(t)$ , ou  $\varphi_X(t) = \varphi_Y(t)$ , pour tout  $t \in \mathbb{R}$ , alors  $\mu_X = \mu_Y$ . Dans ce cas on dit que  $X$  et  $Y$  sont *égales en loi*.

Si la fonction de répartition n'admet pas de généralisation en dimension supérieure, ce n'est pas le cas de la fonction caractéristique d'un vecteur aléatoire  $X$  de  $\mathbb{R}^d$ ,

$$\varphi_X(t) := \mathbb{E}[e^{i\langle t, X \rangle}] = \int e^{i\langle t, x \rangle} d\mu_X(x), \quad t \in \mathbb{R}^d,$$

où  $\langle t, x \rangle = \mathbf{t}tx = \sum_{i=1}^d t_i x_i$  est le produit scalaire usuel de  $\mathbb{R}^d$ , qui caractérise également la loi de  $X$ .

**Définition 2.3.** Deux variables aléatoires  $X_1 : \Omega \rightarrow (E_1, \mathcal{T}_1)$  et  $X_2 : \Omega \rightarrow (E_2, \mathcal{T}_2)$  sont **indépendantes** si, pour tout  $A_1 \in \mathcal{T}_1$  et  $A_2 \in \mathcal{T}_2$ , on a :

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A_2).$$

De façon équivalente,

$$\begin{aligned} & X_1 \text{ et } X_2 \text{ sont indépendantes} \\ \Leftrightarrow & \mathbb{E}[f(X_1)g(X_2)] = \mathbb{E}[f(X_1)]\mathbb{E}[g(X_2)] \text{ pour tout } f, g \text{ mesurables positives} \\ \Leftrightarrow & \mu_{(X_1, X_2)} = \mu_{X_1} \otimes \mu_{X_2} \\ \Leftrightarrow & \varphi_{(X_1, X_2)}(t) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2) \text{ pour tout } t = (t_1, t_2) \in \mathbb{R}^2. \end{aligned}$$

Si  $X_1, \dots, X_n$  est une suite de variables de même loi  $\mu$  et que  $X_i$  et  $X_j$  sont indépendantes pour tout  $i \neq j$ , on dira alors que  $X_1, \dots, X_n$  est une suite de variables indépendantes identiquement distribuées, abrégé *i.i.d.*, de loi  $\mu$ .

## 2.2 Variables aléatoires réelles

On dira que  $X$  est une **variable aléatoire réelle** si elle est à valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et, si  $p \geq 1$ , on écrit  $X \in L^p$  si

$$\mathbb{E}[|X|^p] = \int |X(\omega)|^p d\mathbb{P}(\omega) = \int |x|^p d\mu_X(x) < \infty.$$

L'**espérance** ou la **moyenne** de  $X \in L^1$  est définie par

$$\mathbb{E}[X] := \int X(\omega) d\mathbb{P}(\omega) = \int x d\mu_X(x).$$

On définit aussi sa **variance** par

$$\text{Var}[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

dès que  $X \in L^2$ , qui est une mesure de la dispersion de la variable  $X$  autour de sa moyenne  $\mathbb{E}[X]$ , et son **écart-type** par la racine carrée de la variance,  $\sigma_X := \text{Var}[X]^{1/2}$ . Remarquez que  $X \mapsto \text{Var}(X)$  est une forme quadratique de forme bilinéaire associée

$$\text{Cov}(X, Y) := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

qu'on appelle la **covariance** de deux variables aléatoires  $X, Y \in L^2$ . On a d'ailleurs l'inégalité de Cauchy-Schwarz,

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)},$$

qui montre que la **correlation** entre  $X$  et  $Y$  satisfait

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \in [-1, 1].$$

Notez que si  $X$  et  $Y$  sont indépendantes alors  $\text{Cov}(X, Y) = 0$  mais il faut garder en tête que la réciproque est fautive.

Présentons rapidement quelques lois usuelles qu'on ne peut contourner en statistiques.



**Loi gaussiennes**  $\mathcal{N}(\mu, \sigma^2)$ . On dit que  $X$  est une *variable gaussienne* ou *normale* d'espérance (ou de moyenne)  $m$  et de variance  $\sigma^2$ , ce qu'on écrit de façon abrégé  $X \sim \mathcal{N}(m, \sigma^2)$ , si la loi de  $X$  a une densité par rapport à Leb donnée par

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

On voit qu'une variable gaussienne est complètement caractérisée par  $\mathbb{E}[X] = m$  et  $\text{Var}(X) = \sigma^2$ . Aussi, si  $X \sim \mathcal{N}(0, 1)$ , ce qu'on appelle une *variable gaussienne standard*, alors

$$\sigma X + \mu \sim \mathcal{N}(m, \sigma^2). \quad (2.1)$$

On inclut implicitement dans la définition qu'une variable gaussienne de moyenne  $m$  et de variance  $\sigma^2 = 0$  c'est la variable aléatoire de loi  $\delta_m$ , c'est-à-dire constante p.s. égale à  $m$ .

**Lois Gamma**  $\Gamma(k, \theta)$ . On dit que  $X$  suit une *loi Gamma* de paramètre de forme  $k > 0$  et de taux  $\theta > 0$  (ou d'échelle  $\lambda = 1/\theta$ ), et on écrit  $X \sim \Gamma(k, \theta)$ , si la loi de  $X$  a une densité par rapport à Leb donnée par

$$f(x) = \frac{\theta^k}{\Gamma(k)} x^{k-1} e^{-x\theta} \mathbf{1}_{[0, +\infty[}(x),$$

où  $\Gamma(k)$  est la fonction Gamma d'Euler<sup>?</sup>. Ces lois sont stables par addition quand  $\theta$  est fixé : On peut montrer que

$$X \sim \Gamma(k, \theta), \quad Y \sim \Gamma(\ell, \theta), \quad X, Y \text{ indépendantes} \quad \Rightarrow \quad X + Y \sim \Gamma(k + \ell, \theta).$$

Quelques cas particuliers :

- Quand  $k = 1$ , on dit que  $X$  suit une *loi exponentielle* de paramètre  $\theta$  et  $X \sim \mathcal{E}(\theta)$ . Elle a la propriété dite d'*absence de mémoire*  $P(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$  et intervient souvent dans la modélisation de durées de vie (de composants électroniques, d'atomes radioactifs, etc), où le problème est de retrouver le paramètre  $\theta$ .
- Quand  $k = d/2$  et  $\theta = 1/2$ ,  $X$  suit une *loi du  $\chi^2$*  à  $d$  degrés de liberté et  $X \sim \chi_d^2$ . Il s'avère que c'est la loi de la norme euclidienne au carré d'un vecteur gaussien standard : Si  $X_1, \dots, X_d$  sont i.i.d de loi  $\mathcal{N}(0, 1)$ , alors  $X_1^2 + \dots + X_d^2 \sim \chi_d^2$ . Cette loi est utilisée de façon clef dans les test du  $\chi^2$  (test d'adéquation, d'homogénéité et d'indépendance).

## 2.3 Vecteurs aléatoires

Si  $X$  est à valeur dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , on dira que  $X$  est un *vecteur aléatoire*. Si  $X = {}^t(X_1, \dots, X_d)$  avec  $X_j \in L^1$  pour tout  $j$ , on définit son espérance comme le vecteur des espérances de ses entrées,  $\mathbb{E}[X] := {}^t(\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$ , et sa *matrice de covariance* par

$$\Sigma_X := \left[ \text{Cov}(X_i, X_j) \right]_{i,j=1}^d$$

---

?.  $\Gamma(u) := \int_0^\infty x^{u-1} e^{-x} dx$  pour  $u > 0$ . Elle satisfait l'équation  $\Gamma(u + 1) = u\Gamma(u)$  et en particulier  $\Gamma(u + 1) = u!$  si  $u \in \mathbb{N}$  car  $\Gamma(1) = 1$ ; c'est donc une extension continue de la factorielle. On a aussi la formule utile  $\Gamma(1/2) = \sqrt{\pi}$ .

si  $X_j \in L^2$  pour tout  $j$ . Notez que  $\Sigma_X$  est une matrice symétrique semi-définie positive. Il s'avère que toute matrice symétrique semi-définie positive est la matrice de covariance d'un vecteur aléatoire, ce qu'on l'on peut vérifier avec un vecteur gaussien.

On dit que  $X$  est un *vecteur gaussien standard* de  $\mathbb{R}^d$ , et on écrit  $X \sim \mathcal{N}(0, I_d)$ , si  $X = {}^t(X_1, \dots, X_d)$  avec  $X_1, \dots, X_d$  des variables i.i.d  $\mathcal{N}(0, 1)$ . Plus généralement, pour tout matrice symétrique semi-définie positive  $\Sigma$  et  $m \in \mathbb{R}^d$ , le vecteur aléatoire

$$\Sigma^{1/2}X + m$$

où  $X \sim \mathcal{N}(0, I_d)$  est un *vecteur gaussien*  $\mathcal{N}(m, \Sigma)$ . Ici,  $\Sigma^{1/2}$  est n'importe quelle matrice qui satisfait

$$\Sigma^{1/2} \cdot {}^t\Sigma^{1/2} = \Sigma,$$

et la loi d'un vecteur  $\mathcal{N}(m, \Sigma)$  ne dépend pas du choix spécifique de  $\Sigma^{1/2}$ . Une telle matrice peut par exemple être obtenue par diagonalisation  $\clubsuit$ . Un algorithme classique qui fournit  $\Sigma^{1/2}$  est l'algorithme de Cholesky. On fera le lien avec (2.1).

**Proposition 2.4.** (a)  $X$  est un vecteur gaussien  $\Leftrightarrow$  pour tout  $\alpha = {}^t(\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$ , la variable réelle  $\langle X, \alpha \rangle = \alpha_1 X_1 + \dots + \alpha_d X_d$  est gaussienne.

(b) Pour toute matrice  $A$  de taille appropriée, on a :

$$X \sim \mathcal{N}(m, \Sigma) \quad \Rightarrow \quad AX \sim \mathcal{N}(Am, A\Sigma {}^tA).$$

(c) Si  $\Sigma$  est inversible, alors  $X \sim \mathcal{N}(m, \Sigma)$  a une densité par rapport à  $\text{Leb}^{\otimes d}$ ,

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-{}^t(x-m)\Sigma^{-1}(x-m)/2}.$$

(d) Si  $X = {}^t(X_1, \dots, X_n)$  est un vecteur gaussien, alors

$$X_i \text{ et } X_j \text{ sont indépendants} \quad \Leftrightarrow \quad \text{Cov}(X_i, X_j) = 0.$$

Remarquez que  $\langle X, \alpha \rangle$  représente la projection de  $X$  sur la droite vectorielle engendrée par  $\alpha$ . Ainsi, (a) dit qu'un vecteur aléatoire est gaussien si et seulement toutes ses projections sont des gaussiennes unidimensionnelles. En particulier, en projetant sur chaque droite engendrée par élément de la base canonique  $e_j$  de  $\mathbb{R}^d$ , on voit que les entrées d'un vecteur gaussien sont des gaussiennes. Cependant on peut construire des vecteurs dont chaque entrée est gaussienne qui n'est pas gaussien (et même de façon à ce que sa matrice de covariance soit l'identité), cf. Exercice 4.

– Fin du cours 3 –

On va maintenant prouver la Proposition 2.4. La clef est d'utiliser la forme explicite de la fonction caractéristique d'un vecteur gaussien que l'on calcule maintenant.

---

$\clubsuit$ . En effet, comme  $\Sigma_X$  est symétrique semi-définie positive on peut diagonaliser  $\Sigma_X$  dans une base orthonormée,  $\Sigma_X = {}^tO \text{diag}(\lambda_1, \dots, \lambda_d)O$  avec  $\lambda_j \geq 0$  et  $O {}^tO = I_d$ , et définir  $\Sigma_X^{1/2} := {}^tO \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})O$  qui satisfait bien  $\Sigma_X^{1/2} \cdot {}^t\Sigma_X^{1/2} = \Sigma_X$ . De plus, notez que  ${}^t\Sigma_X^{1/2} = \Sigma_X^{1/2}$

D'abord, rappelons que si  $X \sim \mathcal{N}(0, 1)$  alors  $\varphi_X(t) = e^{-t^2/2}$ , cf. Exercice 5. Maintenant, si  $X = {}^t(X_1, \dots, X_d) \sim \mathcal{N}(0, I_d)$ , on a par indépendance des entrées de  $X$  pour tout  $t = {}^t(t_1, \dots, t_d) \in \mathbb{R}^d$  que

$$\varphi_X(t) = \prod_{i=1}^n \varphi_{X_j}(t_j) = \prod_{j=1}^n e^{-t_j^2/2} = e^{-\langle t, t \rangle / 2}.$$

Du coup, si  $X = \Sigma^{1/2}Z + m \sim \mathcal{N}(m, \Sigma)$  où  $Z \sim \mathcal{N}(0, I_d)$ , on a

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = e^{i\langle t, m \rangle} \mathbb{E}[e^{i\langle t, \Sigma^{1/2}Z \rangle}] = e^{i\langle t, m \rangle} \varphi_Z({}^t\Sigma^{1/2}t),$$

et donc, comme

$$\varphi_Z({}^t\Sigma^{1/2}t) = e^{-\langle {}^t\Sigma^{1/2}t, {}^t\Sigma^{1/2}t \rangle / 2} = e^{-\langle t, \Sigma^{1/2}{}^t\Sigma^{1/2}t \rangle / 2} = e^{-\langle t, \Sigma t \rangle / 2},$$

on obtient

$$\boxed{X \sim \mathcal{N}(m, \Sigma) \Leftrightarrow \varphi_X(t) = e^{i\langle t, m \rangle - \langle t, \Sigma t \rangle / 2}, \quad t \in \mathbb{R}^d.} \quad (2.2)$$

En particulier, en dimension  $d = 1$ , on obtient

$$\boxed{X \sim \mathcal{N}(m, \sigma^2) \Leftrightarrow \varphi_X(t) = e^{itm - \sigma^2 t^2 / 2}, \quad t \in \mathbb{R}.} \quad (2.3)$$

*Démonstration de la Proposition 2.4(a).* Si  $\alpha \in \mathbb{R}^d$  et  $X \sim \mathcal{N}(m, \Sigma)$ , alors on a

$$\mathbb{E}[e^{it\langle \alpha, X \rangle}] = \mathbb{E}[e^{i\langle t\alpha, X \rangle}] = e^{i\langle t\alpha, m \rangle - \langle t\alpha, \Sigma t\alpha \rangle} = e^{it\langle \alpha, m \rangle - t^2\langle \alpha, \Sigma \alpha \rangle}, \quad t \in \mathbb{R},$$

et donc  $\langle \alpha, X \rangle \sim \mathcal{N}(\langle \alpha, m \rangle, \langle \alpha, \Sigma \alpha \rangle)$ . Réciproquement, supposons maintenant que  $\langle \alpha, X \rangle$  est gaussien pour tout  $\alpha \in \mathbb{R}^d$ , mettons  $\langle \alpha, X \rangle \sim \mathcal{N}(m_\alpha, \sigma_\alpha^2)$ . Comme l'espérance est linéaire, on voit que  $\alpha \mapsto m_\alpha$  est une forme linéaire, et donc il existe  $m \in \mathbb{R}^d$  tel que  $m_\alpha = \langle \alpha, m \rangle$ ; en fait  $m = {}^t(m_{e_1}, \dots, m_{e_d})$  convient, où  $e_1, \dots, e_d$  est la base canonique de  $\mathbb{R}^d$ . De façon similaire,  $\alpha \mapsto \sigma_\alpha^2$  est une forme quadratique et on note  $\Sigma$  sa matrice dans la base canonique, qui satisfait  $\sigma_\alpha^2 = {}^t\alpha \Sigma \alpha$  pour tout  $\alpha \in \mathbb{R}^d$ ; plus précisément, on prend  $\Sigma_{ij} := b(e_i, e_j)$  où  $b(\alpha, \beta) := \frac{1}{2}(\sigma_{\alpha+\beta}^2 - \sigma_\alpha^2 - \sigma_\beta^2)$ . On a donc

$$\mathbb{E}[e^{i\langle \alpha, X \rangle}] = e^{im_\alpha - \sigma_\alpha^2 / 2} = e^{i\langle \alpha, m \rangle - \langle \alpha, \Sigma \alpha \rangle / 2}$$

pour tout  $\alpha \in \mathbb{R}^d$ , et donc  $X$  est bien un vecteur gaussien, de loi  $\mathcal{N}(m, \Sigma)$ .  $\square$

*Démonstration de la Proposition 2.4(b).* On a pour toute matrice  $A$  de taille  $d \times d$ ,

$$\varphi_{AX}(t) = \mathbb{E}[e^{i\langle t, AX \rangle}] = \varphi_X({}^tAt) = e^{i\langle {}^tAt, m \rangle + \langle {}^tAt, \Sigma {}^tAt \rangle} = e^{i\langle t, Am \rangle + \langle t, A\Sigma {}^tAt \rangle},$$

ce qui prouve (b).  $\square$

On prouvera (c) dans l'Exercice 9.

*Démonstration de la Proposition 2.4(d).* Si on pose  $\alpha := ue_i + ve_j$ , alors on a

$$\varphi_{(X_i, X_j)}(u, v) = \mathbb{E}[e^{i(uX_i + vX_j)}] = \mathbb{E}[e^{i\langle \alpha, X \rangle}] = e^{i\langle \alpha, m \rangle} e^{-\langle \alpha, \Sigma \alpha \rangle / 2}$$

et on calcule alors que, pour tout  $u, v \in \mathbb{R}$ ,

$$\varphi_{(X_i, X_j)}(u, v) = e^{i u m_i + i v m_j} e^{-(u^2 \Sigma_{ii} + 2 u v \Sigma_{ij} + v^2 \Sigma_{jj})/2} = \varphi_{X_i}(u) \varphi_{X_j}(v) e^{-u v \Sigma_{ij}}.$$

ou l'on a utilisé que, comme on l'a vu dans la preuve du point (a) de la proposition,  $X_i = \langle X, e_i \rangle \sim \mathcal{N}(\langle m, e_i \rangle, \langle e_i, \Sigma e_i \rangle) = \mathcal{N}(m_i, \Sigma_{ii})$  ainsi que (2.3). On voit alors que  $X_i$  et  $X_j$  sont indépendants si et seulement si  $\Sigma_{ij} = \text{Cov}(X_i, X_j) = 0$ .  $\square$

Une remarque concernant le point (c) : Si  $\Sigma$  n'est pas inversible,  $X$  n'a pas de densité par rapport à  $\text{Leb}^{\otimes d}$  mais est portée par le sous-espace affine  $\Sigma^{1/2}(\mathbb{R}^d) + \mu$  de dimension le rang de  $\Sigma$ .

On utilise souvent en statistiques l'observation suivante, attribuée à Cochran, qui s'intéresse au cas particulier où  $\Sigma = \Pi_V$  est une matrice de projection sur un sous-espace vectoriel  $V$  de  $\mathbb{R}^d$  et  $m = 0$ . Rappelez-vous qu'une matrice de projection orthogonale satisfait  $\Pi_V = {}^t \Pi_V = \Pi_V^2$  et  $\Pi_V$  est donc sa propre racine carrée.

**Théorème 2.5** (Cochran). *Soit  $X \sim \mathcal{N}(0, I_d)$ . Si  $V \subset \mathbb{R}^d$  est un sous-espace vectoriel de dimension  $k$  et  $\Pi_V$  est la matrice de la projection orthogonale sur  $V$ , alors  $\|\Pi_V X\|^2 \sim \chi_k^2$ . De plus,  $\Pi_V X$  et  $\Pi_{V^\perp} X$  sont indépendants.*

Notez que si  $V = \text{Vect}(e_1, \dots, e_k)$  avec  $e_1, \dots, e_n$  la base canonique de  $\mathbb{R}^d$  ce résultat est évident (c'est clair?). Pour traiter le cas général, on utilise que la loi d'un vecteur gaussien standard  $\mathcal{N}(0, I_d)$  est invariante sous les isométries (on dit qu'il est isotrope) pour se ramener à ce cas particulier.

