

---

# THEORIE DES PROBABILITES

---

Auguste AMAN

Professeur Titulaires, CAMES

UFR Mathématiques et Informatique

Université Félix H. Boigny, Cocody

Côte d'Ivoire

# Table des matières

<b>1 Espace de probabilité</b>	<b>1</b>
1.1 Rappel de théorie de la mesure . . . . .	2
1.1.1 Tribus . . . . .	2
1.1.2 Mesures . . . . .	3
1.2 Espace de probabilité . . . . .	4
1.3 Classe monotone . . . . .	8
1.4 Extension de mesure . . . . .	11
1.5 Vocabulaire probabiliste . . . . .	12
1.6 Liminf et limsup d'ensembles . . . . .	14
<b>2 Variables aléatoires</b>	<b>18</b>
2.1 Définitions et propriétés . . . . .	18
2.2 Loi d'une variable aléatoire . . . . .	21
2.3 Fonction de répartition . . . . .	22
2.4 Fonction quantile et simulation par inversion . . . . .	27
2.5 Exemples de variable aléatoire . . . . .	29
2.5.1 Variables aléatoires discrètes . . . . .	29
2.5.2 Variables aléatoires à densité . . . . .	35
2.5.3 Lois de probabilité usuelles . . . . .	43
<b>3 Espérance d'une variable aléatoire</b>	<b>48</b>
3.1 Rappels d'intégration . . . . .	48
3.1.1 Théorèmes de Fubini . . . . .	49
3.1.2 Changement de variable . . . . .	50
3.2 Espérance d'une variable aléatoire réelle . . . . .	53
3.3 Convergences monotone et dominée . . . . .	60
3.4 Moments des variables aléatoires . . . . .	62
3.5 Variance, covariance . . . . .	66
3.6 Tableau comparatif . . . . .	71
<b>4 Fonction caractéristique</b>	<b>72</b>
4.1 Définition et premières propriétés . . . . .	72
4.2 Propriétés et exemples . . . . .	75

4.3	Régularité de la fonction caractéristique . . . . .	81
4.4	Autres transformées caractéristiques . . . . .	82
4.4.1	Fonction génératrice . . . . .	82
4.4.2	Transformée de Laplace . . . . .	83
<b>5</b>	<b>Indépendance</b>	<b>85</b>
5.1	Concept d'indépendance . . . . .	85
5.2	Critères et exemples . . . . .	88
5.3	Non-corrélation et indépendance . . . . .	93
5.4	Évènements asymptotiques . . . . .	94
5.4.1	Tribus du futur et tribu asymptotique . . . . .	94
5.4.2	Lemmes de Borel-Cantelli . . . . .	96
<b>6</b>	<b>Somme de deux variables aléatoires indépendantes</b>	<b>99</b>
6.1	Convolution de mesures . . . . .	99
6.2	Loi d'une somme de variables aléatoires à densité indépendantes . . . . .	101
6.3	Variables aléatoires à densité indépendantes . . . . .	102
6.4	Cas de variables aléatoires discrètes indépendantes . . . . .	104
<b>7</b>	<b>Convergences de variables aléatoires</b>	<b>105</b>
7.1	Convergence presque sûre . . . . .	105
7.2	Convergence en probabilité . . . . .	107
7.3	Convergence en norme $p$ . . . . .	114
7.3.1	Définition et propriétés . . . . .	114
7.3.2	Uniforme intégrabilité . . . . .	115
7.4	Convergence en loi . . . . .	118
7.4.1	Convergence en loi et autres convergences . . . . .	127
7.4.2	Liens entre les différents modes de convergence . . . . .	129
<b>8</b>	<b>Théorèmes limite</b>	<b>130</b>
8.1	Lois des grands nombres (LGN) . . . . .	130
8.1.1	Version faible de la LGN . . . . .	130
8.1.2	Version forte (presque sûre) de la LGN . . . . .	136
8.1.3	Applications de la LGN . . . . .	142
8.2	Théorème central limite (TCL) . . . . .	144
8.2.1	TCL classique et applications . . . . .	144
<b>9</b>	<b>Vecteurs gaussiens</b>	<b>147</b>
9.1	Variables aléatoires gaussiennes . . . . .	147
9.2	Vecteurs gaussiens . . . . .	148
9.3	Applications . . . . .	155
9.3.1	TCL multidimensionnel . . . . .	155
9.3.2	Estimations : moyenne, variance empiriques . . . . .	157

*Table des matières*

iii

9.3.3 Décomposition de vecteurs gaussiens et test du $\chi^2$ . . . . .	159
---	-----

# Introduction

Ces notes sont un support (enrichi) d'un cours de probabilités de base. Elles sont redéposables de plusieurs sources d'inspiration, parmi elles : [Suq] et [Gra]. Ce cours ne nécessite que des notions de théorie de la mesure et d'intégrale de Lebesgue. Des références classiques pour compléter un cours de probabilités de ce niveau sont [Ouv], [FF] (en français) et [Chung], [Fel], [Dur], [Kal] (en anglais). (La référence [Kal] est complète mais plus difficile d'accès.) D'autres références en ligne sont [JCB-Leb], [Suq].

Le contenu de ces notes est le suivant :

Dans le Chapitre 1, on donne quelques rappels de théorie de la mesure. On définit un espace de probabilité, une mesure de probabilité et on en rappelle les principales propriétés.

La notion de variable aléatoire est définie dans le Chapitre 2. On y décrit la loi d'une variable aléatoire, sa fonction de répartition et on donne les exemples de lois classiques (discrètes et à densité).

Dans le Chapitre 3, on présente les notions d'espérance, de variance et plus généralement de moments de variables aléatoires.

La fonction caractéristique est un outil très utile qui caractérise la loi d'une variable aléatoire. Cet outil est introduit au Chapitre 4 où on en étudie les principales propriétés.

Le concept d'indépendance est fondamental en probabilités. Il est introduit au Chapitre 5 où on en donne aussi plusieurs caractérisations.

Dans le Chapitre 6, on étudie la somme de variables aléatoires indépendantes et on en détermine la loi à l'aide de convolution.

Il existe plusieurs modes de convergence en probabilités. Ils sont présentés dans le Chapitre 7 où leurs propriétés et relations sont étudiées.

Dans le Chapitre 8, on s'intéresse aux sommes de variables aléatoires indépendantes identiquement distribuées et à leur comportement limite. On y présente les deux premiers résultats fondamentaux des probabilités : la loi des grands nombres (LGN) et le théorème central limite (TCL).

On termine dans le Chapitre 9 avec la description de vecteurs aléatoires gaussiens pour lesquels beaucoup de calculs se ramènent à des calculs matriciels, on parle alors de calcul gaussien.

# Chapitre 1

## Espace de probabilité

### Introduction

Quelques jalons historiques formalisent le concept de probabilités. D'après l'[article dédié de wikipedia](#) :

- La notion de probabilité remonte à Aristote (4ème siècle avant J.-C.), il ne s'agit pas alors de quantifier l'*aléa*, le terme probabilité désigne plutôt l'adhésion à une idée : ce qui est probable est ce qui est généralement admis comme vrai.
- Au 16ème et 17ème siècles, la notion de probabilité est une notion morale. D'abord théologie morale catholique, le terme désignera par glissement sémantique le caractère vraisemblable d'une idée.
- Le traitement mathématique des probabilités remonte à Blaise Pascal (17ème siècle) notamment avec sa correspondance avec Pierre de Fermat (1654). Avec ce traitement mathématique, la notion de probabilité ne concerne plus seulement les idées ou les opinions mais aussi les faits. Le concept de probabilité se rapproche alors de la notion de hasard.
- Le calcul des probabilités se développe autour de questions liées à la théorie des jeux. Des contributions marquantes sont celles de Christian Huygens (espérance, 1657), Jacques Bernoulli (variable aléatoire, LGN, 1713 posthume), Abraham de Moivre (combinatoire, 1718), Pierre-Simon de Laplace (TCL, 1812).
- La théorie classique des probabilités se développe avec le début du 20ème siècle et le fondement de la théorie de la mesure par Émile Borel (1897), Henri Lebesgue (1902-1904).
- Andreï Markov (1902) introduit les chaînes de Markov pour généraliser la LGN dans un cadre sans indépendance.
- La théorie moderne des probabilités est axiomatisée par Andreï Kolmogorov (1933), considéré comme le père des probabilités modernes.
- Les probabilités se développent encore dans les années 1940 avec Paul Lévy et Kiyoshi Itô qui relient les probabilités avec l'analyse et les EDP. C'est l'apparition du calcul stochastique et de l'analyse stochastique.

Quelques références classiques pour compléter ce cours de Probabilités sont [Bil1], [Chung], [Fel], [FF], [Kal], [Ouv].

## 1.1 Rappel de théorie de la mesure

Dans cette section, on se borne à rappeler quelques concepts et outils fondamentaux de la théorie de la mesure nécessaires pour ce cours de Probabilités. On renvoie à [Rud] ou à un cours de théorie de la mesure complet pour des détails.

### 1.1.1 Tribus

Pour un ensemble  $X$ , on note  $\mathcal{P}(X) = \{A : A \subset X\}$  l'ensemble de ses parties.

**Définition 1.1.1 (Tribu)**  $\mathcal{A} \subset \mathcal{P}(X)$  est une tribu (ou une  $\sigma$ -algèbre) si

- $X \in \mathcal{A}$ ;
- Si pour tout  $i \in \mathbb{N}$ ,  $A_i \in \mathcal{A}$  alors  $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$  :  $\mathcal{A}$  est stable par réunion dénombrable;
- Si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$  :  $\mathcal{A}$  est stable par complémentaire.

Si la condition sur l'union est remplacée par la stabilité par union finie, on définit alors une algèbre (ou algèbre de Boole).

La plus petite tribu est  $\{\emptyset, X\}$  (tribu grossière), la plus grande est  $\mathcal{P}(X)$  (tribu totale). En général, les tribus intéressantes sont intermédiaires entre les deux.

On appelle (ensemble) mesurable tout ensemble  $A$  élément d'une tribu  $\mathcal{A}$ . Un ensemble muni d'une tribu  $(X, \mathcal{A})$  s'appelle un espace mesurable.

**Définition 1.1.2 (Tribu borélienne)** Lorsque  $X$  est un espace topologique (c'est à dire muni d'une famille d'ouverts), la plus petite tribu contenant tous les ouverts est appelée la tribu borélienne. Elle est notée  $\mathcal{B}(X)$ .

Les mesurables  $A \in \mathcal{B}(X)$  s'appellent aussi les boréliens. Les boréliens typiques sont les ouverts, les fermés.

En général, pour un ensemble  $X$  dénombrable, une bonne tribu à considérer est  $\mathcal{P}(X)$  et pour un espace  $X$  topologique, c'est la tribu borélienne  $\mathcal{B}(X)$ .

**Définition 1.1.3 (Mesurabilité)** Une application  $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  est dite mesurable si  $\forall B \in \mathcal{B}$ ,  $f^{-1}(B) \in \mathcal{A}$ .

### Exemples.

- La fonction  $\mathbf{1}_A$  est mesurable si et seulement si  $A \in \mathcal{A}$ .
- Lorsqu'on travaille avec les tribus boréliennes, les fonctions  $f : X \rightarrow Y$  continues sont mesurables (implicitement par rapport aux tribus boréliennes  $\mathcal{B}(X)$  de  $X$  et  $\mathcal{B}(Y)$  de  $Y$ ), cf. Proposition 2.1.1.

- Les applications mesurables sont stables par la plupart des opérations : addition, multiplication, multiplication par un scalaire, inverse, quotient, composition, max, min, sup, inf, parties positive et négative, passage à la limite simple, cf. Proposition 2.1.2.

**Définition 1.1.4 (Tribu engendrée)** — Soit  $\mathcal{M}$  une famille de parties de  $X$ . La tribu engendrée par  $\mathcal{M}$ , notée  $\sigma(\mathcal{M})$ , est la plus petite tribu de  $X$  contenant  $\mathcal{M}$  :  $\sigma(\mathcal{M}) = \bigcap_{\mathcal{A} \supset \mathcal{M}} \mathcal{A}$ .

- Soit  $f$  une application de  $X$  dans  $(Y, \mathcal{B})$ , espace mesurable. La tribu engendrée par  $f$  sur  $X$  est la plus petite tribu  $\mathcal{A}_f$  sur  $X$  rendant  $f : (X, \mathcal{A}_f) \rightarrow (Y, \mathcal{B})$  mesurable. C'est la tribu formée des  $\{f^{-1}(B) : B \in \mathcal{B}\}$ . On la note  $\mathcal{A}_f$  ou  $\sigma(f)$ .

**Définition 1.1.5 (Tribu produit)** Soient  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$  des espaces mesurables. La tribu produit  $\mathcal{A} \otimes \mathcal{B}$  sur l'espace produit  $X \times Y = \{(x, y) : x \in X, y \in Y\}$  est la tribu engendrée par les produits de mesurables  $A \times B$ ,  $A \in \mathcal{A}$ ,  $B \in \mathcal{B}$ .

Cf. Section 3.1 pour plus de rappels sur les espaces produits en liaison avec les Théorèmes de Fubini-Tonelli et Fubini (Théorèmes 3.1.1 et 3.1.2).

**Remarque 1.1.1** Sur  $\mathbb{R}^2 = \mathbb{R} \otimes \mathbb{R}$ , on peut considérer la tribu borélienne associée à la topologie produit sur  $\mathbb{R}^2$  et le produit des tribus boréliennes sur chaque espace  $\mathbb{R}$ . En utilisant que tout ouvert de  $\mathbb{R}^2$  peut s'écrire comme une réunion dénombrable de pavés d'intervalles ouverts, on montre que les deux coïncident :  $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R})^{\otimes 2}$ . Plus généralement, on montre pour  $\mathbb{R}^n$  qu'on a  $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R})^{\otimes n}$ .

## 1.1.2 Mesures

On considère  $(X, \mathcal{A})$  un espace mesurable.

**Définition 1.1.6 (Mesure)** Une mesure  $\mu$  sur  $(X, \mathcal{A})$  est une application de  $\mathcal{A} \rightarrow [0, +\infty]$  telle que

- $\mu(\emptyset) = 0$  ;
- si  $(A_n)_{n \geq 1}$  est une suite dénombrable d'ensembles de  $\mathcal{A}$  deux à deux disjoints alors

$$\mu \left( \bigcup_{n=1}^{+\infty} A_n \right) = \sum_{n=1}^{+\infty} \mu(A_n) \quad \sigma\text{-additivité.}$$

Le triplet  $(X, \mathcal{A}, \mu)$  est appelé un espace mesuré (espace mesurable + mesure).

### Exemples de mesure.

- Mesure de Dirac sur  $(X, \mathcal{P}(X))$  : soit  $a \in X$ ,

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

- Mesure de Lebesgue sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  : c'est la mesure qui généralise la notion de longueur des intervalles. Elle est invariante par translation :

$$\lambda([a, b]) = b - a, \quad \lambda(A + x) = \lambda(A).$$

- Mesure finie : mesure  $\mu$  de poids total (ou masse) fini(e) :  $\mu(X) < +\infty$ .
- Mesure  $\sigma$ -finie :  $\mu$  est  $\sigma$ -finie sur  $(X, \mathcal{A})$  si  $X = \bigcup_{n \geq 1} A_n$  avec  $A_n \in \mathcal{A}$  et  $\mu(A_n) < +\infty$  (exemple : la mesure de Lebesgue sur  $\mathbb{R} = \bigcup_{n \in \mathbb{N}} [-n, n]$ ) car  $\lambda([-n, n]) = 2n$ .
- Mesure de probabilité : une mesure  $\mu$  est dite de probabilité sur  $(X, \mathcal{A})$  si  $\mu(X) = 1$ . Traditionnellement, pour un espace de probabilité, on note  $(\Omega, \mathcal{F}, \mathbb{P})$  au lieu de  $(X, \mathcal{A}, \mu)$ , cf. ci-dessous.
- Mesure image : Soit  $f : (X, \mathcal{A}, \mu) \rightarrow (Y, \mathcal{B})$  une fonction mesurable. On définit sur  $(Y, \mathcal{B})$  la mesure image de  $f$  notée  $\mu_f$  par :

$$\mu_f(B) = \mu(f^{-1}(B)).$$

- Mesure produit : si  $\mu$  et  $\nu$  sont des mesures sur  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$ , on considère la mesure produit sur l'espace produit  $X \times Y$  muni de la tribu produit  $\mathcal{A} \otimes \mathcal{B}$ . C'est la mesure  $\mu \otimes \nu$ , défini sur les produits de mesurables par

$$(\mu \otimes \nu)(A \times B) = \mu(A) \times \nu(B), \quad A \in \mathcal{A}, B \in \mathcal{B}$$

et s'étend sur toute la tribu produit  $\mathcal{A} \otimes \mathcal{B}$  par un argument classique (par exemple de classe monotone, cf. Section 1.3 ou par le théorème de Carathéodory).

## 1.2 Espace de probabilité

On définit une probabilité  $\mathbb{P}$  sur un espace mesurable qu'on note traditionnellement  $(\Omega, \mathcal{F})$ . Dans ce contexte, les ensembles mesurables  $A \in \mathcal{F}$  s'appellent les évènements.

**Définition 1.2.1 (Probabilité)** Soient  $(\Omega, \mathcal{F})$  un espace mesurable. On appelle probabilité sur  $(\Omega, \mathcal{F})$  toute mesure  $\mathbb{P}$  sur  $(\Omega, \mathcal{F})$  de poids total 1, c'est à dire  $\mathbb{P}$  est une application de  $\mathcal{F}$  dans  $[0, 1]$  qui vérifie :

- (i)  $\mathbb{P}(\Omega) = 1$  ;
- (ii) (Propriété de  $\sigma$ -additivité) Pour toute suite  $(A_n)_{n \in \mathbb{N}^*}$  d'évènements, deux à deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Le triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  s'appelle un espace probabilisé ou **espace de probabilité**.

- Remarque 1.2.1**
- Lorsque l'espace  $\Omega$  est discret (c'est à dire fini ou dénombrable, par exemple  $\mathbb{N}$  ou une partie de  $\mathbb{N}$ ), on utilise  $\mathcal{F} = \mathcal{P}(\Omega)$  et tous les ensembles sont évènements. C'est la raison pour laquelle cette restriction aux évènements n'apparaît pas lors de cours de Probabilités en espaces finis ou discrets.
  - Lorsque l'espace est  $\mathbb{R}$ , pour les évènements typiques sont les intervalles (fermés ou ouverts ou mixtes).