*Démonstration.* Soit  $e_1, \dots, e_d$  la base canonique de  $\mathbb{R}^d$  et

$$\Pi_k := \text{diag}(\underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k}).$$

Si  $V$  est un sous-espace quelconque de dimension  $k$ , il admet une base orthonormée  $v_1, \dots, v_k$ , que l'on peut compléter en une base orthonormée  $v_1, \dots, v_d$  de  $\mathbb{R}^d$ . Soit  $O$  la matrice de passage de la base canonique  $(e_j)$  à la base  $(v_j)$ ; c'est une matrice orthogonale car ces deux bases sont orthonormées, et donc  $O {}^t O = {}^t O O = I_d$ . En fait, on peut même vérifier que  $O_{ij} = \langle e_i, v_j \rangle$ . On a alors  $\Pi_V = O \Pi_k {}^t O$ . La clef de la preuve est que  ${}^t O X \sim \mathcal{N}(0, {}^t O O) = \mathcal{N}(0, I_d)$ , et en particulier que  ${}^t O X = (Y_1, \dots, Y_j)$  avec des variables  $Y_j \sim \mathcal{N}(0, 1)$  indépendantes. On voit alors que  $\Pi_k {}^t O X = (Y_1, \dots, Y_k, 0, \dots, 0)$  et  $(I_d - \Pi_k) {}^t O X = (0, \dots, 0, Y_{k+1}, \dots, Y_n)$  sont indépendants. Comme  $\Pi_V X = O \Pi_k {}^t O X$  et  $\Pi_{V^\perp} X = O(I_n - \Pi_V) {}^t O X$ , on obtient que  $\Pi_V X$  et  $\Pi_{V^\perp} X$  sont indépendants. De plus, en utilisant encore que  $O$  est une isométrie, on a  $\|\Pi_V X\|^2 = \|O \Pi_k {}^t O X\|^2 = \|\Pi_k {}^t O X\|^2 = Y_1^2 + \dots + Y_k^2 \sim \chi_k^2$ .  $\square$

## 2.4 Quelques inégalités importantes

**Théorème 2.6 (Inégalité de Markov).** *Si  $X$  est une variable aléatoire positive,*

$$\mathbb{P}(X > \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}.$$

*Démonstration.* Comme  $X \geq \varepsilon \mathbf{1}_{X > \varepsilon}$  (dessin) pour tout  $\varepsilon > 0$ , on a

$$\mathbb{E}[X] = \int X \, d\mathbb{P} \geq \varepsilon \int \mathbf{1}_{X > \varepsilon} \, d\mathbb{P} = \varepsilon \mathbb{P}(X > \varepsilon).$$

□

En écrivant  $\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) = \mathbb{P}(|X - \mathbb{E}[X]|^2 > \varepsilon^2)$  et en appliquant l'inégalité de Markov à la variable  $(X - \mathbb{E}[X])^2$ , on obtient :

**Corollaire 2.7 (Inégalité de Tchebychev).** *Si  $X$  est une variable réelle  $L^1$  alors*

$$\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}.$$

L'inégalité de Tchebychev montre que la variance contrôle l'étalement de la variable autour de son espérance. Notez que si  $X$  est un vecteur aléatoire  $L^1$ , au sens où chacune de ses entrées  $X_i$  est  $L^1$ , et  $|\cdot|$  représente la norme euclidienne, alors le même argument montre que :

$$\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \frac{\mathbb{E}(|X - \mathbb{E}[X]|^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{i=1}^d \text{Var}[X_i].$$

**Théorème 2.8 (Inégalité de Jensen).** *Si  $X : \Omega \rightarrow \mathbb{R}^d$  est une variable  $L^1$  et  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  est une fonction convexe telle que  $\varphi(X) \in L^1$ , alors*

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

Rappelons qu'une *fonction convexe*  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfait par définition, pour tout  $\alpha \in [0, 1]$  et tout  $x, y \in \mathbb{R}^d$ , l'inégalité  $\varphi(\alpha x + (1 - \alpha)y) \leq \alpha\varphi(x) + (1 - \alpha)\varphi(y)$ . On peut aussi montrer qu'une fonction convexe  $\mathbb{R}^d \rightarrow \mathbb{R}$  est forcément continue ♣. Si  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  est différentiable, alors  $\varphi$  est convexe si et seulement si  $\langle \nabla\varphi(x) - \nabla\varphi(y), x - y \rangle \geq 0$  pour tout  $x, y \in \mathbb{R}^d$ . En particulier, si la matrice Hessienne de  $\varphi$  est bien définie en tout point et est semi-définie positive, alors  $\varphi$  est convexe. En dimension  $d = 1$ , cela revient à dire que  $\varphi'$  est croissante ou que  $\varphi'' \geq 0$ .

En appliquant l'inégalité de Jensen à  $\varphi(x) = |x|$ , on retrouve que  $|\mathbb{E}[X]| \leq \mathbb{E}|X|$ . En prenant  $\varphi(x) = x^2$  on voit que  $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ , qui est équivalent à  $\text{Var}(X) \geq 0$ .

*Démonstration.* On montre par une simple récurrence que la convexité de  $\varphi$  est équivalent à : pour tout  $m \geq 1$ , tout  $\alpha_1, \dots, \alpha_m \in [0, 1]$  tels que  $\sum_{i=1}^m \alpha_i = 1$  et tout  $x_1, \dots, x_m \in \mathbb{R}^d$ ,

$$\varphi\left(\sum_{i=1}^m \alpha_i x_i\right) \leq \sum_{i=1}^m \alpha_i \varphi(x_i). \quad (2.4)$$

Maintenant, si la loi de  $X$  est discrète de la forme

$$\mu_X = \sum_{i=1}^m \alpha_i \delta_{x_i}, \quad (2.5)$$

---

♣. Remarquez que la fonction définie sur  $[0, 1]$  par  $f = \mathbf{1}_{\{0,1\}}$  est convexe mais non continue et que la fonction définie sur  $\mathbb{R}$  par  $f(x) = x$  sur  $[0, +\infty[$  et  $f(x) = +\infty$  sur  $]-\infty, 0[$  est convexe mais pas continue. Pour que convexité implique continuité, il faut que  $f$  soit définie sur un *ouvert convexe* et finie en tous points.

alors l'inégalité de Jensen est exactement (2.4). On va obtenir le cas général par approximation, un type de raisonnement très courant qu'il nous paraît utile de détailler sur cet exemple. Pour cela, on admettra ♣ que pour toute mesure de probabilité  $\mu$  à *support compact*, c'est-à-dire si il existe  $R > 1$  tel que  $\mu(B_R) = 1$  avec  $B_R := \{x \in \mathbb{R}^d : |x| \leq R\}$ , il existe pour tout  $n$  une mesure de probabilité  $\mu_n$  de la forme (2.5) telle que, pour toute fonction  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  continue

$$\lim_{n \rightarrow \infty} \int f \, d\mu_n = \int f \, d\mu.$$

Du coup, si la loi  $\mu_X$  de  $X$  a un support compact, on l'approxime avec une telle suite de mesures discrètes et, en utilisant que  $\varphi$  est continue, on obtient

$$\begin{aligned} \varphi(\mathbb{E}[X]) &= \varphi\left(\int x \, d\mu_X\right) \\ &= \varphi\left(\lim_{n \rightarrow \infty} \int x \, d\mu_n\right) \\ &= \lim_{n \rightarrow \infty} \varphi\left(\int x \, d\mu_n\right) \\ &\leq \lim_{n \rightarrow \infty} \left(\int \varphi(x) \, d\mu_n\right) = \int \varphi \, d\mu_X = \mathbb{E}[\varphi(X)]. \end{aligned}$$

Il reste à nous débarrasser de l'hypothèse supplémentaire que la loi de  $X$  est à support compact. On va encore procéder par approximation : si  $\mu_X$  n'est pas à support compact, posons  $\mu_X^R := \frac{1}{\mu_X(B_R)} \mathbf{1}_{B_R} \mu_X$  pour tout  $R > 1$ , qui est une mesure de probabilité à support compact. On a alors par convergence dominée et continuité de  $\varphi$ ,

$$\begin{aligned} \varphi(\mathbb{E}[X]) &= \varphi\left(\lim_{R \rightarrow \infty} \int x \, d\mu_X^R(x)\right) \\ &= \lim_{R \rightarrow \infty} \varphi\left(\int x \, d\mu_X^R(x)\right) \\ &\leq \lim_{R \rightarrow \infty} \int \varphi(x) \, d\mu_X^R(x) = \mathbb{E}[\varphi(X)]. \end{aligned}$$

□

– Fin du cours 4 –

**Théorème 2.9 (Inégalité de Hölder).** *Si  $X, Y : \Omega \rightarrow \mathbb{R}$  sont des variables aléatoires réelles telle que  $X \in L^p$  et  $Y \in L^q$  avec  $p > 1$  et  $q := p/(p-1) > 1$  alors  $XY \in L^1$  et*

$$\mathbb{E}|XY| \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}.$$

---

♣. Ce résultat est une conséquence de théorèmes classiques d'analyse fonctionnelle et peut être prouvé de la façon suivante : L'espace  $\mathcal{P}(K)$  des mesures de probabilité supporté dans un compact fixé  $K \subset \mathbb{R}^d$  est un sous-ensemble compact et convexe de l'espace des mesures de  $\mathbb{R}^d$ . Ce dernier est un espace localement convexe une fois équipé de la famille de semi-normes  $\mu \mapsto \int f \, d\mu$  indexée par les fonctions continues  $f : K \rightarrow \mathbb{R}$ . Le théorème de Krein-Milman implique alors que  $\mathcal{P}(K)$  est la fermeture de l'enveloppe convexe de ses points extrémaux, qui sont exactement les masses de Dirac  $\delta_x$ .

Notez que  $q$  est défini tel que  $\frac{1}{p} + \frac{1}{q} = 1$ . Aussi, l'inégalité s'entend au cas  $p = 1$  et donc " $q = \infty$ " car on a

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|] \sup |Y|$$

dès que  $Y$  est une variable bornée.

*Démonstration.* Comme  $\varphi(x) := |x|^p$  est convexe dès que  $p > 1$ , l'inégalité de Jensen nous donne pour toute variable  $Z$  réelle  $|\mathbb{E}[Z]|^p \leq \mathbb{E}[|Z|^p]$  et donc  $|\mathbb{E}[Z]| \leq \mathbb{E}[|Z|^p]^{1/p}$ . On va symétriser cette relation : pour  $\alpha > 1$  on écrit

$$\begin{aligned} \mathbb{E}[|XY|] &= \int |X||Y| \, d\mathbb{P} \\ &= \int \frac{|X|}{|Y|^{\alpha-1}} |Y|^\alpha \, d\mathbb{P} \\ &= \mathbb{E}[|Y|^\alpha] \int Z \, d\tilde{\mathbb{P}} \end{aligned}$$

où l'on a posé

$$Z := |X||Y|^{1-\alpha} \quad \text{et} \quad \tilde{\mathbb{P}} := \frac{1}{\mathbb{E}[|Y|^\alpha]} |Y|^\alpha \, d\mathbb{P}.$$

Remarquez que  $\tilde{\mathbb{P}}$  est une mesure de probabilité et que  $Z < \infty$   $\tilde{\mathbb{P}}$ -p.s. L'inégalité de Jensen nous donne alors

$$\int Z \, d\tilde{\mathbb{P}} \leq \left( \int |Z|^p \, d\tilde{\mathbb{P}} \right)^{1/p} = \frac{1}{\mathbb{E}[|Y|^\alpha]^{1/p}} \left( \int |X|^p |Y|^{p(1-\alpha)+\alpha} \, d\mathbb{P} \right)^{1/p}$$

et finalement

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|Y|^\alpha]^{1-1/p} \left( \int |X|^p |Y|^{p(1-\alpha)+\alpha} \, d\mathbb{P} \right)^{1/p}.$$

Maintenant,  $p(1-\alpha) + \alpha = 0$  si et seulement si  $\alpha = p/(1-p)$ , et dans ce cas l'inégalité de Hölder est prouvée.  $\square$

## 2.5 Convergence de variables aléatoires

On considère ici une suite de variables aléatoires  $X_n : \Omega \rightarrow E$ , c'est-à-dire toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  et à valeurs dans le même espace métrique complet  $E$ , équipé d'une distance  $d$  et de la tribu des Boreliens  $\mathcal{B}(E)$  associée aux boules ouvertes pour  $d$ ; dans la suite,  $E$  sera souvent  $\mathbb{R}$  ou  $\mathbb{R}^d$  et on prendra pour distance la distance (et donc la topologie) euclidienne  $d(x, y) := \|x - y\|$ . On considère aussi une variable supplémentaire  $X$  définie et à valeur sur ces mêmes espaces.

— On dit que  $X_n$  *converge presque sûrement* vers  $X$  quand  $n \rightarrow \infty$  si

$$\lim_{n \rightarrow \infty} d(X_n(\omega), X(\omega)) = 0$$

pour  $\mathbb{P}$ -presque tout  $\omega \in \Omega$ . On écrit

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

C'est assez exigeant comme convergence et elle dépend de comment on a construit  $\Omega$ . Une notion de convergence un peu plus faible est la suivante :

— On dit que  $X_n$  *converge en probabilité* vers  $X$  quand  $n \rightarrow \infty$  si

$$\lim_{n \rightarrow \infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0$$

pour tout  $\varepsilon > 0$ . On écrit

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

En effet, la convergence presque sûre implique la convergence en probabilités. La convergence en probabilité n'implique pas forcément la convergence p.s mais si  $X_n \rightarrow X$  en probabilités alors il existe une sous-suite  $X_{\varphi(n)}$  qui converge p.s vers  $X$  (c'est une conséquence de l'exercice 12).

— On dit que  $X_n$  *converge en loi* ou *en distribution* vers  $X$  quand  $n \rightarrow \infty$  si

$$\mathbb{E}[f(X_n)] = \int f \, d\mu_{X_n} \xrightarrow[n \rightarrow \infty]{} \int f \, d\mu_X = \mathbb{E}[f(X)] \quad (2.6)$$

pour toute fonction  $f : E \rightarrow \mathbb{R}$  continue bornée. On écrit

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

Si les convergences p.s. et en probabilité dépendent fortement de l'espace  $(\Omega, \mathcal{F}, \mathbb{P})$  sur lequel sont construites les variables aléatoires, ce n'est pas le cas de la convergence en loi. La convergence en probabilité implique la convergence en loi mais la réciproque est fautive. Par contre, si  $X_n \rightarrow X$  en loi, alors on peut construire un espace probablisé  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$  et des variables aléatoires  $\hat{X}_n : \hat{\Omega} \rightarrow E$  et  $\hat{X} : \hat{\Omega} \rightarrow E$  qui ont respectivement la même loi que  $X_n$  et  $X$ , et telles que  $\hat{X}_n \rightarrow \hat{X}$   $\hat{\mathbb{P}}$ -p.s.; ce résultat est connu comme le théorème de représentation de Skorokhod.

En utilisant que  $X$  est une variable aléatoire, on peut montrer qu'il est suffisant de montrer la convergence (2.7) pour toute fonction  $f$  continue à support compact. Malgré cela, cette caractérisation n'est pas toujours pratique et peut utiliser des fonctions génératrices pour simplifier la tâche. En effet, si  $X_n$  et  $X$  sont à valeurs dans  $\mathbb{R}$ , alors

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \quad \Leftrightarrow \quad F_{X_n}(t) \rightarrow F_X(t) \quad \text{pour tout } t \text{ là où } F \text{ est continue}$$

et plus généralement, si  $X_n$  et  $X$  sont à valeurs dans  $\mathbb{R}^d$ , alors

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \quad \Leftrightarrow \quad \varphi_{X_n}(t) \rightarrow \varphi_X(t) \quad \text{pour tout } t \in \mathbb{R}^d.$$

— On dit que  $X_n$  *converge* vers  $X$  *dans*  $L^p$  quand  $n \rightarrow \infty$  si

$$\lim_{n \rightarrow \infty} \mathbb{E}[d(X_n, X)^p] = 0, \quad (2.7)$$

ce qu'on écrit

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

Si  $(E, d) = (\mathbb{R}, |\cdot|)$  et  $p' \leq p$ , alors l'inégalité de Hölder montre que  $\mathbb{E}[|X|^{p'}] \leq \mathbb{E}[|X|^p]^{p'/p}$ , ce donne l'inclusion  $L^{p'} \subset L^p$ , et de plus

$$X_n \xrightarrow[n \rightarrow \infty]{L^p} X \quad \Rightarrow \quad X_n \xrightarrow[n \rightarrow \infty]{L^{p'}} X.$$



## 2.6 Loi des grands nombres

**Théorème 2.10** (Loi des grands nombres). *Soit  $X$  vecteur aléatoire de  $\mathbb{R}^d$  qui est  $L^1$ . Si  $(X_i)_{i \geq 1}$  est une suite de copies i.i.d de  $X$ , alors*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[X] \quad p.s.$$

*Preuve (sous l'hypothèse supplémentaire  $L^4$ ). Il suffit de traiter le cas  $d = 1$ , car le cas  $d > 1$  s'obtient en traitant composante par composante. On fait l'hypothèse supplémentaire que  $X \in L^4$  pour simplifier la preuve; le cas  $L^1$  demande beaucoup plus de travail. Quitte à travailler avec la variable  $\tilde{X}_i := X_i - \mathbb{E}[X]$  à la place de  $X_i$ , on peut supposer que  $\mathbb{E}[X] = 0$ . On pose  $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ . On va étudier le comportement de  $\mathbb{E}[\bar{X}_n^4]$  en  $n$ . Pour se faire, en développant les produits de sommes,*

$$\left( \sum_{i=1}^n X_i \right)^4 = \sum_{i=1}^n X_i^4 + \sum_{i \neq j} X_i^3 X_j + \sum_{i \neq j} X_i^2 X_j^2 + \sum_{i \neq j \neq k} X_i^2 X_j X_k + \sum_{i \neq j \neq k \neq \ell} X_i X_j X_k X_\ell$$

et en utilisant que les  $X_i$  sont des copies *indépendantes* de  $X$  et que  $\mathbb{E}[X_i] = 0$ , on obtient

$$\mathbb{E} \left[ \left( \sum_{i=1}^n X_i \right)^4 \right] = n \mathbb{E}[X^4] + n(n-1) \mathbb{E}[X^2]^2 \leq \mathbb{E}[X^4] n^2.$$

Maintenant, en utilisant le théorème de Fubini-Tonelli, on a :

$$\mathbb{E} \left[ \sum_{n=1}^{\infty} \bar{X}_n^4 \right] = \sum_{n=1}^{\infty} \mathbb{E} \left[ \bar{X}_n^4 \right] \leq \mathbb{E}[X^4] \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Par conséquent,

$$\sum_{n=1}^{\infty} \bar{X}_n^4 < \infty \quad p.s.$$

car autrement cette somme aurait une espérance infinie. Et comme une série convergente a nécessairement un terme général qui tend vers zéro, on a  $\bar{X}_n^4 \rightarrow 0$  p.s. quand  $n \rightarrow \infty$ , et donc finalement  $\bar{X}_n \rightarrow 0 = \mathbb{E}[X]$  p.s.  $\square$

**Remarque 2.11.** *On pourrait tenter la même approche avec le moment d'ordre 2 en supposant seulement que  $\mathbb{E}[X^2] < \infty$ , mais un calcul montre que  $\mathbb{E}[(\frac{1}{n} \sum_i X_i)^2]$  est d'ordre  $1/n$ , dont la série n'est pas convergente. Les moment d'ordre impair ça ne marche pas non plus car  $\mathbb{E}[(\frac{1}{n} \sum_i X_i)^3]$  n'est pas forcément positif.*

L'argument final est souvent isolé dans un résultat très utilisé en probabilité, attribué à Borel-Cantelli. On définit d'abord la *limite supérieure d'une suite d'évènements* : Étant donné une suite de sous-ensembles  $(A_n)_{n \in \mathbb{N}}$  d'un ensemble  $\Omega$ , on pose :

$$\limsup A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k = \left\{ \omega \in \Omega : \forall n \in \mathbb{N}, \exists k \geq n \text{ tel que } \omega \in A_k \right\}.$$

De façon équivalente,  $\omega \in \limsup A_n \Leftrightarrow$  l'ensemble  $\{n \in \mathbb{N} : \omega \in A_n\}$  est infini.

**Théorème 2.12** (Théorème de Borel-Cantelli).

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}\left(\limsup A_n\right) = 0.$$

*Démonstration.* En utilisant le théorème de Fubini-Tonelli on a, par hypothèse,

$$\mathbb{E}\left[\sum_{n=0}^{\infty} \mathbf{1}_{A_n}\right] = \sum_{n=0}^{\infty} \mathbb{E}[\mathbf{1}_{A_n}] = \sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty.$$

Par conséquent,

$$\mathbb{P}\left(\limsup A_n\right) = \mathbb{P}\left(\sum_{n=0}^{\infty} \mathbf{1}_{A_n} = \infty\right) = 0.$$

□

Le théorème de Borel-Cantelli est l'un des rares outils probabilistes qui donne une condition suffisante pour obtenir une convergence presque sûre.