### Exemples.

- $\mathbb{P} = \sum_{n \geq 1} \frac{6}{\pi^2 n^2} \delta_n$ .
- Soit  $f(x) = \frac{1}{2}e^{-|x|}$ , montrer que

$$\mathbb{P}([a, b]) = \int_a^b f(x) dx$$

définit une mesure de probabilité sur  $\mathcal{B}(\mathbb{R})$ .

Finalement, par évènement, on pourra se contenter de comprendre en pratique (en première approximation) : n'importe quel ensemble si l'espace  $\Omega$  est discret et les intervalles si l'espace est  $\mathbb{R}$ .

### Propriétés des probabilités

Une probabilité satisfait un certain nombre de propriétés de base qu'il faut connaître et pouvoir manipuler facilement.

Toute probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{F})$  vérifie les propriétés suivantes :

- $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .

En effet  $\Omega = A \cup A^c$  avec une réunion disjointe. Par additivité, on a donc

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

- $\mathbb{P}(\emptyset) = 0$ .

En effet  $\emptyset = \Omega^c$  donc  $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$ .

- Additivité (cas particulier du point (ii) de la définition d'une probabilité) :

- Si  $A \cap B = \emptyset, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ ,
- Si les  $A_i$  ( $1 \leq i \leq n$ ) sont deux à deux disjoints,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i).$$

- $\forall A, B \in \mathcal{F}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ .

En effet  $B = (B \setminus A) \cup A$  où la réunion est disjointe. On a donc

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

- $\forall A, B \in \mathcal{F}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

En effet  $A \cup B = (B \setminus A) \cup (A \cap B) \cup (A \setminus B)$  où les ensembles sont deux à deux disjoints. On a donc

$$\mathbb{P}(A \cup B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B). \quad (1.1)$$

Or  $A = (A \setminus B) \cup (A \cap B)$  avec une réunion d'ensembles disjoints donc

$$\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B).$$

Et de même  $B = (B \setminus A) \cup (A \cap B)$  avec une réunion d'ensembles disjoints donc

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

On a donc  $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$  et  $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$  et on conclut en reportant dans (1.1).

- Sous-additivité :

- $\forall A, B \in \mathcal{F} : \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$  ;
- $\forall A_1, A_2, \dots, A_n \in \mathcal{F} : \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n)$  ;
- $\forall A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$  :

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

En effet, pour une réunion de deux ensembles  $A \cup B$ , cela vient du point précédent ; puis, pour le cas d'une réunion finie, d'une récurrence ; et, pour le cas d'une réunion dénombrable, d'un passage à la limite avec la propriété suivante.

- Propriétés de **continuité monotone séquentielle**

(i) Si  $(A_n)_{n \in \mathbb{N}^*}$  est une suite croissante d'évènements (ie. pour tout  $n$   $A_n \subset A_{n+1}$ ) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}(A) \text{ où } A = \bigcup_{n \in \mathbb{N}^*} A_n. \quad (1.2)$$

(ii) Si  $(B_n)_{n \in \mathbb{N}^*}$  est une suite décroissante d'évènements (ie. pour tout  $n$   $B_{n+1} \subset B_n$ ) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}(B) \text{ où } B = \bigcap_{n \in \mathbb{N}^*} B_n. \quad (1.3)$$

En effet dans le cas croissant (i), on note que  $\bigcup_{k=1}^n A_k = A_n$ . Soit  $C_n = A_n \setminus A_{n-1}$   $n \geq 1$  (avec  $A_0 = \emptyset$ ). On montre facilement que  $\bigcup_{k=1}^n C_k = \bigcup_{k=1}^n A_n$  : une inclusion vient de ce que  $C_k \subset A_k$  pour tout  $k$ , l'autre de ce que si  $\omega \in \bigcup_{k=1}^n A_n$  alors en notant  $k$  le plus petit entier tel que  $\omega \in A_k$  alors on a  $\omega \in C_k$  et donc  $\omega \in \bigcup_{k=1}^n C_k$ . De plus les  $C_k$ ,  $k \geq 1$ , sont deux à deux disjoints : si  $\omega \in C_k$  alors  $\omega \notin A_{k-1}$  et donc  $\omega \notin C_l$  pour  $l < k$ .

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=1}^n C_k\right) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mathbb{P}(C_k) \quad (\text{additivité})$$

$$= \sum_{k=1}^{+\infty} \mathbb{P}(C_k) = \mathbb{P}\left(\bigcup_{k=1}^{+\infty} C_k\right) \quad (\sigma\text{-additivité}) \quad = \mathbb{P}\left(\bigcup_{k=1}^{+\infty} A_k\right).$$

Dans le cas décroissant (ii), comme on est en mesure finie, il suffit de passer aux complémentaires.

**Remarque 1.2.2** En général, on ne peut pas calculer  $\mathbb{P}(A \cup B)$  à partir de  $\mathbb{P}(A)$  et de  $\mathbb{P}(B)$  comme le montre la formule  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$  : il faut connaître  $A \cap B$ , on verra que ceci est lié à l'indépendance ou non des événements  $A$  et  $B$ .

Attention, cette formule ne se généralise pas immédiatement pour plus de deux événements, par exemple pour  $A, B, C$ , on a :

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Plus généralement, on a le résultat suivant (exercice : il se montre par récurrence sur  $n$ ) :

**Proposition 1.2.1 (Formule de Poincaré)** Pour tout entier  $n \geq 2$ , et tous événements  $A_1, A_2, \dots, A_n$ , on a :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}).$$

**Démonstration :** On prouve la formule par récurrence. Si  $n = 2$ , cela est dû à l'additivité usuelle. On admet la formule pour  $n$  et on la montre pour  $n + 1$  : soient  $A_i$ ,  $1 \leq i \leq n + 1$ . On a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cup A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^{n+1} (A_i \cap A_{n+1})\right). \end{aligned}$$

Mais par récurrence

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$$

et

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} (A_i \cap A_{n+1})\right) = \sum_{i=1}^n \mathbb{P}(A_i \cap A_{n+1}) + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap A_{n+1}).$$

On reforme alors  $\sum_{i=1}^n \mathbb{P}(A_i) + \mathbb{P}(A_{n+1}) = \sum_{i=1}^{n+1} \mathbb{P}(A_i)$  et

$$\begin{aligned} & \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) + \sum_{i=1}^n \mathbb{P}(A_i \cap A_{n+1}) \\ & + \sum_{k=2}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap A_{n+1}) \\ = & \sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n+1} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) \end{aligned}$$

ce qui prouve la formule pour  $n+1$  et achève la récurrence.  $\square$

### 1.3 Classe monotone

Dans cette section on rappelle un procédé d'extension des définitions de certains objets sur les tribus après les avoir définis sur des classes restreintes d'ensemble.

**Définition 1.3.1 (Classe monotone ou  $\lambda$ -système)** Une famille  $\mathcal{M}$  de parties de  $\Omega$  est appelée classe monotone si

1.  $\Omega \in \mathcal{M}$  ;
2. lorsque  $A, B \in \mathcal{M}$  et  $B \subset A$ , alors  $A \setminus B \in \mathcal{M}$  ;
3.  $\mathcal{M}$  est stable par réunion croissante ( $A_j \in \mathcal{M}$ ,  $j \in \mathbb{N}$ ,  $A_j \subset A_{j+1} \Rightarrow \bigcup_{j \in \mathbb{N}} A_j \in \mathcal{M}$ ).

Pour  $\mathcal{E} \subset \mathcal{P}(\Omega)$ , on appelle classe monotone engendrée par  $\mathcal{E}$ , la plus petite classe monotone contenant  $\mathcal{E}$ , c'est à dire l'intersection de toutes les classes monotones contenant  $\mathcal{E}$ . On la note  $\mathcal{M}(\mathcal{E})$ .

**Remarque 1.3.1** — Une classe monotone est stable par complémentaire : il suffit d'écrire  $A^c = \Omega \setminus A$  pour  $\Omega, A \in \mathcal{M}$ .

— Une classe monotone est stable par intersection décroissante : si  $(B_i)_{i \geq 1}$  est une suite de  $\mathcal{M}$  telle que  $B_i \supset B_{i+1}$ ,  $i \geq 1$ , alors  $A_i = B_i^c \in \mathcal{M}$  car  $A_i = \Omega \setminus B_i$  et  $(A_i)_{i \geq 1}$  forme une suite croissante de  $\mathcal{M}$  pour laquelle  $\bigcup_{i \geq 1} A_i \in \mathcal{M}$  mais alors  $\bigcap_{i \geq 1} B_i = \left( \bigcup_{i \geq 1} A_i \right)^c \in \mathcal{M}$ .

Une classe monotone est donc stable par limite monotone d'ensembles (croissante ou décroissante), cela justifie la terminologie.

**Exemple :**

1. Une intersection d'un nombre quelconque de classes monotones est encore une classe monotone.
2. Une tribu est une classe monotone. Il suffit pour cela de voir que  $A \setminus B = A \cap B^c$ .

3. Une classe monotone stable par intersection finie est une tribu.

En effet cette classe sera aussi stable par réunion finie en vertu de l'axiome 2) de la Définition 1.3.1 on utilise alors la reécriture d'une réunion dénombrable comme une réunion croissante ( $\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} \left( \bigcup_{k \leq j} A_k \right)$  pour toute famille  $A_j, j \in \mathbb{N}$ ).

**Théorème 1.3.1 (des classes monotones)** *Soit  $\mathcal{E}$  une famille de parties de  $\Omega$  stable par intersection finie. Alors  $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$ .*

**Remarque 1.3.2** Ce résultat s'énonce (et s'utilise) encore sous la forme suivante : si  $\mathcal{M}$  est une classe monotone contenant la famille de parties  $\mathcal{E}$ , stable par intersection finie (ie.  $\mathcal{E}$  est un  $\pi$ -système), alors  $\sigma(\mathcal{E}) \subset \mathcal{M}$ .

**Démonstration :** En vertu de l'exemple 2) ci-dessus,  $\sigma(\mathcal{E})$  est une classe monotone qui contient  $\mathcal{E}$  et donc  $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$ . Pour prouver l'inclusion réciproque, on montre que  $\mathcal{M}(\mathcal{E})$  est stable par intersection finie car alors, d'après l'exemple 3) ci-dessus,  $\mathcal{M}(\mathcal{E})$  sera une tribu contenant  $\mathcal{E}$ , et on aura  $\sigma(\mathcal{E}) \subset \mathcal{M}(\mathcal{E})$ . Il suffit donc de prouver que si  $A, B \in \mathcal{M}(\mathcal{E})$ , alors  $A \cap B \in \mathcal{M}(\mathcal{E})$ . Pour cela, soit

$$\mathcal{M}_1 := \{A \in \mathcal{M}(\mathcal{E}) : \forall B \in \mathcal{E}, A \cap B \in \mathcal{M}(\mathcal{E})\}.$$

Comme  $\mathcal{E}$ ,  $\pi$ -système, est stable par intersection finie, on constate que  $\mathcal{M}_1$  contient  $\mathcal{E}$ . Puis on vérifie facilement que  $\mathcal{M}_1$  est une classe monotone car

- $\Omega \in \mathcal{M}_1$  car  $\Omega \in \mathcal{M}(\mathcal{E})$  et pour  $B \in \mathcal{E}$ ,  $B \cap \Omega = B \in \mathcal{E} \subset \mathcal{M}(\mathcal{E})$ .
- si  $A_1, A_2 \in \mathcal{M}_1$  avec  $A_2 \subset A_1$  alors  $A_1 \setminus A_2 \in \mathcal{M}(\mathcal{E})$  car  $\mathcal{M}(\mathcal{E})$  est une classe monotone puis pour  $B \in \mathcal{E}$ , on a  $(A_1 \setminus A_2) \cap B = (A_1 \cap B) \setminus (A_2 \cap B)$ ; mais  $A_1 \cap B, A_2 \cap B \in \mathcal{M}(\mathcal{E})$  car  $A_1, A_2 \in \mathcal{M}_1$ ; puis comme  $\mathcal{M}(\mathcal{E})$  est stable par différence monotone, on a  $(A_1 \setminus A_2) \cap B \in \mathcal{M}(\mathcal{E})$ ; finalement  $A_1 \setminus A_2 \in \mathcal{M}_1$ .
- si  $A_j, j \geq 0$ , est dans  $\mathcal{M}_1$  avec  $A_j \subset A_{j+1}$  alors  $\bigcup_j A_j \in \mathcal{M}(\mathcal{E})$  car  $\mathcal{M}(\mathcal{E})$  est stable par réunion croissante; puis, pour  $B \in \mathcal{E}$ ,  $\left( \bigcup_j A_j \right) \cap B = \bigcup_j (A_j \cap B) \in \mathcal{M}(\mathcal{E})$  car  $A_j \cap B \in \mathcal{M}(\mathcal{E})$  et stabilité par réunion monotone.

Finalement,  $\mathcal{M}_1$  est une classe monotone contenant  $\mathcal{E}$  donc contient aussi  $\mathcal{M}(\mathcal{E})$  et, par définition est contenu dans  $\mathcal{M}(\mathcal{E})$ , ce qui donne  $\mathcal{M}_1 = \mathcal{M}(\mathcal{E})$ .

Soit maintenant

$$\mathcal{M}_2 := \{A \in \mathcal{M}(\mathcal{E}) : \forall B \in \mathcal{M}(\mathcal{E}), A \cap B \in \mathcal{M}(\mathcal{E})\}.$$

L'ensemble  $\mathcal{M}_2$  est aussi une classe monotone (faire comme précédemment avec  $\mathcal{M}_1$  à la place de  $\mathcal{E}$ ). De plus il contient  $\mathcal{E}$  : on doit pour cela montrer que si  $A \in \mathcal{E}$ , alors pour tout  $B \in \mathcal{M}(\mathcal{E})$  on a  $A \cap B \in \mathcal{M}(\mathcal{E})$ . Mais comme  $B \in \mathcal{M}(\mathcal{E}) = \mathcal{M}_1$ , et  $A \in \mathcal{E}$ , on a  $A \cap B = B \cap A \in \mathcal{M}(\mathcal{E})$  (définition de  $\mathcal{M}_1$ ). On a donc  $\mathcal{M}_2$  classe monotone contenant  $\mathcal{E}$  et donc  $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}_2$ . Comme par définition on a aussi  $\mathcal{M}_2 \subset \mathcal{M}(\mathcal{E})$ , il vient  $\mathcal{M}_2 = \mathcal{M}(\mathcal{E})$  ce qui montre que  $\mathcal{M}(\mathcal{E})$  est stable par intersection finie.

D'après l'argument qui commence la preuve, le théorème des classes monotones (Th. 1.3.1) est prouvé.  $\square$

## Applications des classes monotones

**Théorème 1.3.2 (Dynkin)** *Si deux probabilités  $\mathbb{P}_1$  et  $\mathbb{P}_2$  coïncident sur  $\mathcal{A}$  stable par intersections finies alors  $\mathbb{P}_1$  et  $\mathbb{P}_2$  coïncident sur  $\sigma(\mathcal{A})$ .*

**Démonstration :** Soit  $\mathcal{M} = \{A \in \mathcal{F} : \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$ . Alors, à nouveau, on constate facilement que  $\mathcal{M}$  est une classe monotone qui contient  $\mathcal{A}$  :

- $\Omega \in \mathcal{M}$  car  $\mathbb{P}_1(\Omega) = 1 = \mathbb{P}_2(\Omega)$ .
- Si  $A, B \in \mathcal{M}$  avec  $B \subset A$  alors

$$\mathbb{P}_1(A \setminus B) = \mathbb{P}_1(A) - \mathbb{P}_1(B) = \mathbb{P}_2(A) - \mathbb{P}_2(B) = \mathbb{P}_2(A \setminus B).$$

- Si  $A_j \in \mathcal{M}$  avec  $A_j \subset A_{j+1}$ ,  $j \geq 1$ , alors par croissance séquentielle (1.2)

$$\mathbb{P}_1\left(\bigcup_{j \geq 1} A_j\right) = \lim_{j \rightarrow +\infty} \mathbb{P}_1(A_j) = \lim_{j \rightarrow +\infty} \mathbb{P}_2(A_j) = \mathbb{P}_2\left(\bigcup_{j \geq 1} A_j\right).$$

Comme  $\mathcal{A}$  est un  $\pi$ -système, le Théorème 1.3.1 garantit que  $\sigma(\mathcal{A}) \subset \mathcal{M}$  ce qui conclut.  $\square$

En anticipant sur le Chapitre 5, on dit que deux évènements  $A, B$  sont indépendants (et on écrit  $A \perp\!\!\!\perp B$ ) si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Plus généralement deux parties  $\mathcal{A}$  et  $\mathcal{B}$  d'évènements sont indépendantes si pour tout  $A \in \mathcal{A}$  et  $B \in \mathcal{B}$  alors  $A \perp\!\!\!\perp B$ .

**Proposition 1.3.1** *Si une famille  $\mathcal{A}$  d'ensemble est stable par intersections finies et est indépendante d'une tribu  $\mathcal{G}$  alors  $\sigma(\mathcal{A})$  est indépendante de  $\mathcal{G}$ .*

**Démonstration :** Soit  $\mathcal{M} = \{A \in \mathcal{F} : A \perp\!\!\!\perp \mathcal{G}\}$ . Alors on constate facilement que  $\mathcal{M}$  est une classe monotone qui contient  $\mathcal{A}$ .

- On a  $\Omega \in \mathcal{M}$  car  $\Omega \perp\!\!\!\perp \mathcal{G}$  :  $\mathbb{P}(G \cap \Omega) = \mathbb{P}(G) = \mathbb{P}(G)\mathbb{P}(\Omega)$  pour tout  $G \in \mathcal{G}$ .
- Soit  $A, B \in \mathcal{M}$  avec  $B \subset A$ , alors pour tout  $G \in \mathcal{G}$ , on a

$$\begin{aligned} \mathbb{P}((A \setminus B) \cap G) &= \mathbb{P}((A \cap G) \setminus (B \cap G)) \\ &= \mathbb{P}(A \cap G) - \mathbb{P}(B \cap G) = \mathbb{P}(A)\mathbb{P}(G) - \mathbb{P}(B)\mathbb{P}(G) \\ &= (\mathbb{P}(A) - \mathbb{P}(B))\mathbb{P}(G) = \mathbb{P}(A \setminus B)\mathbb{P}(G) \end{aligned}$$

c'est à dire  $(A \setminus B) \perp\!\!\!\perp G$  et donc  $(A \setminus B) \perp\!\!\!\perp \mathcal{G}$ .