– Fin du cours 5 –

## 2.7 Théorème central limite

La loi des grands nombres stipule que la moyenne empirique  $\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$  est p.s. une bonne approximation de la vraie moyenne  $\mathbb{E}[X]$  pourvu que  $n \rightarrow \infty$ . Mais à partir de quand “ $n$  est assez grand” ? Une réponse partielle est apporté par le théorème central limite, qui dit en gros que les fluctuations de  $\overline{X}_n$  autour de  $\mathbb{E}[X]$  sont d’amplitude  $1/\sqrt{n}$  et de nature gaussienne.

**Théorème 2.13** (Théorème central limit). *Si  $(X_i)_{i \geq 1}$  est une suite de copies indépendantes d’un vecteur  $X \in L^2$  de matrice de covariance  $\Sigma$ , alors*

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X] \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \Sigma).$$

*Démonstration.* Quitte à travailler avec  $\tilde{X}_i := (X_i - \mathbb{E}[X])$  à la place de  $X_i$ , on peut supposer que  $\mathbb{E}[X] = 0$ . Comme  $X \in L^2$ , on voit que  $\varphi_X$  est deux fois dérivable par théorème de dérivation sous le signe intégral et

$$\partial_i \varphi_X(t) = i \mathbb{E}[X_i e^{i\langle t, X \rangle}], \quad \partial_i \partial_j \varphi_X(t) = -\mathbb{E}[X_i X_j e^{i\langle t, X \rangle}].$$

Plus précisément, comme  $\varphi_X(t) = \int f(\cdot, t) d\mu_X$  avec  $f(x, t) := e^{i\langle x, t \rangle}$  pour tout  $t \in \mathbb{R}^d$  et on a, puisque  $X \in L^2$  par hypothèse,

- $f(x, t)$  est  $C^\infty$  en  $x, t \in \mathbb{R}^d$ .
- pour tout  $t \in \mathbb{R}^d$ ,  $|\partial_j f(x, t)| = |X_j e^{i\langle x, t \rangle}| = |X_j| \in L^1$
- pour tout  $t \in \mathbb{R}^d$ ,  $|\partial_j \partial_k f(x, t)| = |X_j X_k e^{i\langle x, t \rangle}| = |X_j X_k| \in L^1$  par Cauchy-Schwarz,

et on peut conclure comme dans l'exercice 12 du chapitre précédent. En particulier, on a :

$$\nabla\varphi_X(0) = i\mathbb{E}[X], \quad \text{Hess}[\varphi_X(0)] = -\Sigma.$$

Une propriété importante de la fonction caractéristique est que, si  $X, Y$  sont deux v.a. indépendantes et  $\alpha \in \mathbb{R}$ ,

$$\varphi_{X+Y}(t) = \varphi_X(t) \times \varphi_Y(t), \quad \varphi_{\alpha X}(t) = \varphi_X(\alpha t).$$

Ainsi, on a pour tout  $t \in \mathbb{R}$  fixé, quand  $n \rightarrow \infty$ ,

$$\begin{aligned} \varphi_{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i}(t) &= \varphi_{\sum_{i=1}^n X_i}\left(\frac{t}{\sqrt{n}}\right) \\ &= \varphi_X\left(\frac{t}{\sqrt{n}}\right)^n \\ &= \left(\varphi_X(0) + \frac{1}{\sqrt{n}} \langle t, \nabla\varphi_X(0) \rangle + \frac{1}{2n} \langle t, \text{Hess}[\varphi_X(0)]t \rangle + o\left(\frac{1}{n}\right)\right)^n \\ &= \left(1 - \frac{1}{2n} \langle t, \Sigma t \rangle + o\left(\frac{1}{n}\right)\right)^n \\ &= e^{n \log\left(1 - \frac{1}{2n} \langle t, \Sigma t \rangle + o\left(\frac{1}{n}\right)\right)} \\ &= e^{-\langle t, \Sigma t \rangle/2 + o(1)}. \end{aligned}$$

Comme  $e^{-\langle t, \Sigma t \rangle/2}$  est la fonction caractéristique d'un vecteur  $\mathcal{N}(0, \Sigma)$ , le résultat est prouvé.  $\square$

Ce résultat reste asymptotique : l'erreur  $\sqrt{n}(\bar{X}_n - \mathbb{E}[X])$  converge vers une gaussienne, mais à quelle vitesse ? Une réponse (partielle) est apportée par le Théorème de Berry-Essen.

## 2.8 EXERCICES – Probabilités

**Exercice 1.** Montrer que si  $\text{Var}(X) = 0$  alors  $X$  est constante p.s.

**Exercice 2.** Soit  $X \sim \mathcal{B}(p)$  une variable de Bernoulli de paramètre  $p \in ]0, 1[$ , définie par  $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$ . Quel est sa loi  $\mu_X$ ? Montrer qu'elle a une densité par rapport à  $\mu = \delta_0 + \delta_1$ . Généraliser au cas d'une variable aléatoire quelconque à valeurs dans un espace discret.

**Exercice 3.** Montrer que si  $X \sim \mathcal{E}(\theta)$  alors pour tout  $s, t > 0$ ,

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t).$$

**Exercice 4.** Soit  $X \sim \mathcal{N}(0, 1)$  et  $Z$  définie par  $\mathbb{P}(Z = -1) = \mathbb{P}(Z = 1) = 1/2$  qu'on suppose indépendante de  $X$ . On considère la variable  $Y := ZX$ .

- Montrer que  $Y \sim \mathcal{N}(0, 1)$ .
- Calculer  $\text{Cov}(X, Y)$ . Est-ce que  $X$  et  $Y$  sont indépendantes?
- Est-ce que le vecteur aléatoire  ${}^t(X, Y)$  est un vecteur Gaussien?

**Exercice 5.** Calculer la fonction caractéristique  $\varphi(t) := \mathbb{E}[e^{itX}]$  de  $X \sim \mathcal{N}(0, 1)$ .

*Aide : On pourra montrer que  $\varphi'(t) = -t\varphi(t)$ .*

**Exercice 6.** On admet <sup>♣</sup> que la fonction caractéristique d'un loi Gamma  $\Gamma(k, \theta)$  est

$$\varphi(t) = \left( \frac{\theta}{\theta - it} \right)^k.$$

- Montrer que

$$X \sim \Gamma(k, \theta), \quad Y \sim \Gamma(\ell, \theta), \quad X, Y \text{ indépendantes} \quad \Rightarrow \quad X + Y \sim \Gamma(k + \ell, \theta).$$

- Soit  $X_1, \dots, X_d$  des variables i.i.d  $\mathcal{N}(0, 1)$ . Calculer la densité de la variable  $X_1^2$  puis montrer que  $X_1^2 + \dots + X_d^2 \sim \Gamma(d/2, 1/2)$ .

**Exercice 7.** On suppose que  $X$  est une variable aléatoire réelle.

- Montrer que

$$\lim_{R \rightarrow \infty} \mathbb{P}(|X| > R) = 0.$$

- Si on suppose de plus que  $X \in L^p$ , montrer que

$$\mathbb{P}(|X| > R) \leq \frac{\mathbb{E}|X|^p}{R^p}$$

et que, si il existe  $\alpha > 0$  tel que  $C := \mathbb{E}[e^{\alpha|X|}] < \infty$ , alors

$$\mathbb{P}(|X| > R) \leq \frac{C}{e^{\alpha R}}.$$

Discuter cette dernière condition pour  $X \sim \mathcal{N}(0, 1)$  et  $X \sim \mathcal{E}(\theta)$ .

---

<sup>♣</sup>. mais on pourrait le montrer de la même façon que dans l'exercice 5.

- (c) Que pensez vous de la décroissance de  $\mathbb{P}(|X| > R)$  quand  $R \rightarrow \infty$  pour une variable  $X$  de Cauchy, c'est-à-dire de densité

$$f(x) = \frac{1}{\pi(1+x^2)}$$

par rapport à Lebesgue ?

**Exercice 8.** Soit  $X$  une variable aléatoire réelle. Montrer que

$$\mathbb{E}|X| = \int_0^\infty \mathbb{P}(|X| > t) dt.$$

*Aide : On pourra utiliser le théorème de Fubini-Tonelli.*

Plus généralement, montrer que pour  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$  croissante et dérivable p.p,

$$\mathbb{E}[\varphi(|X|)] = \int_0^\infty \varphi'(t) \mathbb{P}(|X| > t) dt + \varphi(0)$$

et donner des formules explicites pour  $\varphi(x) = x^p$  et  $\varphi(x) = e^{\alpha x}$ .

**Exercice 9 (Densité d'un vecteur gaussien).** Soit  $X_1, \dots, X_d$  i.i.d  $\mathcal{N}(0, 1)$ .

- Donner la densité du vecteur aléatoire  $X = {}^t(X_1, \dots, X_d)$  par rapport à la mesure de Lebesgue de  $\mathbb{R}^d$ .
- Si  $m \in \mathbb{R}^d$  et  $A \in \text{GL}_d(\mathbb{R})$ , donner la densité du vecteur aléatoire  $AX + m$ .
- Que peut-on dire si  $A$  n'est pas inversible ?

**Exercice 10 (Dessine-moi un vecteur gaussien).** A l'aide de R, simuler et représenter graphiquement 100 réalisations indépendantes d'un vecteur gaussien de  $\mathbb{R}^2$  de loi  $\mathcal{N}(0, \Sigma)$  où :

- $\Sigma = I_2$
- $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}$
- $\Sigma = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$
- $\Sigma = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$

On donnera dans chaque cas la densité par rapport à Lebesgue si elle existe.

**Exercice 11 (Convergences de sommes).** Soit  $(X_n)$  et  $(Y_n)$  deux suites de variables aléatoires et  $X, Y$  deux variables aléatoires à valeurs dans  $(\mathbb{R}^d, \|\cdot\|)$ . Montrer que :

- $X_n \rightarrow X$  p.s. et  $Y_n \rightarrow Y$  p.s. implique que  $X_n + Y_n \rightarrow X + Y$  p.s.
- $X_n \rightarrow X$  en probabilité et  $Y_n \rightarrow Y$  en probabilité implique que  $X_n + Y_n \rightarrow X + Y$  en probabilité.
- $X_n \rightarrow X$  en loi et  $Y_n \rightarrow Y$  en loi n'implique *pas* que  $X_n + Y_n \rightarrow X + Y$  en loi.

**Exercice 12 (Un critère important de convergence p.s).** Soit  $(X_n)$  une suite de variables aléatoires et  $X$  une variable aléatoire à valeurs dans le même espace métrique  $(E, d)$ . Montrer que si pour tout  $\varepsilon > 0$  on a

$$\sum_{n=1}^{\infty} \mathbb{P}(d(X_n, X) > \varepsilon) < \infty$$

alors  $X_n \rightarrow X$  p.s.

**Exercice 13 (Borel-Cantelli, le retour).**

(a) Soit une suite  $(A_n)_{n \geq 1}$  de  $\mathcal{F}$ . Montrer que

$$\mathbb{P}(\limsup_n A_n) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right).$$

(b) En déduire que si on a la condition d'indépendance suivante :

$$\mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) = \prod_{k \geq n} \mathbb{P}(A_k^c),$$

alors

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \quad \Rightarrow \quad \mathbb{P}(\limsup_n A_n) = 1.$$

(c) Soit  $(X_i)_{i \geq 1}$  une suite i.i.d de loi exponentielle  $\mathcal{E}(\theta)$  et  $\alpha \geq 0$ . Montrer que

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \geq \alpha \log n \text{ pour une infinité de } n\right) = \begin{cases} 0 & \text{si } \alpha > 2/\theta \\ 1 & \text{si } \alpha \leq 2/\theta \end{cases}.$$

(d) Soit  $X_n$  une suite de variables indépendantes où  $X_n$  est une variable de Bernoulli de paramètre  $1/n$ . Montrer que  $X_n \rightarrow 0$  en probabilité mais pas presque sûrement. Donner une sous-suite  $(X_{\varphi(n)})$  qui converge p.s. vers zéro.

**Exercice 14. (Lemme de Scheffé).** Soit  $X_n$  une suite de variables à valeurs dans  $(E, \mathcal{F})$  de densité  $f_n$  par rapport à une mesure de référence  $\nu$  sur  $E$ . On suppose que  $f_n \rightarrow f$   $\nu$ -p.p et que  $f$  est une densité, c'est-à-dire que  $\int f d\nu = 1$  et  $f \geq 0$ .

(a) Montrer que  $\int |f_n - f| d\nu \rightarrow 0$  quand  $n \rightarrow \infty$ .

(b) En déduire que  $X_n \rightarrow X$  en loi, où  $X$  est une variable de loi  $f d\nu$ , et qu'on a même la convergence uniforme :

$$\sup_{A \in \mathcal{F}} |\mathbb{P}(X_n \in A) - \mathbb{P}(X \in A)| \xrightarrow{n \rightarrow \infty} 0.$$

(c) Montrer que, si  $E$  est discret, alors  $X_n \rightarrow X$  en loi si et seulement si,

$$\forall e \in E, \quad \lim_{n \rightarrow \infty} \mathbb{P}(X_n = e) = \mathbb{P}(X = e).$$

(d) Montrer que, si  $X_n$  suit une loi de Student<sup>1</sup> de paramètre  $n$ , alors  $X_n$  convergence en loi vers une variable  $\mathcal{N}(0, 1)$  quand  $n \rightarrow \infty$ .

---

1. Student de paramètre  $n$  veut dire de loi à densité  $f_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ .

**Exercice 15 (Convolution).** Montrer que si  $X$  et  $Y$  à valeurs dans  $\mathbb{R}^d$  sont indépendantes de densité respectives  $f$  et  $g$  par rapport à Lebesgue, alors  $X + Y$  a une densité par rapport à Lebesgue donnée par le produit de convolution de  $f$  et  $g$ ,

$$f * g(x) := \int f(x - y)g(y)dy.$$

**Exercice 16 (Entropie relative).** Soit  $\mu$  et  $\nu$  deux mesures de probabilité sur  $(E, \mathcal{F})$ . L'entropie relative, ou divergence de Kullback-Leibler, de  $\mu$  par rapport à  $\nu$  est définie de la façon suivante :

$$H(\mu|\nu) := \begin{cases} \int \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} d\nu & \text{si } \mu \text{ est absolument continue par rapport à } \nu, \\ +\infty & \text{sinon.} \end{cases}$$

Si  $X$  et  $Y$  sont des variables aléatoires de lois  $\mu$  et  $\nu$ , on écrira aussi  $H(X|Y)$ .

- (a) Montrer qu'il existe une variable aléatoire  $Z \geq 0$  telle que  $\mathbb{E}[Z] = 1$  et

$$H(\mu|\nu) = \mathbb{E}[Z \log Z],$$

et en déduire que  $H(\mu|\nu) \geq 0$ .

- (b) En admettant qu'on a égalité dans l'inégalité de Jensen  $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$  si et seulement si,  $\varphi(x) = Ax + B$  ou  $X$  est une constante p.s, montrer que  $H(\mu|\nu) = 0$  si et seulement si  $\mu = \nu$ .
- (c) Si  $X$  et  $Y$  ont respectivement pour densité  $f$  et  $g$  par rapport à une mesure de référence  $\eta$ , exprimer  $H(X|Y)$  en terme de  $f$  et  $g$ .
- (d) Si  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$  sont des gaussiennes réelles, calculer  $H(X_1|X_2)$ .

### 3 Espérance conditionnelle

Dans cette section, on fixe un espace probabilisé de référence  $(\Omega, \mathcal{F}, \mathbb{P})$ . Aussi,  $X$  sera toujours une variable à valeurs réelles. On identifiera deux variables aléatoires qui sont égales p.s. Ainsi,  $X \in L^2$  représentera la classe d'équivalences des variables aléatoires de carré intégrable égales à  $X$  p.p.

#### 3.1 Motivations et exemple

**Rappels sur les probabilités conditionnelles.** Si  $A, B \in \mathcal{F}$ , on note

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

L'application  $\mathbb{P}_B : A \mapsto \mathbb{P}(A|B)$  est une mesure de probabilité sur  $(\Omega, \mathcal{F})$  et on note

$$\mathbb{E}[X|B] := \int X \, d\mathbb{P}_B = \frac{1}{\mathbb{P}(B)} \int_B X \, d\mathbb{P}$$

pour une fonction mesurable  $X : \Omega \rightarrow \mathbb{R}_+$ . L'espérance conditionnelle  $\mathbb{E}[X|B]$  correspond à la "valeur moyenne de  $X$  sachant l'information  $B \in \mathcal{F}$ ". Si la famille  $(B_i)_{i \in I}$  forme une partition dénombrable disjointe de  $\Omega$  alors on a, pour tout  $A \in \mathcal{B}(\mathbb{R})$ ,

$$\mathbb{P}(X \in A) = \sum_{i \in I} \mathbb{P}(X \in A | B_i) \mathbb{P}(B_i) \quad \text{et} \quad \mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[X | B_i] \mathbb{P}(B_i).$$

**Le but de ce chapitre.** En probabilités il est important de définir la notion de "valeur moyenne de  $X$  sachant une collection d'évènements  $\mathcal{H}$ " au delà d'un seul évènement  $B$ , qui est formalisé en l'espérance conditionnelle de  $X$  sachant une sous-tribu  $\mathcal{H} \subset \mathcal{F}$  et notée  $\mathbb{E}[X|\mathcal{H}]$ . Une application importante est le cas où  $\mathcal{H} := \sigma(Y)$ . En effet,  $\mathbb{E}[X|\sigma(Y)]$ , qu'on notera plutôt  $\mathbb{E}[X|Y]$ , représente alors la valeur moyenne de  $X$  sachant l'information potentiellement apportée par  $Y$ . Attention, comme  $Y$  est aléatoire,  $\mathbb{E}[X|Y]$  sera aussi une variable aléatoire car  $Y$  n'est pas encore réalisée. Aussi, un autre cas d'utilisation importante est suivant : si  $(X_t)_{t \geq 0}$  est une suite de variables aléatoires indexées par une variable temporelle  $t$  dans  $\mathbb{N}$  ou  $\mathbb{R}_+$  et  $\mathcal{F}_t := \sigma(X_s : s < t)$ , alors  $\mathbb{E}[X_t | \mathcal{F}_{t_0}]$  avec  $0 \leq t_0 < t$  représente la valeur moyenne de  $X_t$  sachant le passé jusqu'au temps  $t_0$ .

**Tribus engendrées par des variables aléatoires.** Rappelons que, si  $Y$  est une variable aléatoire à valeurs dans  $(E, \mathcal{T})$ , alors  $\sigma(Y) \subset \mathcal{F}$  est la plus petite tribu qui rend  $Y$  mesurable, c'est-à-dire (vérifiez-le!)

$$\sigma(Y) = \{Y^{-1}(A) : A \in \mathcal{T}\}.$$

Le résultat suivant est important pour comprendre la suite :

**Lemma 3.1.** *Soit  $Y$  une variable aléatoire à valeurs dans  $(E, \mathcal{T})$ .  $h : \Omega \rightarrow \mathbb{R}$  est  $\sigma(Y)$ -mesurable ssi il existe  $f : E \rightarrow \mathbb{R}$  mesurable tel que  $h = f(Y)$ .*



*Démonstration.* Si  $f : E \rightarrow \mathbb{R}$  est mesurable, alors  $f(Y)$  est  $\sigma(Y)$ -mesurable : pour tout  $B \in \mathcal{B}(\mathbb{R})$ , on a  $f(Y)^{-1}(B) = Y^{-1}(f^{-1}(B)) \in \sigma(Y)$ . Réciproquement, si  $h = 1_B$  pour  $B \in \mathcal{F}$  alors  $h$  est  $\sigma(Y)$ -mesurable ssi  $B \in \sigma(Y)$  i.e.  $B = Y^{-1}(A)$  pour un  $A \in \mathcal{T}$ , et donc  $h = \mathbf{1}_{Y^{-1}(A)} = f(Y)$  où  $f(y) := \mathbf{1}_{y \in A}$  est une fonction mesurable. Par linéarité, cela est toujours vrai pour les fonction  $h$  étagées, puis par approximation croissante à toutes les fonctions mesurables positives, par exemple en écrivant que

$$f = \sup_{n \geq 1} f_n \quad f_n(x) := \min\left(\frac{1}{2^n} \lfloor 2^n f(x) \rfloor, n\right),$$

où  $\lfloor x \rfloor$  est la partie entière d'un nombre  $x$ , et en notant que  $f_n$  est  $\sigma(Y)$ -mesurable pour tout  $n$  dès que  $X$  l'est (il suffit de vérifier que  $\lfloor g \rfloor$  mesurable pour tout  $g : \Omega \rightarrow \mathbb{R}$ ). Ce résultat s'étend à toute fonction mesurables  $f$  en écrivant avec  $f = f^+ - f^-$ .  $\square$

Si  $(Y_i)_{i \in I}$  est une famille quelconque de variables aléatoires à valeurs dans  $(E_i, \mathcal{F}_i)$ , on notera plus généralement

$$\sigma(Y_i : i \in I) := \sigma(Y_i^{-1}(A_i) : A_i \in \mathcal{F}_i, i \in I).$$

Dans le cas où l'ensemble d'indices  $I$  est dénombrable, on peut appliquer le résultat précédent à la variable aléatoire  $(Y_i : i \in I)$  à valeurs dans l'espace produit  $\prod_{i \in I} E_i$  et obtenir que  $X$  est  $\sigma(Y_i : i \in I)$ -mesurable si et seulement si  $X = f(Y_i : i \in I)$  pour une fonction  $f : \prod_{i \in I} E_i \rightarrow \mathbb{R}$  mesurable.