- Soit  $A_j \in \mathcal{G}$  avec  $A_j \subset A_{j+1}$  alors par convergence monotone

$$\begin{aligned} \mathbb{P}(\bigcup_{j \geq 1} A_j \cap G) &= \mathbb{P}(\bigcup_{j \geq 1} (A_j \cap G)) = \lim_{j \rightarrow +\infty} \mathbb{P}(A_j \cap G) = \lim_{j \rightarrow +\infty} \mathbb{P}(A_j)\mathbb{P}(G) \\ &= \mathbb{P}(\bigcup_{j \geq 1} A_j)\mathbb{P}(G) \end{aligned}$$

c'est à dire  $\bigcup_{j \geq 1} A_j \perp\!\!\!\perp G$  et donc  $\bigcup_{j \geq 1} A_j \perp\!\!\!\perp \mathcal{G}$ .

Comme  $\mathcal{A}$  est un  $\pi$ -système, le Théorème 1.3.1 garantit que  $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A}) \subset \mathcal{M}$  ce qui conclut la preuve de la proposition.  $\square$

**Proposition 1.3.2** Soient  $\mathcal{A}_1$  et  $\mathcal{A}_2$  deux familles d'ensemble stables par intersection finie (deux algèbres) indépendantes dans  $(\Omega, \mathcal{F}, \mathbb{P})$ . Alors les tribus engendrées  $\sigma(\mathcal{A}_1)$  et  $\sigma(\mathcal{A}_2)$  sont indépendantes.

**Démonstration :** Soit  $A_1 \in \mathcal{A}_1$ . Considérons la famille

$$\mathcal{M}_1 = \{A_2 \in \sigma(\mathcal{A}_2) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}$$

des événements indépendants de  $A_1$ . Comme pour la preuve précédente, il s'agit d'une classe monotone qui contient, par hypothèse,  $\mathcal{A}_2$ . Elle contient donc la classe monotone engendrée par  $\mathcal{A}_2$ , qui coïncide par le théorème de classe monotone avec  $\sigma(\mathcal{A}_2)$  puisque  $\mathcal{A}_2$  est un  $\pi$ -système. Soit maintenant  $A_2 \in \sigma(\mathcal{A}_2)$ . Considérons la famille

$$\mathcal{M}_2 = \{A_1 \in \sigma(\mathcal{A}_1) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}$$

des événements indépendants de  $A_2$ . Il s'agit encore d'une classe monotone qui contient  $\mathcal{A}_1$  et d'après ce qui précède  $\sigma(\mathcal{A}_1)$  puisque  $\mathcal{A}_1$  est un  $\pi$ -système.  $\square$

## 1.4 Extension de mesure

Le Théorème de Carathéodory permet de prolonger des mesures (ou des ersatz de mesure) sur des familles d'ensemble en de vraies mesures sur des tribus engendrées par ces familles. On en donne plusieurs formulations dans cette section.

**Définition 1.4.1 (Anneau d'ensemble)** Un anneau d'ensemble  $\mathcal{R}$  sur un ensemble  $X$  est une famille de parties de  $X$  qui vérifie

- $\mathcal{R}$  n'est pas vide :  $\mathcal{R} \neq \emptyset$  ;
- $\mathcal{R}$  est stable par différence ensembliste : si  $A, B \in \mathcal{R}$  alors  $A \setminus B \in \mathcal{R}$  ;
- $\mathcal{R}$  est stable par union finie : si pour  $1 \leq i \leq n$  on a  $A_i \in \mathcal{R}$  alors  $\bigcup_{i=1}^n A_i \in \mathcal{R}$ .

**Théorème 1.4.1 (Extension de Carathéodory 1)** Toute mesure sur un anneau d'ensemble  $\mathcal{R}$  admet au moins un prolongement à la tribu  $\sigma(\mathcal{R})$  engendrée par l'anneau. De plus, si la mesure sur l'anneau est  $\sigma$ -finie alors ce prolongement est unique.

**Définition 1.4.2 (Semi-algèbre)** Soit  $\Omega \neq \emptyset$  un ensemble quelconque. Une famille de sous-ensembles  $\mathcal{S}$  de  $\Omega$  est une semi-algèbre si

- $\emptyset, \Omega \in \mathcal{S}$  ;
- $\mathcal{S}$  est stable par intersections finies ;
- si  $A \in \mathcal{S}$  alors il existe  $n$  fini et des ensembles  $C_1, \dots, C_n \in \mathcal{S}$  disjoints tels que  $A^c = C_1 \cup \dots \cup C_n$ .

Les semi-algèbres sont des structures à partir desquelles on peut étendre une probabilité.

**Théorème 1.4.2 (Extension de Carathéodory 2)** Soit  $\mathcal{S}$  une semi-algèbre sur  $\Omega \neq \emptyset$ . On considère  $\mathbb{P}$  une fonction  $\sigma$ -additive définie sur  $\mathcal{S}$  et à valeurs dans  $[0, 1]$  telle que  $\mathbb{P}(\Omega) = 1$ . Alors il existe une unique mesure de probabilité  $\widetilde{\mathbb{P}}$  définie sur  $\sigma(\mathcal{S})$  qui prolonge  $\mathbb{P}$  (ie.  $\mathbb{P}(A) = \widetilde{\mathbb{P}}(A)$  pour  $A \in \mathcal{S}$ ).

**Définition 1.4.3 (Mesure d'algèbre)** Soit  $\mathcal{B}$  une algèbre sur un ensemble  $E$ . Une mesure d'algèbre sur  $(E, \mathcal{B})$  est une application  $m : \mathcal{B} \rightarrow [0, +\infty]$  telle que

- $m(\emptyset) = 0$  ;
- $m$  est additive : pour  $A, B \in \mathcal{B}$  tels que  $A \cap B = \emptyset$  alors  $m(A \cup B) = m(A) + m(B)$  ;
- Il existe une suite croissante  $(E_n)_{n \geq 1}$  de  $\mathcal{B}$  qui croît vers  $E$  telle que  $m(E_n) < +\infty$  et, pour tout  $A \in \mathcal{B}$ ,  $\lim_{n \rightarrow +\infty} m(A \cap E_n) = m(A)$  ;
- (propriété de Carathéodory) Pour toute suite décroissante  $(A_n)_n$  de  $\mathcal{B}$  convergeant vers  $\emptyset$  et telle que  $m(A_0) < +\infty$  on a  $\lim_{n \rightarrow +\infty} m(A_n) = 0$ .

On montre facilement qu'une mesure d'algèbre  $m$  sur  $(E, \mathcal{B})$  vérifie pour tout  $A, B \in \mathcal{B}$  :

- additivité finie :  $m(A) = m(A \setminus B) + m(A \cap B)$  ;
- additivité forte :  $m(A \cup B) + m(A \cap B) = m(A) + m(B)$  ;
- sous-additivité :  $m(A \cup B) \leq m(A) + m(B)$  ;
- croissance : si  $A \subset B$  alors  $m(A) \leq m(B)$ .

**Théorème 1.4.3 (Extension de Carathéodory 3)** Soit  $\mathcal{B}$  une algèbre sur un ensemble  $E$ . Si  $m$  est une mesure d'algèbre sur  $(E, \mathcal{B})$  alors il existe une mesure  $\mu$  sur la tribu  $\sigma(\mathcal{B})$  qui coïncide avec  $m$  sur  $\mathcal{B}$ .

## 1.5 Vocabulaire probabiliste

D'un point de vue modélisation, dans la suite, l'ensemble de base  $\Omega$  va nous permettre de décrire une expérience aléatoire. Cet ensemble va représenter l'ensemble des résultats possibles de l'expérience (aléatoire) que l'on étudie. Nous l'appellerons *l'univers des possibles* ou *espace probabilisé*. Lorsque  $\Omega$  est dénombrable, ses parties seront appelées des événements (ou évènements composés), les éléments  $\omega \in \Omega$  seront les évènements élémentaires, c'est à dire les évènements les plus simples qui ne peuvent pas être exprimés par des évènements encore plus simples.

**Exemple :** On lance un dé à six face. Le résultat *a priori* est aléatoire et les résultats possibles sont 1, 2, 3, 4, 5, 6. L'espace  $\Omega = \{1, 2, 3, 4, 5, 6\}$  décrit bien l'ensemble des résultats. La partie  $A = \{1, 4\}$  est un évènement composé : il s'agit de « le résultat est un 1 ou un 4 ». Par contre  $\{3\}$  est un évènement élémentaire, « observer un 3 » ne peut pas être décrit par des évènements plus simples.

Avec ce mode de représentation, les opérations logiques sur les évènements que sont « ou », « et », « négation » se traduisent par des opérations ensemblistes : réunion  $\cup$ , intersection  $\cap$ , complémentaire  $\{ \}^c$ . Voici le tableau des correspondances entre ces deux langages :

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
$\emptyset$	ensemble vide	événement impossible
$\Omega$	ensemble plein	événement certain
$\omega$	élément de $\Omega$	événement élémentaire
$A$	sous-ensemble de $\Omega$	événement
$\omega \in A$	$\omega$ appartient à $A$	le résultat $\omega$ est une des réalisations possibles de $A$
$A \subset B$	$A$ inclus dans $B$	$A$ implique $B$
$A \cup B$	réunion de $A$ et $B$	$A$ ou $B$
$A \cap B$	intersection de $A$ et $B$	$A$ et $B$
$A^c$	complémentaire de $A$ dans $\Omega$	événement contraire de $A$
$A \cap B = \emptyset$	$A$ et $B$ sont disjoints	$A$ et $B$ sont incompatibles

Pour un  $\Omega$  général, les événements sont seulement les éléments de  $\mathcal{F}$ , la tribu associée.

**Remarque 1.5.1** Il faut retenir que

- une réunion  $\cup$  s'interprète comme un « ou »,
- une intersection  $\cap$  s'interprète comme un « et »,
- un complémentaire  $\{ \}^c$  s'interprète comme « le contraire de ».

Noter enfin qu'en mathématiques le « ou » est un ou inclusif alors que dans le langage usuel il s'agit d'un ou exclusif (dessert ou fromage ? c'est l'un ou l'autre mais pas les deux alors qu'avec le « ou » mathématiques, ça pourrait être les deux).

Les opérations sur les ensembles (ou sur les événements) peuvent faire intervenir plus de deux événements. Ainsi si  $A_1, \dots, A_n$  sont des événements,

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$$

est l'ensemble des  $\omega$  qui sont dans au moins un des  $A_i$ . De même

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$$

est l'ensemble des  $\omega$  qui sont dans tous les  $A_i$ . On étend encore ces définitions aux réunions et intersections dénombrables (*i.e.* en nombre infini mais qu'on peut énumérer) :

$$\bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i=1}^{+\infty} A_i = \{\text{réalisation d'au moins un } A_i\},$$

$$\bigcap_{i \in \mathbb{N}} A_i = \bigcap_{i=1}^{+\infty} A_i = \{\text{réalisation de tous les } A_i\}.$$

**Rappel (dénombrabilité) :** une partie infinie est dénombrable si elle peut être mise en bijection avec  $\mathbb{N}$ , c'est à dire si on peut énumérer tous ses éléments. L'ensemble  $\mathbb{N}$ , bien sûr, est dénombrable mais  $\mathbb{Z}$ ,  $\mathbb{Q}$  le sont aussi. Par contre  $[0, 1]$  ou  $\mathbb{R}$  ne le sont pas.

Comme on peut énumérer aussi les éléments d'une partie finie, il est usage d'inclure le cas fini dans le cas dénombrable, même si d'ordinaire, le terme dénombrable est utilisé pour les parties infinies dénombrables.

Ces opérations logiques sur des suites d'événements sont très utiles pour analyser les événements complexes : il s'agit de les réexprimer comme réunion, intersection, complémentaire d'événements plus simples. Il importe donc de bien traduire en langage ensembliste un énoncé et ses enchaînements logiques.

Étant donnée une suite d'événements  $(A_i)_{i \geq 1}$ , deux événements assez complexes mais fort utiles sont

$$\begin{aligned} \text{la limite supérieure : } & \limsup_{i \rightarrow +\infty} A_i = \bigcap_{i \geq 0} \bigcup_{j > i} A_j \\ \text{et la limite inférieure : } & \liminf_{i \rightarrow +\infty} A_i = \bigcup_{i \geq 0} \bigcap_{j > i} A_j. \end{aligned}$$

Leur intérêt vient de l'interprétation suivante qui permet de « traduire » en langage ensembliste une assertion en français.

## 1.6 Liminf et limsup d'ensembles

'Etant donnée une suite d'événements  $(A_i)_{i \geq 1}$ , deux événements assez complexes mais fort utiles sont la limite inférieure et la limite supérieure. Leur intérêt vient de l'interprétation qui permet de « traduire » en langage ensembliste une assertion en français, cf. Proposition 1.6.1.

**Définition 1.6.1 (lim inf et lim sup)** Soit  $(A_n)_{n \geq 0}$  une suite d'événements observables d'un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . On pose

$$\limsup_{n \rightarrow +\infty} A_n = \bigcap_{n \geq 1} \bigcup_{k > n} A_k, \quad \text{et} \quad \liminf_{n \rightarrow +\infty} A_n = \bigcup_{n \geq 1} \bigcap_{k > n} A_k.$$

On parle respectivement de limites supérieure et inférieure de la suite d'ensembles  $(A_n)_{n \geq 0}$ .

Noter que

$$\liminf_{n \rightarrow +\infty} A_n \subset \limsup_{n \rightarrow +\infty} A_n. \tag{1.4}$$

En effet  $\bigcap_{k>n} A_k \subset A_q$  pour tout  $q > n$ . On a donc  $\bigcap_{k>n} A_k \subset \bigcup_{q>p} A_q$  pour tout  $p$ .  
On a alors pour tout  $n$  :

$$\bigcap_{k>n} A_k \subset \bigcap_{p \geq 1} \bigcup_{q>p} A_q = \limsup_{n \rightarrow +\infty} A_n.$$

Finalement

$$\bigcup_{n \geq 1} \bigcap_{k>n} A_k \subset \limsup_{n \rightarrow +\infty} A_n.$$

C'est à dire (1.4).

Puis, les règles élémentaires sur les  $\bigcup$ ,  $\bigcap$  et  $^c$  donnent sans difficulté :

$$\left( \limsup_{n \rightarrow +\infty} A_n \right)^c = \liminf_{n \rightarrow +\infty} A_n^c.$$

De plus, les limites supérieure et inférieure d'ensembles ont les interprétations suivantes :

**Proposition 1.6.1** Soit  $A_i$ ,  $i \geq 0$ , une collection infinie d'ensembles. Alors

— « À partir d'un certain rang,  $\omega$  est dans tous les  $A_i$  » s'écrit

$$\omega \in \bigcup_{i \geq 0} \bigcap_{j > i} A_j \quad (= \liminf_{n \rightarrow +\infty} A_n).$$

— «  $\omega$  est dans une infinité de  $A_i$  » s'écrit

$$\omega \in \bigcap_{i \geq 0} \bigcup_{j > i} A_j \quad (= \limsup_{n \rightarrow +\infty} A_n).$$

On écrit parfois  $\{A_n\text{ i.s.}\}$  où i.s. signifie « infiniment souvent ».

**Démonstration :**

• Pour le premier point : Soit  $\omega$  qui, à partir d'un certain rang, est dans tous les  $A_i$ . On traduit cela de la façon suivante : il existe un rang  $i$  tel que pour tout rang  $j > i$ ,  $\omega$  est dans  $A_j$ . D'après la signification des symboles  $\forall, \exists, \cap, \cup$ , cela revient à écrire

$$\omega \in \underbrace{\bigcup_{i \geq 0}}_{\text{il existe}} \underbrace{\bigcap_{j > i}}_{\text{pour tout}} \underbrace{A_j}_{\substack{\omega \text{ est} \\ \text{dans } A_j}}$$

• Pour le second point, dire que  $\omega$  est dans une infinité de  $A_i$  est équivalent à dire que

« pour tout  $p$ , il existe  $q > p$  avec  $\omega$  dans  $A_q$ . »

En effet, si tel est le cas,  $\omega$  est bien dans une infinité de  $A_i$  car, d'après cette propriété,

— avec  $p = 0$ , il existe  $p_1 > p$  tel que  $\omega$  est dans  $A_{p_1}$

- avec  $p = p_1$ , il existe  $p_2 > p_1$  tel que  $\omega$  est dans  $A_{p_2}$
- avec  $p = p_2$ , il existe  $p_3 > p_2$  tel que  $\omega$  est dans  $A_{p_3}$
- ...
- avec  $p = p_n$ , il existe  $p_{n+1} > p_n$  tel que  $\omega$  est dans  $A_{p_{n+1}}$
- ...

et finalement,  $\omega$  est dans chaque  $A_{p_n}$ ,  $n \in \mathbb{N}^*$ , c'est à dire dans une infinité de  $A_i$ . Réciproquement, s'il est dans une infinité de  $A_i$ , alors pour tout  $p$ , on trouve  $q > p$  tel que  $\omega \in A_q$ , sinon, ce serait qu'il existe  $p$  tel que pour  $q > p$ ,  $\omega$  n'est pas dans  $A_q$ . Ou encore :  $\omega$  ne peut appartenir qu'aux  $A_i$  d'indice  $i \leq p$ , c'est à dire seulement à un nombre fini d'entre eux, ce qui est faux.