**L'espérance conditionnelle par l'exemple.** Considérons la variable  $X$  définie de la façon suivante : Soit  $Y$  une variable de Bernoulli de paramètre  $p \in ]0, 1[$ . Si  $Y = 1$  alors on tire  $X$  suivant une loi exponentielle  $\mathcal{E}(1/\lambda_1)$  et sinon on tire  $X$  de loi  $\mathcal{E}(1/\lambda_0)$  ; on dit que  $X$  est un mélange de lois exponentielles. On a donc  $\mathbb{E}[X|Y = 1] = \lambda_1$  et  $\mathbb{E}[X|Y = 0] = \lambda_0$ . On définit alors la moyenne conditionnelle de  $X$  sachant  $Y$  comme

$$\mathbb{E}[X|Y] := \lambda_Y = \lambda_1 Y + \lambda_0(1 - Y),$$

qui est une variable aléatoire. Notez que  $\mathbb{E}[X|Y]$  est une fonction mesurable de  $Y$  et donc  $\sigma(Y)$ -mesurable. De plus, pour toute fonction  $f$  mesurable, on a d'une part

$$\mathbb{E}[f(Y) \mathbb{E}[X|Y]] = p f(1) \lambda_1 + (1 - p) f(0) \lambda_0,$$

et d'autre part,

$$\begin{aligned} \mathbb{E}[f(Y)X] &= p \mathbb{E}[f(Y)X|Y = 1] + (1 - p) \mathbb{E}[f(Y)X|Y = 0] \\ &= p f(1) \lambda_1 + (1 - p) f(0) \lambda_0. \end{aligned}$$

On a ainsi

$$\mathbb{E}[f(Y) \mathbb{E}[X|Y]] = \mathbb{E}[f(Y)X]$$

pour toute fonction mesurable  $f$ . En prenant  $f(y) := \mathbf{1}_{y \in B}$  pour  $B \in \mathcal{T}$ , on voit que

$$\int_A \mathbb{E}[X|Y] d\mathbb{P} = \int_A X d\mathbb{P}$$

pour tout  $A \in \sigma(Y)$ . On va voir que cette dernière propriété, avec le fait que  $\mathbb{E}[X|Y]$  est  $\sigma(Y)$ -mesurable, caractérise l'espérance conditionnelle.

### 3.2 Espérance conditionnelle par rapport à une sous-tribu

**Théorème 3.2 (Existence et unicité de l'espérance conditionnelle).** Si  $\mathcal{H} \subset \mathcal{F}$  est une sous-tribu et  $X \in L^1$ , alors il existe une unique variable aléatoire, notée  $\mathbb{E}[X|\mathcal{H}]$ , telle que :

1.  $\mathbb{E}[X|\mathcal{H}]$  est  $\mathcal{H}$ -mesurable.
2. Pour tout  $A \in \mathcal{H}$ ,

$$\int_A \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_A X d\mathbb{P}$$

Rappelons que ici unique veut dire que si  $Z$  est une autre variable aléatoire satisfait

1. et 2. alors  $Z = \mathbb{E}[X|\mathcal{H}]$  p.s.

**Remarque 3.3.** Par approximations, le point 2. est équivalent à : Pour toute fonction  $f : \Omega \rightarrow \mathbb{R}_+$  qui est  $\mathcal{H}$ -mesurable,

$$\int f \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int f X d\mathbb{P}. \quad (3.1)$$

**Notation :** On écrira aussi  $\mathbb{P}(A|\mathcal{H}) := \mathbb{E}[\mathbf{1}_A|\mathcal{H}]$ . Attention, c'est une variable aléatoire ! Aussi, si  $Y$  est une variable aléatoire à valeurs dans  $(E, \mathcal{T})$ , alors on notera  $\mathbb{E}[X|Y]$  plutôt que  $\mathbb{E}[X|\sigma(Y)]$ .

*Démonstration.* On commence par l'unicité : Si  $Z$  et  $Z'$  satisfont 1. et 2. alors, pour tout  $\varepsilon > 0$  on a  $A_\varepsilon := \{\omega \in \Omega : Z(\omega) - Z'(\omega) \geq \varepsilon\} \in \mathcal{H}$ , et

$$0 = \int_{A_\varepsilon} Z d\mathbb{P} - \int_{A_\varepsilon} Z' d\mathbb{P} = \int_{A_\varepsilon} (Z - Z') d\mathbb{P} \geq \varepsilon \mathbb{P}(A_\varepsilon).$$

Ainsi,  $Z < Z' + \varepsilon$  p.s. et, comme cela est vrai pour tout  $\varepsilon > 0$ , on obtient  $Z \leq Z'$  p.s. Comme l'argument est symétrique en  $Z$  et  $Z'$  on obtient de même  $Z = Z'$  p.s.

Pour l'existence, on peut supposer que  $X \geq 0$  p.s. (sinon, on travaille avec  $X^+$  et  $X^-$ ). Notons  $\mathbb{P}|_{\mathcal{H}}$  la restriction de  $\mathbb{P}$  à la tribu  $\mathcal{H}$ , qui est toujours une mesure de probabilité, et on définit la mesure  $\eta$  sur  $(\Omega, \mathcal{H})$  par  $\eta(A) := \int_A X d\mathbb{P}$ . C'est donc une mesure positive et finie, car  $X \in L^1$  par hypothèse. Si  $A$  est tel que  $\mathbb{P}|_{\mathcal{H}}(A) = 0$ , alors  $\mathbf{1}_A = 0$   $\mathbb{P}$ -p.p. et donc  $\eta(A) = 0$ . Le théorème de Radon-Nikodym nous assure alors de l'existence d'une fonction  $\mathcal{H}$ -mesurable  $Z : \Omega \rightarrow \mathbb{R}$  telle que, pour tout  $A \in \mathcal{H}$ ,

$$\int_A X d\mathbb{P} = \eta(A) = \int_A Z d\mathbb{P}|_{\mathcal{H}} = \int_A Z d\mathbb{P},$$

et donc  $Z$  satisfait bien 1. et 2. □

**Remarque 3.4** (Interpretation géométrique). L'espace des variables aléatoires  $L^2$  équipé du produit scalaire  $\langle X, Y \rangle = \int XY d\mathbb{P}$  est un espace de Hilbert (dès qu'on identifie deux variables aléatoires p.s. égales). Si note  $L^2_{\mathcal{H}}$  le sous-espace vectoriel des fonction  $\mathcal{H}$ -mesurables qui sont dans  $L^2$ , alors (3.1) donne que  $\langle f, X - \mathbb{E}[X|\mathcal{H}] \rangle = 0$  pour tout  $f \in L^2_{\mathcal{H}}$ . Si l'on suppose que  $X \in L^2$ , alors  $\mathbb{E}[X|\mathcal{H}] \in L^2_{\mathcal{H}}$  représente donc le projeté orthogonal de  $X$  sur  $L^2_{\mathcal{H}}$ , c'est-à-dire l'unique élément de  $L^2(\mathcal{H})$  qui satisfait

$$\min_{h \in L^2_{\mathcal{H}}} \|X - h\|_{L^2} = \|X - \mathbb{E}[X|\mathcal{H}]\|_{L^2}.$$

En effet, si on note  $P_{\mathcal{H}}(X)$  la projection orthogonale de  $X$  sur  $L^2_{\mathcal{H}}$  est caractérisée par  $X = P_{\mathcal{H}}(X) + X - P_{\mathcal{H}}(X)$  avec  $P_{\mathcal{H}}(X) \in L^2_{\mathcal{H}}$  et  $X - P_{\mathcal{H}}(X) \in (L^2_{\mathcal{H}})^{\perp}$ , c'est-à-dire  $\langle X - P_{\mathcal{H}}(X), f \rangle = 0$  pour tout  $f \in L^2_{\mathcal{H}}$ , et du coup pour tout  $f \in L^2(\mathcal{H})$ , on a :

$$\|X - f\|_{L^2}^2 = \|X - P_{\mathcal{H}}(X)\|_{L^2}^2 + \|P_{\mathcal{H}}(X) - f\|^2 \geq \|X - P_{\mathcal{H}}(X)\|_{L^2}^2.$$

Par unicité, on a bien  $\mathbb{E}[X|\mathcal{H}] = P_{\mathcal{H}}(X)$  et en particulier  $\mathbb{E}[X|\mathcal{H}] \in L^2$ . Ainsi,  $\mathbb{E}[X|Y]$  est la meilleure approximation de  $X$  par une variable  $\mathcal{H}$ -mesurable au sens  $L^2$ .

**Remarque 3.5.** Alternativement, on peut montrer que l'espace  $L^2_{\mathcal{H}}$  est un sous-espace vectoriel (donc convexe) qui est fermé dans  $L^2$ , et le théorème de projection sur un convexe fermé garantit l'existence de  $P_{\mathcal{H}}(X)$  pour tout  $X \in L^2$ . On définit alors  $\mathbb{E}[X|Y] := P_{\mathcal{H}}(X)$  et on étend cette définition aux fonctions  $L^1$  par approximation, qui satisfait alors 2. L'unicité établie en 1. montre que ces deux définitions sont identiques.

### 3.3 Propriétés de base

Quelques propriétés qui découlent de la définition.

**Corollaire 3.6.** Si  $X \in L^1$  alors  $\mathbb{E}[X|\mathcal{H}] \in L^1$  et  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[X]$ .

*Démonstration.* Il suffit de prendre  $A = \Omega$  dans 2. □

**Corollaire 3.7.** Si  $X \in L^1$  est  $\mathcal{H}$ -mesurable, alors  $\mathbb{E}[X|\mathcal{H}] = X$  p.s.

*Démonstration.* La variable  $X$  satisfait 1. et 2. donc le théorème précédent conclut. □

De la même façon, on montre que :

**Corollaire 3.8.** Si  $X, Y$  sont deux variables réelles  $L^1$  : Pour tout  $a, b \in \mathbb{R}$  on a :

$$\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}] \quad p.s.$$

**Corollaire 3.9.** Soient  $X, Y \in L^1$  réelles. Si  $X \leq Y$  alors  $\mathbb{E}[X|\mathcal{H}] \leq \mathbb{E}[Y|\mathcal{H}]$ .

En particulier,  $\mathbb{E}[X|\mathcal{H}] \geq 0$  dès que  $X \geq 0$ .

*Démonstration.* Si  $X \leq Y$  alors pour tout  $A \in \mathcal{H}$  on a  $\mathbb{E}[\mathbf{1}_A X] \leq \mathbb{E}[\mathbf{1}_A Y]$ , et donc  $\mathbb{E}[\mathbf{1}_A \mathbb{E}[X|\mathcal{H}]] \leq \mathbb{E}[\mathbf{1}_A \mathbb{E}[Y|\mathcal{H}]]$  avec 2. En prenant  $A = \{\mathbb{E}[X|\mathcal{H}] > \mathbb{E}[Y|\mathcal{H}]\}$ , on obtient

$$\mathbb{E}\left[\underbrace{\mathbf{1}_A(\mathbb{E}[X|\mathcal{H}] - \mathbb{E}[Y|\mathcal{H}])}_{\geq 0}\right] = 0$$

ce qui implique que  $\mathbb{P}(A) = 0$  ou  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[Y|\mathcal{H}]$ , et donc  $\mathbb{E}[X|\mathcal{H}] \leq \mathbb{E}[Y|\mathcal{H}]$  p.s. □

**Proposition 3.10.** Si  $X \in L^1$  et  $Y$  est une autre variable aléatoire réelle  $\mathcal{H}$ -mesurable telle que  $XY \in L^1$  alors

$$\mathbb{E}[XY|\mathcal{H}] = Y\mathbb{E}[X|\mathcal{H}] \quad p.s.$$

En particulier,

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{H}]]$$

*Démonstration.* On suppose que  $Y$  est une fonction caractéristique,  $Y = \mathbf{1}_B$  avec  $B \in \mathcal{H}$ . D'abord,  $\mathbf{1}_B \mathbb{E}[X|\mathcal{H}]$  est  $\mathcal{H}$ -mesurable car le produit de deux fonctions mesurables. Ensuite, on a pour tout  $A \in \mathcal{H}$ ,

$$\int_A \mathbf{1}_B \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_{A \cap B} \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_{A \cap B} X d\mathbb{P} = \int_A \mathbf{1}_B X d\mathbb{P},$$

ce qui montre bien que  $\mathbf{1}_B \mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbf{1}_B X|\mathcal{H}]$  p.s. Le cas d'une variable  $\mathcal{H}$ -mesurable  $Y$  générale s'obtient par limite de fonction étagées en utilisant le théorème de convergence monotone. □

La situation diamétralement opposée est celle où  $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A]$  pour tout  $A \in \mathcal{H}$ ; on dit alors que  $X$  est *indépendante de  $\mathcal{H}$* .

NB : Si  $\mathcal{H} = \sigma(Y)$ , cela revient à supposer que  $X$  et  $Y$  sont indépendantes.

**Corollaire 3.11.** *Si  $X \in L^1$  est indépendante de  $\mathcal{H}$  alors  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$  p.s. Ainsi, si  $X$  et  $Y$  sont indépendantes, alors  $\mathbb{E}[XY] = \mathbb{E}[X]$ .*

*Démonstration.* En effet, comme  $\mathbb{E}[X]$  est une constante elle est  $\mathcal{H}$ -mesurable et de plus  $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A] = \int_A \mathbb{E}[X] d\mathbb{P}$ . □

**Proposition 3.12.** *Si  $\mathcal{H}' \subset \mathcal{H} \subset \mathcal{F}$  sont des sous-tribus, alors*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{H}|\mathcal{H}']] = \mathbb{E}[X|\mathcal{H}'] \quad \text{p.s.}$$

*Démonstration.* D'abord  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}|\mathcal{H}']]$  est  $\mathcal{H}'$ -mesurable. Ensuite, en utilisant la propriété 2. de  $\mathbb{E}[\cdot|\mathcal{H}]$  on a pour tout  $A \in \mathcal{H}'$ ,

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{H}|\mathcal{H}']] d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \int_A X d\mathbb{P}$$

car  $A \in \mathcal{H}$  également. □

Finalement, si  $X$  est une variable *positive* mais pas forcément  $L^1$ , on définit :

$$\mathbb{E}[X|\mathcal{H}] := \lim_{n \rightarrow \infty} \mathbb{E}[X_n|\mathcal{H}] \quad X_n := \min(X, n).$$

D'abord,  $\mathbb{E}[X|\mathcal{H}]$  est  $\mathcal{H}$ -mesurable car limite de fonctions  $\mathcal{H}$ -mesurables. Ensuite, comme  $X_n$  est croissante,  $\mathbb{E}[X_n|\mathcal{H}]$  aussi et donc, pour tout  $A \in \mathcal{H}$ , par convergence monotone,

$$\int_A \mathbb{E}[X|\mathcal{H}] d\mathbb{P} = \lim_{n \rightarrow \infty} \int_A \mathbb{E}[X_n|\mathcal{H}] d\mathbb{P} = \lim_{n \rightarrow \infty} \int_A X_n d\mathbb{P} = \int_A X d\mathbb{P}.$$

Ainsi,  $\mathbb{E}[X|\mathcal{H}]$  satisfait les points 1. et 2. et en particulier est p.s. unique à les satisfaire.

### 3.4 Inégalités conditionnelles

**Théorème 3.13 (Inégalité de Jensen conditionnelle).** *Sous les hypothèses de l'inégalité de Jensen,*

$$\varphi(\mathbb{E}[X|\mathcal{H}]) \leq \mathbb{E}[\varphi(X)|\mathcal{H}] \quad \text{p.s.}$$

**Théorème 3.14 (Inégalité de Hölder conditionnelle).** *Sous les hypothèses de l'inégalité de Hölder,*

$$\mathbb{E}[|XY||\mathcal{H}] \leq \mathbb{E}[|X|^p|\mathcal{H}]^{\frac{1}{p}} \mathbb{E}[|Y|^q|\mathcal{H}]^{\frac{1}{q}} \quad \text{p.s.}$$

En particulier, en prenant  $p = q = 2$ , on obtient l'inégalité de Cauchy-Schwarz conditionnelle.

### 3.5 Théorèmes de convergence conditionnelle

**Théorème 3.15 (Convergence monotone conditionnelle).** *Si  $X_n$  est une suite croissante de variables aléatoires positives, alors*

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n | \mathcal{H}] = \mathbb{E}[\lim_n X_n | \mathcal{H}] \quad p.s.$$

*Démonstration.*  $\lim_n X_n$  existe p.p. et  $\mathcal{H}$ -mesurable car limite, et on vérifie 2. par convergence monotone usuelle.  $\square$

**Théorème 3.16 (Convergence dominée conditionnelle).** *Si  $X_n$  est une suite de variables aléatoires telle que  $X_n \rightarrow X$  p.s. et qu'il existe  $Z \in L^1$  telle que  $|X_n| \leq Z$  p.s., alors  $X \in L^1$  et*

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X| | \mathcal{H}] = 0 \quad p.s.$$

*En particulier,  $\mathbb{E}[X_n | \mathcal{H}] \rightarrow \mathbb{E}[X | \mathcal{H}]$  p.s.*

### 3.6 Calcul pratique d'espérances conditionnelles

On considère des variables aléatoires  $X$  et  $Y$  à valeurs dans des espaces mesurés respectifs  $(E, \mathcal{F}, \nu)$  avec  $E \subset \mathbb{R}$  et  $(\tilde{E}, \tilde{\mathcal{F}}, \tilde{\nu})$  et on suppose que le couple  $(X, Y)$  a une densité  $f(x, y)$  par rapport à  $\nu \otimes \tilde{\nu}$ . Par conséquent,  $X$  a une densité donnée  $f(x)$  par rapport à  $\nu$  et  $Y$  a une densité  $\tilde{f}(y)$  par rapport à  $\tilde{\nu}$  et on a :

$$f(x) := \int f(x, y) d\tilde{\nu}(y), \quad \tilde{f}(y) := \int f(x, y) d\nu(x).$$

**Théorème 3.17** (Calcul pratique de l'espérance conditionnelle). *On a :*

$$\mathbb{E}[X|Y] = \varphi(Y) \tag{3.2}$$

avec

$$\varphi(y) := \int x \frac{f(x, y)}{\tilde{f}(y)} d\nu(x)$$

On considère parfois

$$f(x|y) := \frac{f(x, y)}{\tilde{f}(y)},$$

qu'on appelle la *densité conditionnelle* de  $X$  par rapport à  $Y$ , et  $f(x|y) d\nu(x)$  sera alors appelée la *loi conditionnelle* de  $X$  par rapport à  $Y$ . Notez que si  $X$  et  $Y$  sont indépendantes alors  $f(x, y) = f(x)\tilde{f}(y)$  (donc  $f(x|y) = f(x)$ ) et on retrouve que  $\mathbb{E}[X|Y] = \mathbb{E}[X]$ . On note aussi

$$\mathbb{E}[X|Y = y] := \int x f(x|y) d\nu(x) = \varphi(y)$$

et on vérifie que cette définition est identique à l'espérance conditionnelle sachant l'évènement  $\{Y = y\}$  quand celui-ci est de probabilité positive. Quand  $\mathbb{P}(Y = y) = 0$ , comme c'est le cas dès que  $Y$  est une variable continue, alors la définition est nouvelle.

Attention, il faut dès fois se méfier de l'écriture  $\mathbb{E}[X|Y = y]$  car on ne peut pas travailler avec  $\mathbb{E}[X|Y] = \varphi(Y) = \mathbb{E}[X|Y = Y]$  sans s'arracher quelques cheveux.