Donc, pour ce deuxième point, pour tout  $p$ , on trouve  $q > p$ , tel que  $\omega \in A_q$ , en langage  $\forall, \exists$ , cela s'écrit

$$\omega \in \underbrace{\bigcap_{\substack{p \geq 0 \\ p \geq 0}}}_{\text{pour tout}} \quad \underbrace{\bigcup_{\substack{q > p \\ q > p}}}_{\text{il existe}} \quad \underbrace{A_q}_{\substack{\omega \text{ est} \\ \text{dans } A_q.}}$$

□

### Lien entre liminf, limsup d'ensembles et de fonctions

**Rappel.** Pour une suite réelle  $u = (u_n)_{n \geq 0}$ , on définit ses limites supérieure et inférieure

$$\begin{aligned}\liminf_{n \rightarrow +\infty} u_n &= \sup_{n \geq 0} \inf_{k \geq n} u_k, \\ \limsup_{n \rightarrow +\infty} u_n &= \inf_{n \geq 0} \sup_{k \geq n} u_k.\end{aligned}$$

Ce sont les plus petite et plus grande valeurs d'adhérence de la suite  $u$ . On a toujours

$$\liminf_{n \rightarrow +\infty} u_n \leq \limsup_{n \rightarrow +\infty} u_n$$

et il y a égalité si et seulement si la suite  $u$  converge ; de plus si tel est le cas

$$\lim_{n \rightarrow +\infty} u_n = \liminf_{n \rightarrow +\infty} u_n = \limsup_{n \rightarrow +\infty} u_n.$$

En plus, en changeant le signe, les limites inférieure et supérieure s'échangent :

$$\begin{aligned}\liminf_{n \rightarrow +\infty} (-u_n) &= -\limsup_{n \rightarrow +\infty} u_n, \\ \limsup_{n \rightarrow +\infty} (-u_n) &= -\liminf_{n \rightarrow +\infty} u_n.\end{aligned}$$

Pour une suite de fonctions  $(f_n)_{n \geq 0}$ , on définit des fonctions limites inférieure et supérieure de la façon suivante :

$$\left( \liminf_{n \rightarrow +\infty} f_n \right) (x) = \liminf_{n \rightarrow +\infty} f_n(x), \quad \left( \limsup_{n \rightarrow +\infty} f_n \right) (x) = \limsup_{n \rightarrow +\infty} f_n(x).$$

## TEXT

17

**Proposition 1.6.2** Soit  $(A_n)_{n \geq 0}$  une suite d'ensembles mesurables, on a

$$\liminf_{n \rightarrow +\infty} \mathbf{1}_{A_n} = \mathbf{1}_{\{\liminf_{n \rightarrow +\infty} A_n\}}, \quad \limsup_{n \rightarrow +\infty} \mathbf{1}_{A_n} = \mathbf{1}_{\{\limsup_{n \rightarrow +\infty} A_n\}}.$$

**Démonstration :** Il suffit de le faire pour  $\limsup_{n \rightarrow +\infty} \mathbf{1}_{A_n}$ . Or  $\omega \in \limsup_{n \rightarrow +\infty} A_n$  si  $\forall k, \exists n \geq k$  tel que  $\omega \in A_n$ . On a donc  $\forall k, \exists n \geq k$  avec  $\mathbf{1}_{A_n}(\omega) = 1$  ou encore  $\inf_{k \geq 0} \sup_{n \geq k} \mathbf{1}_{A_n}(\omega) = 1$ , c'est à dire  $\limsup_{n \rightarrow +\infty} \mathbf{1}_{A_n}(\omega) = 1$ . Manifestement, il s'agit d'une équivalence.  $\square$

De plus, on a :

**Proposition 1.6.3**

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \{f_n \leq t\} &= \{\limsup_{n \rightarrow +\infty} f_n \leq t\}, & \limsup_{n \rightarrow +\infty} \{f_n \leq t\} &= \{\liminf_{n \rightarrow +\infty} f_n \leq t\}, \\ \liminf_{n \rightarrow +\infty} \{f_n \geq t\} &= \{\liminf_{n \rightarrow +\infty} f_n \geq t\}, & \limsup_{n \rightarrow +\infty} \{f_n \geq t\} &= \{\limsup_{n \rightarrow +\infty} f_n \geq t\}. \end{aligned}$$

**Démonstration :** exercice.

**Définition 1.6.2 (Convergence d'ensembles)** On dit qu'une suite d'ensemble  $(A_n)_{n \geq 0}$  converge vers  $A$  et on note  $A_n \rightarrow A$ ,  $n \rightarrow +\infty$ , si

$$\limsup_{n \rightarrow +\infty} A_n = \liminf_{n \rightarrow +\infty} A_n = A.$$

On montre alors aussi :

**Proposition 1.6.4**

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n\right) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right).$$

On en déduit que si  $A_n \rightarrow A$  alors  $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$ ,  $n \rightarrow +\infty$ .

**Démonstration :** Soient  $B_n = \bigcup_{k > n} A_k$  et  $C_n = \bigcap_{k > n} A_k$ . Alors  $B_n$  est une suite décroissante d'ensembles, de limite  $\limsup_{n \rightarrow +\infty} A_n$  et  $C_n$  est une suite croissante d'ensembles, de limite  $\liminf_{n \rightarrow +\infty} A_n$ . D'après la propriété de convergence monotone des mesures (de probabilités, cf. Chapitre 1), on a

$$\begin{aligned} \mathbb{P}(A_n) &\leq \mathbb{P}(B_n) \longrightarrow \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) \\ \mathbb{P}(A_n) &\geq \mathbb{P}(C_n) \longrightarrow \mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n\right) \end{aligned}$$

On en déduit

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n\right) \leq \liminf_{n \rightarrow +\infty} \mathbb{P}(A_n) \quad \text{et} \quad \limsup_{n \rightarrow +\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right).$$

Le second point est clair car les limites supérieure et inférieure coïncident lorsque  $A_n \rightarrow A$ ,  $n \rightarrow +\infty$ .  $\square$

# Chapitre 2

## Variables aléatoires

Les variables aléatoires sont l'objet de base de ce cours. Il s'agit de fonction du hasard dont nous présentons dans ce chapitre les outils clef pour leur étude : loi, fonction de répartition (les moments sont étudiés au Chapitre 3). Nous donnons la plupart des lois usuelles (discrètes et continues) et insistons sur leur interprétation en terme de modéle. Dans tout le chapitre, on considère un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ .

### 2.1 Définitions et propriétés

**Définition 2.1.1 (Variable aléatoire)** *On appelle variable aléatoire (va) toute application mesurable  $X$  d'un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  dans un espace mesurable  $(B, \mathcal{B})$ .*

- En général, une variable aléatoire est à valeurs réelles, ie.  $(B, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .
- Si  $(B, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , on parle de **vecteur aléatoire** de dimension  $n$ . Dans ce cas, on peut écrire  $X = (X_1, \dots, X_n)$  où  $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est une variable aléatoire appelée  $i$ -ème marginale.
- Si  $(B, \mathcal{B}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ , on parle en particulier de **couple aléatoire**.
- Si  $(B, \mathcal{B}) = (\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R}^\mathbb{N}))$ , on parle de **suite aléatoire**.

Une variable aléatoire est donc tout simplement une application mesurable sur un espace de probabilité.

En pratique, lorsque la tribu de l'espace d'arrivée  $\mathcal{B}$  est engendrée par un système de générateurs, pour vérifier qu'une fonction est mesurable (en particulier, variable aléatoire), il suffit de vérifier la propriété caractéristique sur les générateurs. Cela justifie l'importance de connaître les familles de générateurs les plus petites d'une tribu.

**Proposition 2.1.1** 1. Soient  $X : (\Omega, \mathcal{F}) \rightarrow (B, \mathcal{B})$  et  $\mathcal{C} \subset \mathcal{P}(B)$ , famille qui engendre la tribu  $\mathcal{B}$  de  $B$ . Alors

$$X^{-1}(\mathcal{B}) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})) \quad (= \sigma(X^{-1}(C) : C \in \mathcal{C})).$$

2. En particulier pour qu'une fonction  $X : (\Omega, \mathcal{F}) \rightarrow (B, \sigma(\mathcal{C}))$  soit mesurable, il faut et suffit que  $X^{-1}(\mathcal{C}) \subset \mathcal{F}$ .

3. Une fonction  $X : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire réelle si et seulement si pour tout  $t \in \mathbb{R}$ ,  $\{X \leq t\} \in \mathcal{F}$ .

**Démonstration :** 1) Il est clair que  $X^{-1}(\sigma(\mathcal{C}))$  est une tribu qui contient  $X^{-1}(\mathcal{C})$ . Par conséquent  $X^{-1}(\sigma(\mathcal{C})) \supseteq \sigma(X^{-1}(\mathcal{C}))$ .

Pour l'inclusion inverse, on considère

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

On vérifie facilement que  $\mathcal{T}$  est une tribu. Par sa définition  $X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C}))$ . De plus  $\mathcal{C} \subset \mathcal{T}$  car  $X^{-1}(\mathcal{C}) \subset \sigma(X^{-1}(\mathcal{C}))$  et donc  $\sigma(\mathcal{C}) \subset \mathcal{T}$ . On déduit que

$$X^{-1}(\sigma(\mathcal{C})) \subset X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C})).$$

On peut traiter de la même façon le cas d'un nombre quelconque de fonctions.

2) Si  $X^{-1}(\mathcal{C}) \subset \mathcal{F}$  alors  $\sigma(X^{-1}(\mathcal{C})) \subset \mathcal{F}$  car  $\mathcal{F}$  est une tribu. Comme par 1)  $X^{-1}(\mathcal{B}) = \sigma(X^{-1}(\mathcal{C}))$ , la conclusion en découle.