*Démonstration.*  $\varphi(Y)$  est  $\sigma(Y)$ -mesurable car  $\varphi : \tilde{E} \rightarrow \mathbb{R}$  est mesurable. De plus, pour toute fonction  $h : \tilde{E} \rightarrow \mathbb{R}$  mesurable positive, on obtient par Fubini-Tonelli :

$$\begin{aligned}\mathbb{E}[h(Y)\varphi(Y)] &= \int h(y)\varphi(y)\tilde{f}(y)d\tilde{\nu}(y) \\ &= \iint h(y)x \frac{f(x,y)}{\tilde{f}(y)} \tilde{f}(y) d(\nu \otimes \tilde{\nu})(x,y) \\ &= \iint h(y)x f(x,y) d(\nu \otimes \tilde{\nu})(x,y) \\ &= \mathbb{E}[h(Y)X]\end{aligned}$$

et on conclut avec le théorème 3.2. □

**Exemple 1 (variables discrètes).** Si  $X$  et  $Y$  sont des variables discrètes, alors on peut prendre  $\nu$  et  $\tilde{\nu}$  les mesures de comptages appropriées, et  $f(x,y) = \mathbb{P}(X = x, Y = y)$ ,  $f(x) = \mathbb{P}(X = x)$  et  $\tilde{f}(y) = \mathbb{P}(Y = y)$ . On a donc :

$$\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = y]$$

pour tout  $\omega \in \Omega$  tel que  $Y(\omega) = y$ .

Par exemple, supposons que  $Y$  est uniforme sur  $\{1, \dots, 6\}$  (lancé d'un dé équilibré), et que  $X = Y + Z$  avec  $Z$  une Bernoulli de paramètre  $1/2$  indépendante de  $Y$ . On prend alors  $E = \{1, \dots, 7\}$  et  $\tilde{E} = \{1, \dots, 6\}$  équipés de leurs mesures de comptage respectives  $\nu$  et  $\tilde{\nu}$ . On a

$$f(x,y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y, Z = x - y) = \frac{1}{12} (\mathbf{1}_{y=x} + \mathbf{1}_{y=x-1})$$

et  $\tilde{f}(y) = 1/6$  pour tout  $y \in \{1, \dots, 6\}$ . Du coup,

$$\begin{aligned}\mathbb{E}[X|Y] &= \sum_{x \in E} x \frac{f(x,Y)}{\tilde{f}(Y)} \\ &= \frac{1}{2} \sum_{x=1}^7 x (\mathbf{1}_{Y=x} + \mathbf{1}_{Y=x-1}) \\ &= \frac{1}{2} \sum_{x=1}^7 x \mathbf{1}_{Y=x} + \frac{1}{2} \sum_{x=1}^7 x \mathbf{1}_{Y=x-1} \\ &= \frac{1}{2} \sum_{x=1}^7 Y \mathbf{1}_{Y=x} + \frac{1}{2} \sum_{x=1}^7 Y \mathbf{1}_{Y=x-1} \\ &= \frac{1}{2} Y + \frac{1}{2} (Y + 1) = Y + \frac{1}{2}.\end{aligned}$$

Dans ce cas précis, une façon plus rapide de faire ce calcul est d'écrire que :

$$\mathbb{E}[X|Y] = \mathbb{E}[Y + Z|Y] = \mathbb{E}[Y|Y] + \mathbb{E}[Z|Y] = Y + \mathbb{E}[Z] = Y + \frac{1}{2}.$$

**Exemple 2 (variables continues).** Si  $X$  et  $X'$  sont des copies indépendantes de loi exponentielle  $\mathcal{E}(\lambda)$  et  $Y := X + X'$ , alors on peut prendre  $E = E' = \mathbb{R}$  et  $\nu = \tilde{\nu} = \text{Leb}$ . Pour calculer la densité de  $(X, Y)$  on écrit, pour une fonction  $h : (\mathbb{R}_+)^2 \rightarrow \mathbb{R}_+$  mesurable

$$\begin{aligned}\mathbb{E}[h(X, Y)] &= \mathbb{E}[h(X, X + X')] \\ &= \int_{\mathbb{R}_+^2} h(x, x + x') \lambda^2 e^{-\lambda(x+x')} dx dx' \\ &= \int_0^\infty \int_x^\infty h(x, y) \lambda^2 e^{-\lambda y} dx dy\end{aligned}$$

et donc  $f(x, y) = \lambda^2 e^{-\lambda y} \mathbf{1}_{[0, y]}(x) \mathbf{1}_{\mathbb{R}_+}(y)$ . Aussi,

$$\tilde{f}(y) = \int f(x, y) dx = \lambda^2 y e^{-\lambda y} \mathbf{1}_{\mathbb{R}_+}(y)$$

et

$$\mathbb{E}[X|Y] = \int x \frac{f(x, Y)}{\tilde{f}(Y)} dx = \int_0^Y \frac{x}{Y} dx = \frac{Y}{2}.$$

**Exemple 3.** Retournons au mélange d'exponentielles du début du chapitre, où  $Y$  est une Bernoulli de paramètre  $p$  et  $X$  est de loi exponentielle de paramètre  $\lambda_1$  (resp.  $\lambda_0$ ) conditionnellement à  $Y = 1$  (resp.  $Y = 0$ ). On prend  $\nu = \text{Leb}$  sur  $E = \mathbb{R}$  et  $\tilde{\nu} = \delta_0 + \delta_1$  la mesure de comptage de  $\tilde{E} = \{0, 1\}$ . Comme

$$\begin{aligned}\mathbb{E}[h(X, Y)] &= p \mathbb{E}[h(X, 1)] + (1 - p) \mathbb{E}[h(X, 0)] \\ &= \int_0^\infty (ph(x, 1)\lambda_1 e^{-\lambda_1 x} + (1 - p)h(x, 0)\lambda_0 e^{-\lambda_0 x}) dx \\ &= \int p^y (1 - p)^{1-y} h(x, y) \lambda_y e^{-\lambda_y x} \mathbf{1}_{\mathbb{R}_+}(x) d(\nu \otimes \tilde{\nu})(x, y)\end{aligned}$$

on a donc  $f(x, y) = p^y (1 - p)^{1-y} \lambda_y e^{-\lambda_y x} \mathbf{1}_{\mathbb{R}_+}(x)$ . On sait aussi que  $\tilde{f}(y) = \mathbb{P}(Y = y) = p^y (1 - p)^{1-y}$ , ce qui donne

$$\mathbb{E}[X|Y] = \int x \frac{f(x, Y)}{\tilde{f}(Y)} dx = \int_0^\infty \lambda_Y e^{-\lambda_Y x}(x) dx = \lambda_Y.$$

### 3.7 EXERCICES – Espérance conditionnelle

**Exercice 1. (Conditionnement par rapport à une somme i.i.d).** Soit  $X_1, \dots, X_n$  des variables i.i.d de loi exponentielle  $\mathcal{E}(\lambda)$ . Calculer

$$\mathbb{E}[X_1 | X_1 + \dots + X_n].$$

Que pensez-vous du cas général où  $X_1, \dots, X_n$  sont seulement supposées i.i.d  $L^1$  ?

**Exercice 2 (Données censurées).** Supposons que  $X$  suit une loi de Poisson  $\spadesuit$  de paramètre  $\lambda$  et posons  $Y = \min(X, N)$  où  $N > 0$  est fixé. Calculer  $\mathbb{E}[X|Y]$ .

**Exercice 3 (Linéarité en la seconde variable).** Si  $X, Y, Z$  sont des variables réelles  $L^1$ , pensez-vous que  $\mathbb{E}[X|Y + Z] = \mathbb{E}[X|Y] + \mathbb{E}[X|Z]$  ?

**Exercice 4 (Vecteurs gaussiens).** Si  ${}^t(Y, X_1, \dots, X_n)$  est un vecteur gaussien  $\mathcal{N}(m, \Sigma)$  de  $\mathbb{R}^{n+1}$ , montrer qu'il existe  ${}^t(\beta_0, \beta_1, \dots, \beta_n) \in \mathbb{R}^{n+1}$  tel que :

$$\mathbb{E}[Y | X_1, \dots, X_n] = \beta_0 + \sum_{j=1}^n \beta_j X_j.$$

Exprimer les coefficients  $\beta_j$  en termes de  $m$  et  $\Sigma$ .

**Exercice 5 (Variance conditionnelle).** Si  $X \in L^2$ , on définit sa variance conditionnelle par rapport à une variable  $Y$  par

$$\text{Var}[X|Y] := \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2.$$

- (a) Montrer que  $\text{Var}[X|Y] \geq 0$  p.s.
- (b) Montrer que  $\text{Var}[X|Y] = 0$  p.s. si et seulement si  $X$  est  $\sigma(Y)$ -mesurable.
- (c) Montrer la formule “de la variance totale”,

$$\text{Var}[X] = \mathbb{E} \text{Var}[X|Y] + \text{Var} \mathbb{E}[X|Y]$$

puis interpréter géométriquement cette formule dans l'espace  $L^2$ .

- (d) Si  $Y = f(X) + \varepsilon$  avec  $f$  une fonction mesurable et  $\varepsilon$  une variable aléatoire  $L^1$  indépendante de  $X$ , calculer  $\text{Var}[Y|X]$ . Même question si  $Y = f(X) \varepsilon$ .

---

$\spadesuit$ . Rappelons que, si le temps entre deux évènements aléatoires indépendants (par exemple, “une ampoule s'éteint”) suit une loi exponentielle de paramètre  $\lambda$ , alors le nombre d'évènements arrivés pendant un laps de temps  $t$  suit une loi de Poisson de paramètre  $\lambda t$ . Si vous ne vous en rappelez pas, et c'est votre droit, démontrez-le !



## 4 Chaînes de Markov

Les chaînes de Markov sont certainement le type de processus aléatoires le plus simple après les suites de variables indépendantes, tout en étant assez polyvalentes pour inclure de nombreuses situations rencontrées dans les applications. Cette “simplicité” a pour conséquence l’omniprésence des chaînes de Markov dans le monde des algorithmes stochastiques, et de nombreux résultats théoriques disponibles assurent des garanties sérieuses quant à leur utilisation. Un exemple important, que l’on présentera ici : l’algorithme PageRank de Google, qui renvoie des pages web pertinentes ordonnées lorsque l’on tape une requête dans le célèbre moteur de recherche.

### 4.1 Processus aléatoires (généralités)

**Processus aléatoires.** Un *processus aléatoire* (ou *stochastique*) est une suite  $X = (X_n)_{n \in \mathbb{N}}$  de variables aléatoires définie sur le même espace  $(\Omega, \mathcal{F}, \mathbb{P})$  à valeurs dans un espace mesurable  $(E, \mathcal{E})$ .

La variable  $n$  est interprétée comme l’état du processus au temps  $n$ . Une réalisation  $X(\omega)$  est donc une trajectoire  $(X_0(\omega), X_1(\omega), X_2(\omega) \dots)$  indexée par la variable temps.

**Remarque 4.1.** Ici  $n \in \mathbb{N}$ , mais on pourrait aussi considérer des processus indexés par  $n \in \mathbb{Z}$  [passé infini]; c’est le cas des séries temporelles. Aussi on pourrait considérer des processus en temps continu,  $n \in [0, \infty[$  ou  $n \in \mathbb{R}$  (on appellera alors  $t$  la variable temporelle plutôt que  $n$ ); c’est le cas du mouvement Brownien ou des objets qui apparaissent en mathématiques financières.

**Exemple 4.2.** Soit  $(\xi_i)_{i \geq 1}$  une suite *i.i.d* telles que  $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$ , alors

$$\begin{aligned} X_0 &:= 0 \\ X_n &:= X_{n-1} + \xi_n = \sum_{i=1}^n \xi_i, \quad n \geq 1 \end{aligned}$$

définit un processus aléatoire sur  $\mathbb{Z}$  : la marche aléatoire simple sur  $\mathbb{Z}$ .

**Filtrations.** Associée à un processus aléatoire il y a, pour tout  $n \geq 0$ , la notion “d’information contenue par ce qu’il s’est passé jusqu’au temps  $n$ ”, mathématiquement définie par la *filtration naturelle*  $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$  associée à  $X$ ,

$$\mathcal{F}_n^X := \sigma(X_i : i \leq n) = \sigma(X_0, \dots, X_n).$$

Notez que, comme vu au chapitre précédent, on a :

$$\begin{aligned} A \in \mathcal{F}_n^X &\Leftrightarrow \mathbf{1}_A \text{ est } \mathcal{F}_n^X \text{ mesurable} \\ &\Leftrightarrow \mathbf{1}_A \text{ est une fonction mesurable de } X_0, \dots, X_n. \end{aligned}$$

Ainsi, si  $A \in \mathcal{F}_n^X$ , on peut dire à la seule connaissance d’une réalisation  $X_0(\omega), \dots, X_n(\omega)$  si oui ou non  $\omega \in A$ .

Plus généralement, une *filtration adaptée* à  $X$  est une suite de sous-tribu  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  où, pour tout  $n$ , on a  $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$  et  $X_n$  est  $\mathcal{F}_n$ -mesurable. On vérifiera que  $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$  est une filtration adaptée à  $X$ . Ici, on associera toujours au processus  $X$  sa filtration naturelle.

**Temps d'arrêt.** Une variable aléatoire  $T : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{N} \cup \{\infty\}$  est un *temps d'arrêt* si, pour tout  $n \in \mathbb{N}$ , on a  $\{T \leq n\} \in \mathcal{F}_n^X$ .

**Exemple 4.3.** Si  $A \in \mathcal{T}$ , le temps d'atteinte  $T_A := \inf\{n \geq 0 : X_n \in A\}$ , où par convention  $\inf \emptyset := \infty$ , est un temps d'arrêt. En effet,

$$\{T_A \leq n\} = \bigcup_{k=0}^n \underbrace{\{X_n \in A\}}_{\in \mathcal{F}_k^X \subset \mathcal{F}_n^X} \in \mathcal{F}_n^X.$$

En guise d'illustration, si on prend  $A = \{100, 101, 102, \dots\}$  dans l'exemple de la marche aléatoire simple sur  $\mathbb{Z}$ , le temps d'arrêt  $T_A$  représente le premier  $n$  tel que  $X_n \geq 100$ , qui est bien sûr une variable aléatoire.

Une remarque utile :

**Lemma 4.4.**  $T$  est un temps d'arrêt  $\Leftrightarrow \{T = n\} \in \mathcal{F}_n^X$  pour tout  $n \in \mathbb{N}$ .

*Démonstration.* Si  $\{T \leq k\} \in \mathcal{F}_k^X$  pour tout  $k \in \mathbb{N}$ , alors pour tout  $n \in \mathbb{N}$ ,

$$\{T = n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n^X} \cap \underbrace{\{T \leq n-1\}^c}_{\in \mathcal{F}_{n-1}^X \subset \mathcal{F}_n^X} \in \mathcal{F}_n^X.$$

Réciproquement, si  $\{T = k\} \in \mathcal{F}_k^X$  pour tout  $k \in \mathbb{N}$ , alors pour tout  $n \in \mathbb{N}$ ,

$$\{T \leq n\} = \bigcup_{k=0}^n \underbrace{\{T = k\}}_{\in \mathcal{F}_k^X \subset \mathcal{F}_n^X} \in \mathcal{F}_n^X.$$

□

**Exemple 4.5.** Si  $A \in \mathcal{T}$ , le dernier temps de passage  $\tilde{T}_A := \sup\{n \geq 0 : X_n \in A\}$  n'est pas un temps d'arrêt, car

$$\{\tilde{T}_A = n\} = \{X_n \in A\} \cap \bigcap_{k=n+1}^{\infty} \{X_k \in A\}^c.$$

## 4.2 Chaînes de Markov

*Rappel :* pour  $B \in \mathcal{F}$  et une sous-tribu  $\mathcal{H} \subset \mathcal{F}$ , on avait défini  $\mathbb{P}(B|\mathcal{H}) := \mathbb{E}[\mathbf{1}_B|\mathcal{H}]$ .

**Définition 4.6.** Un processus aléatoire  $X = (X_n)_{n \in \mathbb{N}}$  est une *chaîne de markov* si pour tout  $A \in \mathcal{T}$  et tout  $n \in \mathbb{N}$ ,

$$\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n^X) = \mathbb{P}(X_{n+1} \in A | X_n). \quad (4.1)$$

Autrement dit, la loi de  $X_{n+1}$  ne dépend du passé de la chaîne qu'au travers  $X_n$ .

*A partir de maintenant, sauf mention contraire, nous supposons dans ce chapitre que l'espace des réalisations  $E$  est discret, fini ou infini dénombrable et que  $\mathcal{T} = \mathcal{P}(E)$ .*

Dans ce contexte, (4.1) est équivalent à : pour tout  $n \geq 1$  et tout  $x_0, \dots, x_{n+1}$ ,

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

**Hypothèse d'homogénéité et matrices de transition.** On dit qu'une chaîne de Markov est *homogène* si, pour tout  $x, y \in E$ ,

$$P_{xy} := \mathbb{P}(X_{n+1} = y | X_n = x)$$

ne dépend pas de  $n$ .

*Sauf mention contraire, nous supposons toujours qu'une chaîne de Markov est homogène.*

On appelle alors

$$P := [P_{xy}]_{x,y \in E}$$

la *matrice de transition* de la chaîne de Markov. C'est une matrice stochastique, c'est à dire qui satisfait :

- (a)  $P_{xy} \geq 0$  pour tout  $x, y \in E$ .
- (b) Pour tout  $x \in E$ ,

$$\sum_{y \in E} P_{xy} = 1.$$

On vérifie  $\spadesuit$  que les opérations usuelles de “produits de matrices” et de “multiplication d'un vecteur par une matrice” peuvent être définies quand  $E$  est infini dénombrable comme dans le cas où  $E$  est fini si on se restreint aux vecteurs  $\ell^2$ . Plus précisément, Si  $A, B$  sont des matrices et  $u, v$  sont respectivement des vecteurs ligne et colonne dont les entrées sont indexées par  $E$ , alors on pose :

$$\begin{aligned} (Av)_x &:= \sum_{y \in E} A_{xy} v_y \\ (uA)_y &:= \sum_{x \in E} u_x A_{xy} \\ uAv &:= \sum_{x,y \in E} u_x A_{xy} v_y \\ (AB)_{xy} &:= \sum_{z \in E} A_{xz} B_{zy}. \end{aligned}$$

Ces quantités sont bien définies et finies pour toutes matrices stochastiques  $A, B$  et vecteurs  $u, v$  qui sont  $\ell^2$ , c'est-à-dire satisfaisant  $\sum_{x \in E} u_x^2 < \infty$  et de même pour  $v$ . On définit la puissance d'une matrice  $P$  indexée par  $E$  par  $P^0 := I$  où  $I_{xy} := \mathbf{1}_{x=y}$  et  $P^{n+1} := P P^n$ . Si  $\mu$  est une mesure de probabilité sur  $E$ , on l'identifie à un vecteur ligne dont les entrées sont indexées par  $E$ , c'est-à-dire qu'on identifie  $\mu = \sum_{x \in E} \mu_x \delta_x$  avec  $[\mu_x]_{x \in E}$ , où  $\mu_x := \mu(\{x\})$ . On notera ce vecteur ligne  $\mu$  également par commodité. On voit que  $\mu$  est  $\ell^2$  car  $\sum_{x \in E} \mu_x^2 \leq \sum_{x \in E} \mu_x = 1$ . Aussi, si  $g : E \rightarrow \mathbb{R}$  est une fonction, alors on l'identifiera au vecteur colonne  $g = [g_x]_{x \in E}$  avec  $g_x := g(x)$  et on dira que  $g$  est  $\ell^2$  si le vecteur associé l'est.

Ce formalisme permet de décrire simplement une chaîne de Markov (homogène), comme expliqué dans la proposition suivante.

---

$\spadesuit$ . Utiliser l'inégalité de Cauchy-Schwarz et le fait que comme  $P_{xy}^2 \leq P_{xy}$  car  $P_{xy} \in [0, 1]$ .

**Proposition 4.7.** *Si  $X$  est une chaîne de Markov (homogène) alors sa loi est complètement caractérisée par sa matrice de transition  $P$  et la loi  $\mu$  de  $X_0$  : Pour tout  $n \geq 1$  et  $x_0, \dots, x_n \in E$ ,*

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n}$$

*De plus, pour tout  $n \geq 1$  la loi de  $X_n$  est  $\mu P^n$  : si  $x \in E$ ,*

$$\mathbb{P}(X_n = x) = (\mu P^n)_x.$$

*En particulier, si  $g : E \rightarrow \mathbb{R}$  est  $\ell^2$ , alors*

$$\mathbb{E}[g(X_n)] = \sum_{x \in E} (\mu P^n)_x g(x) = \mu P^n g.$$

*Démonstration.* Par conditionnements successifs, on obtient en utilisant que  $X$  est une chaîne de Markov :

$$\begin{aligned} & \mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \mathbb{P}(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= P_{x_{n-1} x_n} \mathbb{P}(X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ & \quad \vdots \\ &= P_{x_{n-1} x_n} \cdots P_{x_0 x_1} \mathbb{P}(X_0 = x_0) \\ &= \mu_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x_n}. \end{aligned}$$

Ensuite, il vient pour les lois marginales :

$$\begin{aligned} \mathbb{P}(X_n = x) &= \sum_{x_{n-1}, \dots, x_0 \in E} \mathbb{P}(X_n = x, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= \sum_{x_{n-1}, \dots, x_0 \in E} \mu_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x} \\ &= (P^n \mu)_x. \end{aligned}$$

□

**Exemple 4.8.** *Si  $E = \mathbb{Z}$  et  $X$  est la marche aléatoire simple, on voit que*

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = \frac{1}{2} \mathbf{1}_{|x_{n+1} - x_n| = 1} = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

*et donc c'est une chaîne de Markov (homogène) de matrice de transition  $P_{xy} = \frac{1}{2} \mathbf{1}_{|x-y|=1}$ , qui est bi-infinie (car indexée par  $\mathbb{Z}$ ).*

### 4.3 Propriété de Markov (forte)

**Notations.** *On est souvent amené à considérer différentes chaînes de Markov avec la même matrice de transition  $P$  mais différentes lois initiales  $\mu$ . On écrira alors  $\mathbb{P}_\mu$  la loi de la chaîne de Markov ayant  $\mu$  pour loi de  $X_0$  quand on voudra souligner cette dépendance. Aussi, on écrira  $\mathbb{P}_x$  au lieu de  $\mathbb{P}_{\delta_x}$  pour la chaîne qui commence en  $x$  p.s. Enfin, on écrira  $\mathbb{E}_\mu$  (et  $\mathbb{E}_x$ ) quand on prendra l'espérance par rapport à  $\mathbb{P}_\mu$  (et  $\mathbb{P}_x$ ).*

Le résultat suivant explique que si  $X = (X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov, alors pour tout  $n \geq 1$  et  $x \in E$  fixé, conditionnellement à  $X_n = x$  le processus aléatoire  $\tilde{X} := (X_{n+m})_{m \in \mathbb{N}}$  est une chaîne de Markov de matrice de transition  $P$  partant de  $x$  qui est indépendante de  $(X_0, \dots, X_n)$ .

**Théorème 4.9** (Propriété de Markov). *Si  $X$  est une chaîne de Markov, alors pour tout  $A \in \mathcal{F}_n^X$ , pour tout  $n, m \geq 1$  et tout  $x, y_1, \dots, y_m \in E$ ,*

$$\mathbb{P}(A \cap \{X_{n+1} = y_1, \dots, X_{n+m} = y_m\} | X_n = x) = \mathbb{P}(A | X_n = x) \mathbb{P}_x(X_1 = y_1, \dots, X_m = y_m).$$

*Démonstration.* Comme  $A \in \mathcal{P}(E)$  est discret, il suffit de montrer le résultat pour un ensemble de la forme  $A = \{X_0 = x_0, \dots, X_n = x_n\}$  puisqu'on retrouve le cas général par réunion dénombrable de tels ensembles. On peut de plus supposer que  $x_n = x$  sinon l'égalité est triviale vu que les deux membres sont nuls. On a alors :

$$\begin{aligned} & \mathbb{P}(A \cap \{X_{n+1} = y_1, \dots, X_{n+m} = y_m\} | X_n = x) \\ &= \frac{1}{\mathbb{P}(X_n = x)} \mathbb{P}(X_0 = x_0, \dots, X_n = x, X_{n+1} = y_1, \dots, X_{n+m} = y_m) \\ &= \frac{1}{\mathbb{P}(X_n = x)} \mu_{x_0} P_{x_0 x_1} \cdots P_{x_{n-1} x} P_{x y_1} \cdots P_{y_{m-1} y_m} \\ &= \frac{1}{\mathbb{P}(X_n = x)} \mathbb{P}_x(X_1 = y_1, \dots, X_m = y_m) \mathbb{P}(X_0 = x_0, \dots, X_n = x) \\ &= \mathbb{P}(A | X_n = x) \mathbb{P}_x(X_1 = y_1, \dots, X_m = y_m). \end{aligned}$$

□

**Théorème 4.10** (Propriété de Markov forte). *Sous les mêmes conditions que le théorème précédent, si  $T$  est un temps d'arrêt alors conditionnellement à  $\{X_T = x\} \cap \{T < \infty\}$  le processus  $\tilde{X} := (X_{T+m})_{m \in \mathbb{N}}$  est une chaîne de Markov de matrice de transition  $P$  et de loi initiale  $\delta_x$  qui est indépendante de  $X_1, \dots, X_T$ .*

*Démonstration.* Il suffit de prouver cette identité en conditionnant de plus par  $\{T = n\}$  pour tout  $n \in \mathbb{N}$  fixé car le théorème suit alors en sommant sur  $n \in \mathbb{N}$ , mais la preuve est alors exactement la même que celle du théorème précédent. □

#### 4.4 Mesures invariantes

Quand on laisse évoluer une chaîne de Markov dans le temps, une question naturelle est de savoir si la chaîne admet une forme d'équilibre.

Considérons une chaîne de Markov  $X = (X_n)_{n \in \mathbb{N}}$  de matrice de transition  $P$ . Une mesure  $\pi$  sur  $E$  est une *mesure invariante* pour  $X$  si  $\pi P = \pi$ . Notez qu'une mesure invariante est un vecteur propre à gauche associé à la valeur propre 1 de la matrice de transition  $P$  (au moins dans le cas où  $E$  est fini) :  $\pi(P - I) = 0$ . Si  $\pi$  est une mesure de probabilité invariante alors pour tout  $x \in E$  et tout  $n \in \mathbb{N}$ , on a :

$$\mathbb{P}_\pi(X_n = x) = (\pi P^n)_x = (\pi P^{n-1})_x = \cdots = \pi_x = \mathbb{P}_\pi(X_0 = x),$$

c'est-à-dire que la loi de  $X_n$  est  $\pi$  pour tout  $n \in \mathbb{N}$ . En d'autres termes, la dynamique de la chaîne de Markov laisse invariante la loi  $\pi$  de l'état initial. Notez que si  $\pi$  est une

mesure invariante, alors  $c\pi$  pour tout  $c > 0$  est aussi une mesure invariante. Ainsi, si  $\pi$  est une mesure invariante de masse finie,  $\pi(E) := \sum_{x \in E} \pi_x < \infty$ , alors on peut construire une mesure de probabilité invariante  $\tilde{\pi}$  en posant  $\tilde{\pi}_x := \pi_x / \pi(E)$ .

Si elle existe, une telle mesure de probabilité représente un état d'équilibre de la chaîne de Markov. Plus précisément, on a le résultat suivant :

**Proposition 4.11.** *Soit  $\pi$  une mesure de probabilité sur  $E$ . On a l'équivalence entre :*

- (a) *Il existe une mesure de probabilité  $\mu$  sur  $E$  telle que, si  $X_0$  a pour loi  $\mu$ , alors  $X_n$  converge quand  $n \rightarrow \infty$  en loi vers une variable aléatoire de loi  $\pi$ .*
- (b)  *$\pi$  est une mesure invariante.*

*Démonstration.* On a déjà montré que (b) $\Rightarrow$ (a) en prenant  $\mu = \pi$ . Dans l'autre sens, (a) est équivalent à : pour toute fonction  $g$  (continue) bornée, on a :

$$\mathbb{E}[g(X_n)] = \mu P^n g \xrightarrow{n \rightarrow \infty} \int g d\pi = \pi g. \quad (4.2)$$

D'un côté, cela implique que  $\mu P^{n+1} g \rightarrow \pi g$  quand  $n \rightarrow \infty$ , mais en appliquant (4.2) à la fonction bornée  $Pg$ , on a aussi  $\mu P^{n+1} g = \mu P^n (Pg) \rightarrow \pi Pg = \pi g$ .  $\square$

On a déjà le résultat d'existence suivant.

**Théorème 4.12.** *Si  $E$  est fini, alors toute chaîne de Markov admet au moins une mesure de probabilité invariante.*

*Démonstration.* Notons  $N := \#E < \infty$  et introduisons l'ensemble

$$\mathcal{P} = \left\{ \mu \in \mathbb{R}^N : \mu_x \geq 0 \text{ pour tout } x \in E, \quad \sum_{x \in E} \mu_x = 1 \right\}$$

des mesures de probabilité sur  $E$ , qui est un compact de  $\mathbb{R}^N$  (fermé borné). Notons aussi  $\mu$  la loi de  $X_0$  et considérons la suite  $(\mu^{(n)})_{n \geq 1}$  de  $\mathcal{P}$  définie par

$$\mu^{(n)} := \frac{1}{n+1} (\mu + \mu P + \mu P^2 + \dots + \mu P^n).$$

Par compacité, quitte à extraire une sous-suite on peut supposer que  $\mu^{(n)}$  a une limite  $\pi \in \mathcal{P}$  quand  $n \rightarrow \infty$ . Comme

$$\mu^{(n)} P = \mu^{(n)} - \frac{1}{n+1} \mu + \frac{1}{n+1} \mu P^{n+1}$$

et que la suite  $(\mu P^{n+1})_{n \geq 1}$  est bornée, en prenant  $n \rightarrow \infty$  on obtient que  $\pi P = \pi$ .  $\square$

Cet argument ne marche plus quand  $E$  est infini car  $\mathcal{P}$  n'est plus compact. Prenez par exemple la suite de mesures de probabilité  $\nu^{(n)}$  sur  $\mathbb{N}$  définie par

$$\nu_k^{(n)} := \begin{cases} 1/n & \text{si } 0 \leq k \leq n-1 \\ 0 & \text{sinon.} \end{cases}$$

On voit que  $\nu^{(n)}(A) \rightarrow 0$  quand  $n \rightarrow \infty$  pour tout  $A \in \mathcal{P}(\mathbb{N})$  borné, ce qui exclut la convergence vers une mesure de probabilité sur  $\mathbb{N}$ .

Un premier résultat d'existence quand  $E$  est infini vient avec la notion de réversibilité.

**Chaînes réversibles.** Etant donné une mesure  $\mu$  sur  $E$ , supposons que la matrice de transition satisfait pour tout  $x, y \in E$ ,

$$\mu_x P_{xy} = \mu_y P_{yx}.$$

On dit alors que la chaîne de Markov associée à  $P$  est *réversible par rapport à  $\mu$* . Il s'avère que  $\mu$  est forcément une mesure invariante pour cette chaîne. En effet, comme  $P$  est une matrice stochastique, on a pour tout  $x \in E$  :

$$(\mu P)_x = \sum_{y \in E} \mu_y P_{yx} = \mu_x \sum_{y \in E} P_{xy} = \mu_x.$$

En particulier, si la matrice de transition est symétrique,  $P_{xy} = P_{yx}$  pour tout  $x, y \in E$ , auquel cas on dit que la chaîne est *réversible* (tout court), on voit que la mesure uniforme sur  $E$  définie par  $\mu_x := 1$  pour tout  $x \in E$  est une mesure invariante. Ainsi, une chaîne réversible sur un espace d'état  $E$  fini admet toujours pour mesure de probabilité invariante la loi uniforme sur  $E$ , définie par  $\pi_x := 1/\#(E)$  pour tout  $x \in E$ .

**Remarque 4.13.** *Toute chaîne de Markov réversible n'admet pas de mesure de probabilité invariante. Par exemple, la marche aléatoire simple sur  $\mathbb{Z}$  de matrice de transition  $P_{xy} = \frac{1}{2}\mathbf{1}_{|x-y|}$  est bien réversible, mais si  $\pi P = \pi$  alors, pour tout  $n \in \mathbb{Z}$ ,*

$$\pi_n = \frac{1}{2}\pi_{n-1} + \frac{1}{2}\pi_{n+1}$$

et donc  $\pi_{n+1} - \pi_n = \pi_n - \pi_{n-1}$ . Ainsi, pour tout  $n, m \in \mathbb{Z}$ , on a  $\pi_{n+1} - \pi_n = \pi_{m+1} - \pi_m$ . Si  $\sum_{n \in \mathbb{Z}} \pi_n = 1$ , en sommant sur  $n \in \mathbb{Z}$  l'identité précédente nécessairement  $\pi_{m+1} = \pi_m$ , pour tout  $m \in \mathbb{Z}$ , ce qui contredit l'hypothèse  $\sum_{m \in \mathbb{Z}} \pi_m = 1$ .

## 4.5 Récurrence

Pour savoir si une chaîne de Markov admet une probabilité invariante quand  $E$  est infini, il est important de quantifier si la chaîne de Markov "visite suffisamment souvent" un état donné, ce qui mène à la notion de récurrence. Plus précisément, pour  $x \in E$ , on considère le temps d'arrêt

$$\tau_x := \inf\{n \geq 1 : X_n = x\}$$

et la probabilité  $\mathbb{P}_x(\tau_x < \infty)$  que la chaîne de Markov qui part de  $x$  revienne en  $x$  après un temps fini. On dit que :

- $x \in E$  est un état *récurrent* si  $\mathbb{P}_x(\tau_x < \infty) = 1$ .
- $x \in E$  est état *transitoire* si  $\mathbb{P}_x(\tau_x < \infty) < 1$ .

Ainsi, si  $x$  est état transitoire, la probabilité  $\mathbb{P}_x(\tau_x = \infty)$  de ne jamais revenir en  $x$  est positive. Si  $x$  est transitoire, alors clairement  $\mathbb{E}_x[\tau_x] = \infty$ , mais  $x$  récurrent n'implique pas forcément que  $\mathbb{E}_x[\tau_x] < \infty$ . On introduit alors la notion suivante :

- $x \in E$  est un état *récurrent positif* si  $\mathbb{E}_x[\tau_x] < \infty$ .

Notez que  $x$  récurrent positif implique  $x$  récurrent.

Il existe une relation simple entre  $\mathbb{P}_x(\tau_x < \infty)$  et le nombre de retours en  $x$ , défini par

$$N_x := \sum_{n=1}^{\infty} \mathbf{1}_{X_n=x}.$$

**Lemma 4.14.** *Pour tout  $k \geq 1$ , on a :  $\mathbb{P}_x(N_x \geq k) = \mathbb{P}_x(\tau_x < \infty)^k$ .*

*Démonstration.* Par définition, on a

$$\mathbb{P}_x(N_x \geq 1) = \mathbb{P}(\tau_x < \infty).$$

Pour  $k \geq 1$ , on définit le temps de  $k$ -ème retour en  $x$ ,

$$\begin{aligned} \tau_x^{(k)} &:= \inf\{n > \tau_x^{(k-1)} : X_n = x\} = \tau_x^{(k-1)} + \inf\{n \geq 1 : X_{\tau_x^{(k-1)}+n} = x\} \\ \tau_x^{(0)} &:= 0 \end{aligned}$$

Notez que  $\tau_x^{(1)} = \tau_x$  et on vérifie (cf. Exercice 1) que c'est un temps d'arrêt. On a alors :

$$\mathbb{P}_x(N_x \geq k) = \mathbb{P}(\tau_x^{(k)} < \infty).$$

On calcule ensuite,

$$\begin{aligned} \mathbb{P}_x(\tau_x^{(k+1)} < \infty) &= \mathbb{P}_x(\tau_x^{(k+1)} < \infty \mid \tau_x^{(k)} < \infty) \mathbb{P}_x(\tau_x^{(k)} < \infty) \\ &= \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x^{(k+1)} = \tau_x^{(k)} + n \mid \tau_x^{(k)} < \infty) \mathbb{P}_x(\tau_x^{(k)} < \infty) \end{aligned}$$

puis, en utilisant la propriété de Markov forte,

$$\begin{aligned} \mathbb{P}_x(\tau_x^{(k+1)} = \tau_x^{(k)} + n \mid \tau_x^{(k)} < \infty) &= \mathbb{P}_x(X_{\tau_x^{(k)}+1} \neq x, \dots, X_{\tau_x^{(k)}+n-1} \neq x, X_{\tau_x^{(k)}+n} = x \mid \tau_x^{(k)} < \infty) \\ &= \mathbb{P}_x(X_1 \neq x, \dots, X_{n-1} \neq x, X_n = x) \\ &= \mathbb{P}_x(\tau_x^{(1)} = n), \end{aligned}$$

pour finalement obtenir

$$\mathbb{P}_x(\tau_x^{(k+1)} < \infty) = \sum_{n=1}^{\infty} \mathbb{P}_x(\tau_x^{(1)} = n) \mathbb{P}_x(\tau_x^{(k)} < \infty) = \mathbb{P}_x(\tau_x^{(1)} < \infty) \mathbb{P}_x(\tau_x^{(k)} < \infty).$$

Comme cela est vrai pour tout  $k \geq 1$ , on obtient  $\mathbb{P}_x(\tau_x^{(k+1)} < \infty) = \mathbb{P}_x(\tau_x < \infty)^{k+1}$  ce qui termine la preuve.  $\square$

Après avoir remarqué que

$$\mathbb{E}_x[N_x] = \sum_{k=1}^{\infty} \mathbb{P}_x(X_k = x) = \sum_{n=1}^{\infty} (P^n)_{xx}.$$

on obtient alors le corollaire utile suivant.



**Proposition 4.15.**  *$x$  est un état récurrent si et seulement si*

$$\sum_{n=1}^{\infty} (P^n)_{xx} = \infty.$$

*Démonstration.* Si  $x$  est un état récurrent, alors le lemme précédent nous montre que  $\mathbb{P}_x(N_x \geq k) = 1$  pour tout  $k \geq 1$  et donc  $\mathbb{P}_x(N_x = \infty) = 1$  ce qui entraîne  $\mathbb{E}_x[N_x] = \infty$ . Si  $x$  n'est pas récurrent alors c'est un état transitoire et  $\Pi_x := \mathbb{P}_x(\tau_x < \infty) < 1$ . Comme

$$\mathbb{P}_x(N_x = k) = \mathbb{P}_x(N_x \geq k) - \mathbb{P}_x(N_x \geq k + 1) = (1 - \Pi_x)\Pi_x^k$$

on obtient alors dans ce cas que :

$$\mathbb{E}_x[N_x] = \sum_{k=1}^{\infty} k \mathbb{P}_x(N_x = k) = \frac{\Pi_x}{1 - \Pi_x} < \infty.$$

□

La notion de récurrence est en fait une notion globale pour les états qui communiquent entre eux. Plus précisément, on dit que *deux états  $x$  et  $y$  communiquent*, et on écrit  $x \sim y$  si il existe  $n, m \geq 0$  tels que  $(P^n)_{xy} > 0$  et  $(P^m)_{yx} > 0$  (la probabilité de passer en  $x$  en  $y$  en  $m$  sauts et de  $y$  en  $x$  en  $n$  sauts avec probabilité positive). Notez que, par convention, on a  $x \sim x$  car on autorise  $n = 0$ . On vérifie facilement que  $\sim$  est une relation d'équivalence sur  $E$ , et donc qu'on peut partitionner  $E$  en classes disjointes d'états communicants.

**Lemma 4.16.** *Si  $x \sim y$  alors  $x$  récurrent (positif)  $\Leftrightarrow y$  récurrents (positif).*

Bien sûr, comme un état est soit récurrent soit transitoire, si  $x \sim y$  alors  $x$  transitoire  $\Leftrightarrow y$  transitoire.

*Démonstration.* Le cas récurrent positif est admis : montrons que  $x$  récurrent  $\Leftrightarrow y$  récurrent. Supposons que  $x \neq y$  sinon c'est trivial. Il suffit de montrer que  $x$  transitoire  $\Rightarrow y$  transitoire. Comme  $x \sim y$  il existe  $m, n \geq 1$  tels que  $(P^n)_{xy} > 0$  et  $(P^m)_{yx} > 0$ . On a alors pour tout  $k \geq 0$ ,

$$(P^{n+k+m})_{xx} \geq (P^n)_{xy}(P^k)_{yy}(P^m)_{yx}$$

et donc

$$\sum_{k=1}^{\infty} (P^k)_{yy} \leq \frac{1}{(P^n)_{xy}(P^m)_{yx}} \sum_{k=0}^{\infty} (P^{n+k+m})_{xx} \leq \frac{1}{(P^n)_{xy}(P^m)_{yx}} \sum_{k=0}^{\infty} (P^k)_{xx} < \infty.$$

□

Si tous les états communiquent, on dit que la chaîne est *irréductible*. Dans ce cas, soit tous les états sont récurrents, auquel cas on dit que la chaîne est récurrente, soit aucun état ne l'est, auquel cas la chaîne est transitoire. Si un état d'une chaîne irréductible est récurrent positif, on dit aussi que la chaîne est récurrente positive.

**Proposition 4.17.** *Si  $E$  est fini alors toute chaîne de Markov sur  $E$  admet un état  $x$  récurrent positif. En particulier, si cette chaîne de Markov est irréductible, alors elle est récurrente positive.*

*Démonstration.* Comme  $E$  est fini, il existe au moins un état  $x \in E$  visité une infinité de fois avec probabilité positive, donc  $\mathbb{E}_x[N_x] = \infty$  et  $x$  est récurrent.  $\square$

Il existe un lien fort entre récurrence (positive) et existence de mesures (de probabilité) invariantes, et un lien fort entre irréductibilité et unicité de la mesure invariante. Pour mettre en évidence ces liens on introduit, étant donné une chaîne de Markov  $X$  de matrice de transition  $P$  et  $x \in E$ , la mesure  $\pi^{(x)}$  définie par

$$\pi_y^{(x)} := \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x-1} \mathbf{1}_{X_n=y} \right], \quad y \in E.$$

Ainsi,  $\pi_y^{(x)}$  représente le temps moyen passé en  $y$  partant de  $x$  avant de revenir en  $x$ .

**Lemme 4.18.** *On a les propriétés clefs suivantes :*

- (a)  $x$  est récurrent  $\Leftrightarrow \pi^{(x)}P = \pi^{(x)}$ .
- (b)  $x$  est récurrent positif  $\Leftrightarrow \pi^{(x)}(E) < \infty$ .
- (c)  $\pi_x^{(x)} = 1$  et si  $x \sim y$  alors  $\pi_y^{(x)} > 0$ .
- (d) Si la chaîne est irréductible et si il existe une mesure invariante  $\pi$  telle que  $\pi_x = 1$  pour un  $x \in E$  récurrent, alors nécessairement  $\pi = \pi^{(x)}$ .

La preuve du lemme est reportée en fin de section. On va déjà voir ses implications quant à l'existence et l'unicité de mesures invariantes.

**Théorème 4.19.** *Si  $(X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov irréductible et récurrente, alors elle admet une mesure invariante  $\pi$  qui est de plus strictement positive, au sens où  $\pi_x > 0$  pour tout  $x \in E$ . Cette mesure invariante positive  $\pi$  est unique à constante multiplicative près.*

*Démonstration.* Pour tout  $x \in E$  fixé, comme  $(X_n)_{n \in \mathbb{N}}$  est récurrente, Lemme 4.18(a) donne que  $\pi^{(x)}$  est une mesure invariante et, par hypothèse d'irréductibilité,  $\pi^{(x)}$  est strictement positive par Lemme 4.18(c). Pour prouver l'unicité à une constante multiplicative près, considérons une mesure invariante  $\pi$  positive quelconque. En particulier  $\pi_x > 0$  et la mesure  $\tilde{\pi} := \frac{1}{\pi_x} \pi$  est une mesure invariante strictement positive qui satisfait  $\tilde{\pi}_x = 1$ . Le Lemme 4.18(d) implique alors que  $\tilde{\pi} = \pi^{(x)}$ , c'est-à-dire  $\pi = \pi_x \pi^{(x)}$ , et donc que  $\pi$  est égal à  $\pi^{(x)}$  à une constante multiplicative près.  $\square$

**Remarque 4.20.** *Toute chaîne de Markov irréductible récurrente n'admet pas de mesure de probabilité invariante. Par exemple, la marche aléatoire simple sur  $\mathbb{Z}$  de matrice de transition  $P_{xy} = \frac{1}{2} \mathbf{1}_{|x-y|=1}$  est bien irréductible, et il n'est pas dur de montrer qu'elle est récurrente (cf. Exercice 4), mais on a vu qu'elle ne pouvait pas avoir de mesure de probabilité invariante.*

**Théorème 4.21.** *Si la chaîne est irréductible et récurrente positive, alors :*

- (a) Elle admet une unique mesure de probabilité invariante  $\pi$  donnée par

$$\pi_x = \frac{1}{\mathbb{E}_x[\tau_x]}.$$

- (b) (Théorème ergodique) Pour toute fonction bornée  $f : E \rightarrow \mathbb{R}$ , on a :

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} \int f(x) \pi(dx).$$

Si une chaîne de Markov admet une mesure de probabilité invariante  $\pi$ , on s'attend à ce qu'elle donne une masse  $\pi_x$  à un état  $x \in E$  d'autant plus grande que l'état est visité souvent, c'est-à-dire dont le temps moyen de retour  $\mathbb{E}_x[\tau_x]$  est assez petit. Le théorème précédent montre qu'en fait  $\pi_x$  est précisément inversement proportionnel à  $\mathbb{E}_x[\tau_x]$ .

Le théorème ergodique est donc l'analogie de la loi des grands nombres pour les chaînes de Markov. Ce résultat, que l'on admettra, est un cas particulier d'un théorème concernant des systèmes dynamiques plus généraux que les chaînes de Markov, connu comme le théorème ergodique de Birkhoff.

*Démonstration.* On admet (b) et prouve (a) : Par le théorème précédent, on sait que la chaîne admet une mesure invariante strictement positive, unique à constante multiplicative près. Comme la chaîne est récurrente positive,  $\pi^{(x)}$ , et donc toutes les autres mesures invariantes, sont des mesures finies. Cela entraîne qu'il existe une unique mesure de probabilité invariante, strictement positive, que nous notons  $\pi$ . Pour  $x \in E$  fixé, la mesure  $\nu := \frac{1}{\pi_x} \pi$  est une mesure invariante strictement positive qui satisfait  $\nu_x = 1$ , donc égale à  $\pi^{(x)}$  par le Lemme 4.18(d). Comme  $\pi(E) = 1$  on a  $\nu(E) = \frac{1}{\pi_x}$  et comme  $\nu(E) = \pi^{(x)}(E) = \mathbb{E}_x[\tau_x]$  on obtient finalement  $\pi_x = 1/\mathbb{E}_x[\tau_x]$ .  $\square$

Ce théorème est particulièrement utilisé dans le cas où  $E$  est fini car ils suffit d'être en présence d'une chaîne irréductible. Plus précisément, on a la version plus forte de la Proposition 4.17 :

**Proposition 4.22.** *Toute chaîne de Markov irréductible sur un espace d'état  $E$  fini est récurrente positive.*

*Démonstration.* On a déjà montré en proposition Proposition 4.17 qu'une telle chaîne est récurrente. Le Théorème 4.12 nous assure aussi de l'existence d'une mesure de probabilité invariante  $\pi$ . Soit  $x \in E$  tel que  $\pi_x > 0$ . Comme la mesure  $\nu := \frac{1}{\pi_x} \pi$  est invariante et satisfait  $\nu_x = 1$ , le Lemme 4.18(d) implique que  $\nu = \pi^{(x)}$  et donc que  $\pi^{(x)}(E) = \pi(E)/\pi_x = 1/\pi_x < \infty$  ce qui montre la proposition grâce au Lemme 4.18(b).  $\square$

**Convergence des itérés de la matrice de transition :** En prenant  $f(y) := \mathbf{1}_{y=x}$  dans le théorème ergodique, on obtient sous les hypothèses du théorème

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=x} \xrightarrow[n \rightarrow \infty]{p.s.} \pi_x.$$

Quel que soit  $y \in E$ , en intégrant contre la mesure  $\mathbb{P}_y$ , le théorème de convergence dominée implique la convergence des sommes de Cesàro :

$$\frac{1}{n} \sum_{k=1}^n (P^k)_{yx} \xrightarrow[n \rightarrow \infty]{} \pi_x$$

pour tout  $x, y \in E$ . Peut-on supprimer la moyenne de Cesàro, c'est-à-dire avoir simplement

$$(P^n)_{xy} \xrightarrow[n \rightarrow \infty]{} \pi_x ?$$

En général non. Par exemple, si  $E = \{0, 1\}$  et  $X_0 := 0$  et  $X_{n+1} = X_n + \xi_n \pmod{2}$  avec  $\xi_i$  i.i.d Rademacher  $\pm 1$ , alors  $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  et on voit que  $P^{2n} = I_2$  et  $P^{2n+1} = P$  pour tout  $n \in \mathbb{N}$ , ce qui exclue la convergence des  $(P^n)_{yx}$  vers  $\pi_x$ .

On dit que la chaîne est *apériodique* si, pour tout  $x \in E$ , il existe  $N \in \mathbb{N}$  tel que  $(P^n)_{xx} > 0$  quelque soit  $n \geq N$ . Si on rajoute la condition d'apériodicité aux hypothèses du théorème précédent, alors on peut montrer (admis) qu'on a bien pour tout  $x, y \in E$ ,

$$(P^n)_{yx} \xrightarrow[n \rightarrow \infty]{} \pi_x.$$

En utilisant le théorème de convergence dominée, cela implique que  $(\mu P^n)_x \rightarrow \pi_x$  pour toute mesure de probabilité sur  $E$  et  $x \in E$ .

Pour les applications, une question importante est de savoir à quelle vitesse  $(\mu P^n)_x$  converge vers  $\pi_x$ , car nous pourrions alors utiliser  $(\mu P^n)_x$  comme approximation de la mesure invariante avec un contrôle de l'erreur. Il existe de nombreux résultats, mais un premier qui concerne essentiellement le cas  $E$  fini est le suivant :

**Théorème 4.23.** *Soit  $(X_n)$  une chaîne récurrente qui satisfait la condition de Doeblin : Il existe  $n_0 \geq 1$ ,  $\beta \in ]0, 1[$  et une probabilité  $\nu$  sur  $E$  telle que, pour tout  $x, y \in E$ ,*

$$(P^{n_0})_{xy} \geq \beta \nu_y.$$

Alors il existe une unique probabilité invariante  $\pi$  et pour toute probabilité  $\mu$  sur  $E$  et  $n \in \mathbb{N}$ ,

$$\sum_{x \in E} |(\mu P^n)_x - \pi_x| \leq 2(1 - \beta)^{\lfloor n/n_0 \rfloor}. \quad (4.3)$$

Ici  $\lfloor x \rfloor$  représente la partie entière de  $x$ .

*Démonstration.* On prouve seulement la borne (4.3) quand  $n_0 = 1$ , c'est-à-dire que l'on suppose que  $P_{xy} \geq \beta \nu_y$  pour tout  $x, y \in E$ . On peut écrire

$$P_{xy} = \beta \nu_y + (1 - \beta) S_{xy}$$

où

$$S_{xy} := \frac{1}{1 - \beta} (P_{xy} - \beta \nu_y)$$

est une matrice stochastique ; notez que  $S_{xy} \geq 0$  est équivalent à la condition de Doeblin.

L'espace des mesures de probabilité  $\mathcal{P}$  sur  $E$  est un sous-espace fermé de l'espace  $\ell^1(E)$  des suites  $u = [u_x]_{x \in E}$  qui satisfont  $\|u\|_{\ell^1} := \sum_{x \in E} |u_x| < \infty$ , qui est complet. Si  $\mu^1, \mu^2 \in \mathcal{P}$ , on a

$$\begin{aligned} (\mu^1 P)_y - (\mu^2 P)_y &= \sum_{x \in E} (\mu_x^1 - \mu_x^2) P_{xy} \\ &= \beta \nu_y \sum_{x \in E} (\mu_x^1 - \mu_x^2) + (1 - \beta) \sum_{x \in E} (\mu_x^1 - \mu_x^2) S_{xy} \\ &= (1 - \beta) \sum_{x \in E} (\mu_x^1 - \mu_x^2) S_{xy} \end{aligned}$$

et donc, en utilisant le théorème de Fubini-Tonelli et que  $S$  est une matrice stochastique,

$$\begin{aligned} \|(\mu^1 - \mu^2)P\|_{\ell^1} &= (1 - \beta) \sum_{y \in E} \left| \sum_{x \in E} (\mu_x^1 - \mu_x^2) S_{xy} \right| \\ &\leq (1 - \beta) \sum_{y \in E} \sum_{x \in E} |\mu_x^1 - \mu_x^2| S_{xy} \\ &= (1 - \beta) \sum_{x \in E} |\mu_x^1 - \mu_x^2| = (1 - \beta) \|\mu^1 - \mu^2\|_{\ell^1}. \end{aligned}$$

Ainsi, comme  $1 - \beta \in ]0, 1[$ , l'application  $F : \mathcal{P} \rightarrow \mathcal{P}$ ,  $F(\mu) := \mu P$  est une application contractante et le théorème du point fixe de Banach nous assure l'existence d'un unique point fixe  $\pi$  pour  $F$ , i.e.  $F(\pi) = \pi$ . De plus, pour tout  $\mu \in \mathcal{P}$ , on a

$$\begin{aligned} \|\mu P^n - \pi\|_{\ell^1} &= \|\mu P^n - \pi P\|_{\ell^1} \\ &\leq (1 - \beta) \|\mu P^{n-1} - \pi\|_{\ell^1} \\ &\quad \vdots \\ &\leq (1 - \beta)^n \|\mu - \pi\|_{\ell^1} \leq 2(1 - \beta)^n. \end{aligned}$$

□

On termine cette section en donnant une preuve du Lemme clef 4.18.

**Preuve du Lemme 4.18.**

*Démonstration.* Pour prouver (a), remarquons que déjà, pour tout  $y \in E$ , on a par le théorème de Fubini-Tonelli,

$$\begin{aligned} \pi_y^{(x)} &= \mathbb{E}_x \left[ \sum_{n=0}^{\infty} \mathbf{1}_{X_n=y} \mathbf{1}_{n < \tau_x} \right] \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, \tau_x > n) \end{aligned}$$

et puis, en utilisant encore Fubini-Tonelli et que  $\{\tau_x > n\} \in \mathcal{F}_n^X$ ,

$$\begin{aligned}
(\pi^{(x)}P)_y &= \sum_{z \in E} \pi_z^{(x)} P_{zy} \\
&= \sum_{z \in E} P_{zy} \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = z, \tau_x > n) \\
&= \sum_{z \in E} P_{zy} \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = z | \tau_x > n) \mathbb{P}_x(\tau_x > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_x > n) \sum_{z \in E} \mathbb{P}(X_{n+1} = y | X_n = z) \mathbb{P}_x(X_n = z | \tau_x > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_x > n) \sum_{z \in E} \mathbb{P}(X_{n+1} = y | X_n = z, \tau_x > n) \mathbb{P}_x(X_n = z | \tau_x > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_x > n) \mathbb{P}(X_{n+1} = y | \tau_x > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_{n+1} = y, \tau_x > n) \\
&= \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x-1} \mathbf{1}_{X_{n+1}=y} \right] \\
&= \mathbb{E}_x \left[ \sum_{n=1}^{\tau_x} \mathbf{1}_{X_n=y} \right] \\
&= \mathbb{E}_x \left[ \sum_{n=1}^{\tau_x} \mathbf{1}_{X_n=y} \mathbf{1}_{\tau_x < \infty} + \sum_{n=1}^{\infty} \mathbf{1}_{X_n=y} \mathbf{1}_{\tau_x = \infty} \right].
\end{aligned}$$

Comme on peut alors écrire

$$\begin{aligned}
\pi_y^{(x)} &= \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x-1} \mathbf{1}_{X_{n+1}=y} \mathbf{1}_{\tau_x < \infty} + \sum_{n=0}^{\infty} \mathbf{1}_{X_{n+1}=y} \mathbf{1}_{\tau_x = \infty} \right] \\
&= (\pi^{(x)}P)_y + \mathbb{E}_x \left[ (\mathbf{1}_{X_0=y} - \mathbf{1}_{X_{\tau_x=y}}) \mathbf{1}_{\tau_x < \infty} + \mathbf{1}_{X_0=y} \mathbf{1}_{\tau_x = \infty} \right] \\
&= (\pi^{(x)}P)_y + \mathbf{1}_{x=y} \mathbb{P}_x(\tau_x = \infty),
\end{aligned}$$

on voit qu'on a prouvé (a).

Le point (b) suit directement du calcul :

$$\pi^{(x)}(E) = \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x-1} \mathbf{1}_{X_{n+1} \in E} \right] = \mathbb{E}_x[\tau_x].$$

Pour le point (c), on a clairement (que  $\tau_x = \infty$  ou non)

$$\pi_x^{(x)} = \mathbb{E}_x \left[ \underbrace{\mathbf{1}_{X_0=x}}_{=1 \text{ p.s.}} + \underbrace{\mathbf{1}_{X_1=x} + \dots + \mathbf{1}_{X_{\tau_x-1}=x}}_{=0 \text{ p.s.}} \right] = 1$$

et ensuite, si  $n \in \mathbb{N}$  est tel que  $(P^n)_{xy} > 0$ , on a :

$$0 < (P^n)_{xy} = (P^n)_{xy} \pi_x^{(x)} \leq \sum_{z \in E} \pi_z^{(x)} (P^n)_{zy} = (\pi^{(x)}P^n)_y = \pi_y^{(x)}.$$

Montrons alors (d) : Supposons donc que  $\pi$  est une mesure invariante d'une chaîne irréductible telle que  $\pi_x = 1$  où  $x$  est récurrent et montrons que  $\pi = \pi^{(x)}$ . On a pour tout  $y \in E$ ,

$$\pi_y = (\pi P)_y = P_{xy} + \sum_{\substack{z \in E \\ z \neq x}} \pi_z P_{zy}.$$

En réutilisant cette expression pour  $\pi_z$  et en réitérant  $N$  fois l'opération, on obtient pour tout  $N \geq 1$ ,

$$\begin{aligned} \pi_y &= P_{xy} + \sum_{\substack{z_1 \in E \\ z_1 \neq x}} \pi_{z_1} P_{z_1 y} \\ &= P_{xy} + \sum_{\substack{z_1 \in E \\ z_1 \neq x}} P_{xz_1} P_{z_1 y} + \sum_{\substack{z_1, z_2 \in E \\ z_1, z_2 \neq x}} \pi_{z_2} P_{z_2 z_1} P_{z_1 y} \\ &= \vdots \\ &= \sum_{n=0}^N \sum_{\substack{z_1, \dots, z_n \in E \\ z_1, \dots, z_n \neq x}} P_{xz_n} \cdots P_{z_2 z_1} P_{z_1 y} + \sum_{\substack{z_1, \dots, z_{N+1} \in E \\ z_1, \dots, z_{N+1} \neq x}} \pi_{z_{N+1}} P_{z_{N+1} z_N} \cdots P_{z_2 z_1} P_{z_1 y} \\ &\geq \sum_{n=0}^N \sum_{\substack{z_1, \dots, z_n \in E \\ z_1, \dots, z_n \neq x}} P_{xz_n} \cdots P_{z_2 z_1} P_{z_1 y}. \end{aligned}$$

En prenant la limite  $N \rightarrow \infty$ , on obtient donc

$$\begin{aligned} \pi_y &\geq \sum_{n=0}^{\infty} \sum_{\substack{z_1, \dots, z_n \in E \\ z_1, \dots, z_n \neq x}} P_{xz_n} \cdots P_{z_2 z_1} P_{z_1 y} \\ &= \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n+1} = y, \tau_x \geq n+1) \\ &= \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x - 1} \mathbf{1}_{X_{n+1} = y} \right] \\ &= (\pi^{(x)} P)_y \\ &= \pi_y^{(x)}. \end{aligned}$$

Par conséquent  $\pi_y \geq \pi_y^{(x)}$  pour tout  $y \in E$ , et  $\nu := \pi - \pi^{(x)}$  est une mesure (signée) qui est invariante et satisfait  $\nu_x = 0$ . Par hypothèse d'irréductibilité, pour tout  $y \in E$  il existe  $n \geq 0$  tel que  $(P^n)_{yx} > 0$ , ce qui donne finalement,

$$0 = \nu_x = (\nu P)_x = \sum_{z \in E} \nu_z (P^n)_{zx} \geq \nu_y (P^n)_{yx}$$

et donc nécessairement  $\nu_y = 0$ . Ainsi,  $\nu$  est la mesure nulle, c'est-à-dire  $\pi = \pi^{(x)}$ . □

## 4.6 Application : L'algorithme PageRank

Comment, lorsqu'on envoie une requête sur un moteur de recherche, obtient-on une liste pertinente de pages web ? Rares sont les fois où l'on passe la première page de suggestions de pages web. On pourrait mesurer la "pertinence" d'une page comme le nombre d'autres pages qui pointent vers elle, ce qui était utilisé initialement, mais il est facile de corrompre cette approche en créant de nombreuses pages web fictives pour booster son score de pertinence. Il faudrait prendre en compte le poids de ces pages qui pointent vers la page pertinente, puis les pages qui pointent vers les pages qui pointent vers ses pages, etc.

L'algorithme initial du moteur de recherche de Google, PageRank ♠, s'est imposé ♡ comme une solution populaire à ce problème et a révolutionné l'efficacité des moteurs de recherche, et ce à l'aide d'une utilisation des propriétés basiques des chaînes de Markov. L'algorithme exact utilisé de nos jours par Google est secret dans les détails mais reste une variation de PageRank.

L'idée est la suivante : on considère le graphe de toutes les pages web existantes, où les sommets  $E$  sont les pages web, et on dessine une arête de la page  $x \in E$  vers la page  $y \in E$  si il y a un lien de la  $x$  qui pointe vers  $y$  ; on note dans ce cas  $x \rightarrow y$ . Notons  $d_x := |\{y \in E : x \rightarrow y\}|$  le nombre de liens émanants de la page  $x$ . On considère alors la matrice stochastique

$$L_{xy} = \frac{1}{d_x} \mathbf{1}_{x \rightarrow y}$$

qui est la matrice de transition d'une chaîne de Markov (homogène) sur  $E$ . La dynamique associée est celle d'un surfeur qui, partant d'une page initiale  $x_0$ , sélectionne un lien au hasard uniformément parmi les liens possibles sur la page  $x_0$ , et réitère ce procédé indéfiniment. Le principe de PageRank est d'associer à une page  $x$  le rang  $\pi_x \in [0, 1]$  satisfaisant  $\sum_{x \in E} \pi_x = 1$  prenant en compte l'importance relative des liens pointant vers  $x$  via l'équation :

$$\pi_x = \sum_{\substack{y \in E \\ y \rightarrow x}} \frac{\pi_y}{d_y}.$$

En d'autres termes,  $\pi$  est une mesure de probabilité invariante pour  $L$ . On a alors l'interprétation suivante :  $\pi_x$  représente le temps moyen passé par le surfeur ivre sur la page  $x$  si il surfait infiniment longtemps. Comme  $E$  est fini, une telle mesure de probabilité existe mais rien ne garantit son unicité car la chaîne n'est pas forcément irréductible. De plus, ce vecteur  $\pi$  n'est approchable en pratique qu'à partir des approximations définies par récurrence  $\mu^{(n+1)} := L\mu^{(n)}$  en partant d'un  $\mu^{(0)}$  arbitraire. Il faut alors se demander au bout de combien d'itérations est-on assez proche (par rapport à un seuil d'erreur fixé en avance) de la vraie mesure invariante  $\pi$ . Pour répondre à ces problèmes, Brin et Page ont modifié la matrice de transition en la matrice, dite de Google, suivante :

$$G_{xy} := \alpha L_{xy} + (1 - \alpha) \frac{1}{|E|}$$

où  $\alpha \in ]0, 1[$  est un paramètre, appelé facteur d'amortissement. Il semble que la paramètre  $\alpha$  de PageRank était à la base  $\alpha = 0.85$ .  $G$  est aussi une matrice de transition et la chaîne de Markov associée est décrite de la façon suivante : A chaque étape, on joue à

---

♠. développé par Larry Page et Sergey Brin en 1998, qui ont fondé Google la même année.  
 ♡. 90% des internautes environs utilisent Google comme moteur de recherche en 2018.



pile ou face, avec probabilité  $\alpha$  de faire pile. Si on obtient “pile”, alors le surfeur choisit une page comme avant, avec la matrice de transition  $L$ . Si on obtient “face”, le surfeur se téléporte sur une page au hasard uniformément parmi toutes les pages existantes. La chaîne de Markov associée à  $G$  est maintenant irréductible récurrence et satisfait de plus la condition de Doeblin : pour tout  $x, y \in E$ ,

$$G_{xy} \geq (1 - \alpha) \frac{1}{|E|}$$

L’unicité d’une mesure de probabilité invariante est assurée et les garanties théoriques du Théorème 4.23 peuvent être appliquées pour estimer  $\pi$  à une erreur fixée et donner ainsi un classement des pages web référencées par Google (cf. Exercice 11).

## 4.7 EXERCICES – chaînes de Markov

**Exercice 1.** Soit  $X = (X_n)_{n \in \mathbb{N}}$  un processus aléatoire sur un espace d'états  $E$  discret et  $x \in E$ . On suppose que  $X_0 = x$  p.s. Si on définit par récurrence

$$\tau_x^{(0)} := 0, \quad \tau_x^{(k+1)} := \inf\{n > \tau_x^{(k)} : X_n = x\},$$

que représente  $\tau_x^{(k)}$ ? Montrer que  $\tau_x^{(k)}$  est un temps d'arrêt pour tout  $k \in \mathbb{N}$ .

**Exercice 2 (Suites récurrentes avec innovation).** Soit  $(\varepsilon_n)_{n \in \mathbb{N}}$  une suite i.i.d à valeurs dans  $(\tilde{E}, \tilde{\mathcal{F}})$  de loi  $\nu$ . On considère le processus aléatoire  $X = (X_n)_{n \in \mathbb{N}}$  défini par

$$X_{n+1} = F(X_n, \varepsilon_n)$$

pour une fonction  $F : E \times \tilde{E} \rightarrow E$  mesurable, et dont la loi  $\mu$  de  $X_0$  est spécifiée, avec  $X_0$  indépendant de la suite  $(\varepsilon_n)_{n \in \mathbb{N}}$ .

- Montrer que la marche aléatoire simple (sur  $\mathbb{Z}$ ) est un exemple de tels processus.
- Montrer que  $X$  est une chaîne de Markov homogène et donner sa matrice de transition.
- On considère une file d'attente avec 0 clients au temps  $n = 0$  et on suppose que, à chaque instant  $n \geq 1$ , un client est retiré de la file d'attente et  $q_n$  nouveaux clients se rajoutent, où  $(q_n)_{n \in \mathbb{N}}$  est une suite de variables i.i.d de loi  $\nu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{6}\delta_2$ . Montrer que le processus  $(X_n)_{n \in \mathbb{N}}$  du nombre de clients à l'instant  $n$  est un processus de Markov dont on donnera la matrice transition.

**Exercice 3 (On-Off).** On considère le système dynamique  $(X_n)_{n \in \mathbb{N}}$  sur  $\{0, 1\}$  de loi initiale  $\mu = (\mu_0, \mu_1)$  et de matrice transition

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

- Quelles conditions doivent satisfaire  $a, b, c, d$ ?
- Calculer la loi de  $X_n$  puis la valeur de  $\lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0)$  en fonction de  $a$  et  $b$ .

**Exercice 4 (Classes communicantes et récurrence).** Donner les classes communicantes de la chaîne de Markov associée à la matrice de transition

$$P = \begin{bmatrix} 1/2 & 0 & 1/3 & 0 & 1/2 \\ 1/2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 1/6 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 \end{bmatrix}$$

après avoir dessiné le graphe associé.

**Exercice 5 (Récurrence de la marche aléatoire simple sur  $\mathbb{Z}$ ).** On considère  $(X_n)_{n \in \mathbb{N}}$  la marche aléatoire simple sur  $\mathbb{Z}$ , de matrice de transition  $P_{xy} = \frac{1}{2}\mathbf{1}_{|x-y|=1}$ .

- Montrer que cette chaîne de Markov est irréductible.
- Calculer  $(P^n)_{00}$  pour tout  $n \geq 1$ .

- (c) A l'aide de la formule de Stirling <sup>?</sup>, montrer que cette chaîne de Markov est récurrente.
- (d) Que pensez-vous du cas de la marche aléatoire simple non-symétrique, définie par  $X_0 := 0$  et  $X_{n+1} = X_n + \varepsilon_n$  où  $(\varepsilon_n)_n \in \mathbb{N}$  est une suite de variables i.i.d. telle que  $\mathbb{P}(\varepsilon_n = 1) = 1 - \mathbb{P}(\varepsilon_n = -1) = p$  où  $p \in ]0, 1[$  et  $p \neq 1/2$ ?

**Exercice 6 (Loi des excursions).** Soit  $(X_n)$  une chaîne de Markov. On fixe  $x \in E$  que l'on suppose récurrent et on considère pour tout  $k \geq 1$  le temps d'arrêt  $\tau_x^{(k)}$  introduit à l'Exercice 1, à savoir le  $k$ -ième temps de retour en  $x \in E$ . On s'intéresse à la loi des excursions successives  $(E_0, E_1, \dots)$  où

$$E_k := (X_{\tau_x^{(k)}}, X_{\tau_x^{(k)}+1}, \dots, X_{\tau_x^{(k+1)}}).$$

- (a) Décrivez l'ensemble  $\mathcal{E}$  où les variables  $E_k$  prennent leurs valeurs.
- (b) Si  $X_0 := x$  p.s, montrer que la suite  $(E_0, E_1, \dots)$  est i.i.d.

**Exercice 7 (Recurrence et invariance).** Pour  $x \in E$  fixé, on considère la mesure  $\pi^{(x)}$  sur  $E$  définie par

$$\pi_y^{(x)} := \mathbb{E}_x \left[ \sum_{n=0}^{\tau_x-1} \mathbf{1}_{X_n=y} \right], \quad y \in E.$$

- (a) Montrer que  $\pi^{(x)}(E) < \infty \Leftrightarrow x$  est récurrent positif.
- (b) On veut montrer que  $P\pi^{(x)} = \pi^{(x)} \Leftrightarrow x$  est récurrent.
- (c) Si la chaîne est irréductible et si il existe une mesure invariante  $\pi$  telle que  $\pi_x = 1$  et  $x$  est récurrent, montrer que nécessairement  $\pi = \pi^{(x)}$ .

**Exercice 8.** Montrer que si  $x \sim y$  alors  $x$  récurrent positif  $\Leftrightarrow y$  récurrent positif.

**Exercice 9.** Montrer la marche aléatoire simple sur  $\mathbb{Z}$  n'est pas récurrente positive.

**Exercice 10.** Étant donné  $0 < p < 1$ , on considère la chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  sur  $E = \{1, 2, 3, 4\}$  de matrice de transition

$$P = \begin{bmatrix} p & 0 & p & 0 \\ 1-p & 0 & 1-p & 0 \\ 0 & p & 0 & p \\ 0 & 1-p & 0 & 1-p \end{bmatrix}.$$

- (a) Montrer que la chaîne est irréductible récurrente positive.
- (b) Calculer son unique probabilité invariante.
- (c) Calculer  $P^2$  et en déduire la loi de  $X_n$  pour tout  $n \geq 2$ .
- (d) Calculer  $\mathbb{E}_4[\tau_4]$ .

**Exercice 11 (PageRank ; application numérique).** A partir de quel  $n$  la matrice  $G^n$  donne une approximation à  $10^{-5}$  du classement théorique des pages web ?

?. Formule de Stirling :  $n! \sim \sqrt{2\pi n}(n/e)^n$  quand  $n \rightarrow \infty$ .

**Exercice 12 (Algorithme de Metropolis-Hastings).** Soit  $\pi$  une mesure de probabilité sur un ensemble  $E$  fini. On cherche à construire un algorithme qui renvoie une approximation numérique de  $\int f d\pi$  pour une fonction  $f : E \rightarrow \mathbb{R}$  donnée.

- Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov irréductible sur  $E$  de matrice de transition  $Q$  telle que  $Q_{xy} > 0 \Leftrightarrow Q_{yx} > 0$ .
- Pour tout  $x, y \in E$  tels que  $Q_{xy} > 0$ , on définit

$$R_{xy} := \min \left( \frac{\pi_x Q_{yx}}{\pi_y Q_{xy}}, 1 \right).$$

- On construit finalement la chaîne de Markov  $(\tilde{X}_n)_{n \in \mathbb{N}}$  :
  - on tire la variable initiale  $\tilde{X}_0$  de façon arbitraire.
  - pour tout  $n \geq 1$ , on construit  $\tilde{X}_n$  à partir de  $\tilde{X}_{n-1}$  et d'une variable aléatoire  $U_n$  uniforme sur  $[0, 1]$  indépendante de  $(X_n, \tilde{X}_{n-1})$  comme :

$$\tilde{X}_n := X_n \mathbf{1}_{U_n \leq R_{X_n \tilde{X}_{n-1}}} + \tilde{X}_{n-1} \mathbf{1}_{U_n > R_{X_n \tilde{X}_{n-1}}}$$

- (a) Expliquez la construction de  $\tilde{X}_n$  comme si vous vouliez l'implémenter dans un script, puis donner la matrice de transition  $P$  de  $(\tilde{X}_n)_{n \in \mathbb{N}}$ .
- (b) Montrer que, pour tout  $f : E \rightarrow \mathbb{R}$ , on a :

$$\frac{1}{n} \sum_{k=1}^n f(\tilde{X}_k) \xrightarrow[n \rightarrow \infty]{p.s.} \int f d\pi.$$

- (c) Est-il important que  $\pi$  soit une mesure de probabilité ?

**Exercice 13 (Marche aléatoire simple sur  $\mathbb{Z}^d$ ).** On considère  $(X_n)_{n \in \mathbb{N}}$  la marche aléatoire simple sur  $\mathbb{Z}^d$  définie par  $X_0 := 0$  et  $X_n = X_{n-1} + \xi_n$  pour  $n \geq 1$ , où  $(\xi_n)_{n \geq 1}$  sont i.i.d sur  $\mathbb{Z}^d$  telles que  $\mathbb{P}(\xi_n = \pm e_j) = 1/2d$  avec  $e_1, \dots, e_d$  la base canonique.

- (a) Montrer que si  $X$  est une variable aléatoire à valeur dans  $\mathbb{Z}^d$  alors

$$\mathbb{P}(X = 0) := \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \varphi_X(t) dt.$$

- (b) Montrer que la fonction caractéristique de  $\xi_n$  est donnée par

$$\varphi(t) = \frac{1}{d} \sum_{j=1}^d \cos(t_j)$$

et en déduire une expression pour  $(P^n)_{00}$  pour tout  $n \geq 1$ .

- (c) Montrer que pour tout  $0 < z < 1$ , on a :

$$\sum_{n=0}^{\infty} z^n (P^n)_{00} = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1}{1 - z\varphi(t)} dt =: I(z).$$

- (d) Montrer que  $(X_n)_{n \in \mathbb{N}}$  est récurrente  $\Leftrightarrow \lim_{z \rightarrow 1} I(z) = \infty$ .

- (e) On pose  $F(z, t) := \frac{1}{1 - z\varphi(t)}$ . Montrer que :

— pour tout  $c > 0$ ,  $(z, t) \mapsto F(z, t)$  est bornée sur  $]0, 1] \times \{t \in \mathbb{Z}^d : \|t\| \geq c\}$ .

- donner le comportement asymptotique de  $F(1, t)$  quand  $t \rightarrow 0$ .
  - si  $\|t\|$  est assez petit, l'application  $z \mapsto F(z, t)$  est croissante.
- (f) Montrer que  $(X_n)_{n \in \mathbb{N}}$  est récurrente si  $d = 1$  ou  $2$  mais transitoire pour  $d \geq 3$ .

CORRECTION :

- (a) Si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{Z}^d \subset \mathbb{R}^d$ , on a :

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \sum_{n \in \mathbb{Z}^d} \mathbb{P}(X = n) e^{i\langle t, n \rangle}, \quad t \in \mathbb{R}^d.$$

Pour tout  $n = (n_1, \dots, n_d) \in \mathbb{Z}^d$ , on calcule

$$\int_{[-\pi, \pi]^d} e^{i\langle t, n \rangle} dt = \int_{[-\pi, \pi]^d} \prod_{j=1}^d e^{it_j n_j} dt_j = \prod_{j=1}^d \int_{-\pi}^{\pi} e^{it_j n_j} dt_j = \prod_{j=1}^d 2\pi \mathbf{1}_{n_j=0} = (2\pi)^d \mathbf{1}_{n=0}$$

ce qui donne

$$\frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \varphi_X(t) dt = \sum_{n \in \mathbb{Z}^d} \mathbb{P}(X = n) \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} e^{i\langle t, n \rangle} dt = \mathbb{P}(X = 0)$$

grâce au théorème de Fubini ; en effet  $(n, t) \mapsto \mathbb{P}(X = n) e^{itn}$  est  $L^1$  par rapport à la mesure  $(\sum_{n \in \mathbb{Z}^d} \delta_n) \otimes \mathbf{1}_{[-\pi, \pi]^d} dt$ .

- (b) On calcule

$$\varphi(t) = \mathbb{E}[e^{i\langle t, \xi_n \rangle}] = \frac{1}{2^d} \sum_{j=1}^d (e^{i\langle t, e_j \rangle} + e^{-i\langle t, e_j \rangle}) = \frac{1}{d} \sum_{j=1}^d \cos(t_j)$$

puis, en utilisant (a) et l'indépendance des  $\xi_n$ ,

$$(P^n)_{00} = \mathbb{P}(X_n = 0) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \varphi_{X_n}(t) dt = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \varphi^n(t) dt.$$

- (c) Puisque pour tout  $0 < z < 1$  l'application  $(n, t) \mapsto z^n \varphi(t)^n$  est  $L^1$  par rapport à la mesure  $(\sum_{n \in \mathbb{Z}^d} \delta_n) \otimes \mathbf{1}_{[-\pi, \pi]^d} dt$ , le théorème de Fubini et (b) nous donnent pour tout  $0 < z < 1$ ,

$$\sum_{n=0}^{\infty} z^n (P^n)_{00} = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \left( \sum_{n=0}^{\infty} z^n \varphi^n(t) \right) dt = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \frac{1}{1 - z\varphi(t)} dt.$$

- (d) Remarquons d'abord que la marche aléatoire sur  $\mathbb{Z}^d$  est clairement irréductible et donc que  $(X_n)$  est récurrente  $\Leftrightarrow 0$  est récurrent. Le théorème de convergence monotone donne

$$\sum_{n=0}^{\infty} (P^n)_{00} = \lim_{z \rightarrow 1} \sum_{n=0}^{\infty} z^n (P^n)_{00} = \lim_{z \rightarrow 1} I(z)$$

et on conclut par l'équivalence :  $0$  récurrent  $\Leftrightarrow \sum_{n=0}^{\infty} (P^n)_{00} = \infty$ .

- (e) — D'un côté, comme  $\cos(x) \geq -1$ , on a  $F(z, t) \geq 1/(1+z) \geq 1$  ; en particulier  $F$  est positive. Comme  $\varphi$  est continue et que, pour  $t \in [-\pi, \pi]^d$ , on  $\varphi(t) = 1 \Leftrightarrow t = 0$ , on voit que  $F$  est bornée supérieurement dès que l'on restreint  $t$  à un sous ensemble de la forme  $\{t \in \mathbb{R}^d : \|t\| \geq c\}$  avec  $c > 0$ .

— Pour tout  $t \neq 0$ , on a

$$\varphi(t) = \frac{1}{d} \sum_{j=1}^d \cos(t_j) = \frac{1}{d} \sum_{j=1}^d \left(1 - \frac{t_j^2}{2} + o(t_j^2)\right) = 1 - \frac{1}{2d} \|t\|^2 + o(\|t\|^2), \quad t \rightarrow 0,$$

et donc on a l'équivalence

$$F(1, t) \sim \frac{2d}{\|t\|^2}, \quad t \rightarrow 0.$$

— Comme  $\varphi(t) > 0$  dès que  $\|t\|$  est suffisamment petit on voit que  $z \mapsto 1/(1-z\varphi(t))$  est croissante pour tout  $t$  suffisamment petit.

(f) On fixe  $c > 0$  assez petit tel que  $z \mapsto F(z, t)$  est croissante et

$$\frac{d}{\|t\|^2} \leq F(1, t) \leq \frac{4d}{\|t\|^2} \quad (\star)$$

pour tout  $t$  tel que  $\|t\| \leq c$ . En utilisant (e), on voit que

$$\int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| > c} F(z, t) dt$$

reste borné pour tout  $z \in ]0, 1]$  et donc

$$\lim_{z \rightarrow 1} I(z) = \infty \quad \Leftrightarrow \quad \lim_{z \rightarrow 1} \int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} F(z, t) dt = \infty.$$

Par convergence monotone, on a

$$\lim_{z \rightarrow 1} \int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} F(z, t) dt = \int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} F(1, t) dt.$$

L'encadrement  $(\star)$  nous donne aussi que

$$\int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} F(1, t) dt = \infty \quad \Leftrightarrow \quad \int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} \frac{1}{\|t\|^2} dt.$$

Finalement, en passant en coordonnées polaires  $t \in \mathbb{R}^d \mapsto (r, \sigma)$  avec  $r = \|t\|$  et  $\sigma = t/\|t\|$  (de Jacobien  $dt = r^{d-1} dr d\sigma$  où  $d\sigma$  est la mesure uniforme sur la sphere de  $\mathbb{R}^d$ ) on voit que

$$\begin{aligned} & \int_{[-\pi, \pi]^d} \mathbf{1}_{\|t\| \leq c} \frac{1}{\|t\|^2} dt = \infty \\ & \Leftrightarrow \int_0^c r^{d-3} dr = \infty \\ & \Leftrightarrow d \leq 2. \end{aligned}$$