3) Cela vient de 2) et du fait que la tribu borélienne  $\mathcal{B}(\mathbb{R})$  est engendrée par les demi-droites  $] -\infty, t]$ ,  $t \in \mathbb{R}$ .  $\square$

**Proposition 2.1.2** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires de  $\Omega$  dans un espace métrique  $(E, d)$ . Si cette suite de variable aléatoire converge ponctuellement vers  $X$  (c'est à dire pour tout  $\omega \in \Omega$ ,  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ ) alors  $X$  est une variable aléatoire à valeurs dans  $E$ .

**Démonstration :** D'après la Proposition 2.1.1, il suffit de montrer que si  $O$  est un ouvert de  $E$  alors  $X^{-1}(O) \in \mathcal{F}$ . Pour cela, on pose

$$O_r = \{x \in O : d(x, E \setminus O) > 1/r\}, \quad r \in \mathbb{N}^*.$$

L'ensemble  $O_r$  est ouvert donc borélien de  $E$  et

$$\bigcup_{r \geq 1} O_r = \{x \in O : d(x, E \setminus O) > 0\} = O.$$

On a

$$\begin{aligned} X^{-1}(O) &= \{\omega \in \Omega : X(\omega) \in O\} = \{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) \in O\} \\ &= \left\{ \omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) \in \bigcup_{r \geq 1} O_r \right\} \\ &= \{\omega \in \Omega : \exists r \geq 1, \exists m \geq 1, \forall n \geq m : X_n(\omega) \in O_r\} \\ &= \bigcup_{r, m \in \mathbb{N}^*} \bigcap_{n \geq m} X_n^{-1}(O_r) \end{aligned}$$

est un événement de  $\mathcal{F}$ .  $\square$

**Définition 2.1.2 (Variable aléatoire simple ou étagée)** On appelle variable aléatoire étagée ou simple (à valeurs dans  $\mathbb{R}^d$ ) une variable aléatoire définie sur  $\Omega$  de la forme  $X = \sum_{j=1}^r a_j \mathbf{1}_{A_j}$  où les  $A_j \in \mathcal{F}$  sont disjoints et où  $a_j \in \mathbb{R}^d$ .

**Proposition 2.1.3 (Approximation)** Toute variable aléatoire  $X$  est limite simple de variables aléatoires étagées. Si de plus  $X$  est réelle positive, la limite peut être choisie croissante.

**Démonstration :** Soit d'abord  $X$  réelle positive. On définit pour tout  $n$  et  $k \in \mathbb{N}^*$

$$\begin{aligned} A_{n,k} &= \left\{ \omega : \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n} \right\}, \quad k \leq n2^n \\ B_n &= \{\omega : X(\omega) > n\}. \end{aligned}$$

Les ensembles  $A_{n,k}$  et  $B_n$  sont des images réciproques par la variable aléatoire  $X$  d'intervalles. Ils sont donc dans  $\mathcal{F}$ . La suite

$$X_n(\omega) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}}(\omega) + n \mathbf{1}_{B_n}(\omega)$$

converge en croissant vers  $X(\omega)$ . En effet, comme

$$\left[ \frac{k-1}{2^n}, \frac{k}{2^n} \right] \subset \left[ \frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}} \right] \cup \left[ \frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}} \right]$$

on a  $A_{n,k} = A_{n+1,2k-1} \cup A_{n+1,2k}$ . Ainsi pour  $X_n(\omega) = \frac{k-1}{2^n}$  on a soit  $X_{n+1}(\omega) = \frac{2k-2}{2^{n+1}} = X_n(\omega)$  soit  $X_{n+1}(\omega) = \frac{2k-1}{2^{n+1}} \geq X_n(\omega)$  donc en tout cas  $X_{n+1}(\omega) \geq X_n(\omega)$ .

Puis si  $X(\omega) > n$  alors soit  $X(\omega) > n+1$  et alors  $X_{n+1}(\omega) = n+1 \geq X_n(\omega)$ , soit  $X(\omega)$  est dans

$$[n, n+1[ = \bigcup_{k=2^{n+1}n+1}^{2^{n+1}(n+1)} \left[ \frac{k-1}{2^{n+1}}, \frac{k}{2^{n+1}} \right].$$

Pour  $k \in [2^{n+1}n+1, 2^{n+1}(n+1)]$ , on a alors  $X_{n+1}(\omega) = \frac{k-1}{2^{n+1}} \geq \frac{2^{n+1}n}{2^{n+1}} = n = X_n(\omega)$  et donc  $X_{n+1}(\omega) \geq X_n(\omega)$ .

Pour la convergence : si  $X(\omega) = +\infty$  alors on a clairement  $X_n(\omega) = n \rightarrow +\infty$  soit  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ . Tandis que si  $X(\omega) < +\infty$  alors pour  $n \geq [X(\omega)] + 1$ , on a  $X(\omega) \leq n$  et donc  $X_n(\omega) = \frac{k-1}{2^n}$  pour  $X(\omega) \in [\frac{k-1}{2^n}, \frac{k}{2^n}[$ , ie.  $|X(\omega) - X_n(\omega)| \leq 2^{-n}$ . On a donc  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ .

Si  $X$ , réelle, est de signe quelconque, on écrit  $X = X^+ - X^-$  avec  $X^+ = \max(X, 0)$ ,  $X^- = \max(-X, 0)$  et on approxime  $X^+$  et  $X^-$  comme précédemment.

Si  $X$  est à valeurs dans  $\mathbb{R}^d$ , on applique la stratégie précédente à chaque marginale.  $\square$

## 2.2 Loi d'une variable aléatoire

**Définition 2.2.1 (Loi)** Soit  $X : \Omega \rightarrow (B, \mathcal{B})$ , on appelle loi de  $X$  la mesure  $\mathbb{P}_X$ , mesure image sur  $(B, \mathcal{B})$  de  $\mathbb{P}$  par  $X$  :

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\omega \in \Omega : X(\omega) \in A), \quad A \in \mathcal{B}.$$

En effet, on vérifie facilement que  $\mathbb{P}_X$  est  $\sigma$ -additive : pour des ensembles  $A_i$ ,  $i \geq 1$ , mesurables disjoints, on a :

$$\begin{aligned} \mathbb{P}_X\left(\bigcup_{i=1}^{+\infty} A_i\right) &= \mathbb{P}\left(X \in \bigcup_{i=1}^{+\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{+\infty} \{X \in A_i\}\right) \\ &= \sum_{i=1}^{+\infty} \mathbb{P}(X \in A_i) = \sum_{i=1}^{+\infty} \mathbb{P}_X(A_i) \end{aligned}$$

en utilisant la  $\sigma$ -additivité de  $\mathbb{P}$ . En fait, la loi  $\mathbb{P}_X$  d'une variable aléatoire  $X$  définit une mesure de probabilité sur  $(B, \mathcal{B})$  :  $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = 1$ .

- En général, pour une variable aléatoire on a  $(B, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et  $\mathbb{P}_X$  est une mesure sur  $\mathbb{R}$ .
- Si  $X = (X_1, \dots, X_n)$  est un vecteur, ie.  $(B, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ , la loi  $\mathbb{P}_{(X_1, \dots, X_n)}$  s'appelle la **loi jointe** des variables aléatoires  $X_1, \dots, X_n$ . C'est une loi sur  $\mathbb{R}^n$ . Les lois des variables marginales s'appellent les lois marginales.

**Proposition 2.2.1** Si  $(X, Y)$  est un couple de loi  $\mathbb{P}_{X,Y} = \mu$  alors les lois marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  de  $X$  et  $Y$  s'obtiennent par

$$\mathbb{P}_X(A) = \mu(A \times \mathbb{R}), \quad \mathbb{P}(Y \in A) = \mu(\mathbb{R} \times A), \quad A \in \mathcal{B}(\mathbb{R}).$$

De même pour un vecteur  $X = (X_1, \dots, X_n)$  :

$$\mathbb{P}_{X_i}(A) = \mathbb{P}_{(X_1, \dots, X_n)}(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{n-i}), \quad A \in \mathcal{B}(\mathbb{R}).$$

**Démonstration :** C'est évident si on remarque par exemple que  $\{X \in A\} = \{(X, Y) \in A \times \mathbb{R}\}$ . De même pour un vecteur.  $\square$

**Définition 2.2.2 (Atome)** Soit  $X$  une variable aléatoire. On appelle atome de  $X$  (ou de sa loi) tout  $x \in B$  tel que  $\mathbb{P}(X = x) \neq 0$ .

Si  $x$  est un atome de  $X$ , on lui associe la probabilité  $\mathbb{P}(X = x)$ , dite probabilité ponctuelle associée à l'atome  $x$ .

**Définition 2.2.3 (Support)** On appelle support (topologique) d'une variable aléatoire  $X : \Omega \rightarrow B$  le plus petit ensemble  $F$  fermé de  $B$  tel que  $\mathbb{P}(X \in F) = 1$ . Avec un abus de notation, on notera dans la suite  $\mathcal{S}(X)$  ce support.

Pour une variable aléatoire réelle, la connaissance de la loi est importante : dans la suite, la plupart des calculs de probabilité pour  $X$  seront transférés en des calculs sur  $\mathbb{R}$  où il faut utiliser la mesure  $\mathbb{P}_X$  (cf. Théorème de transfert). Dans la suite, on considère en général des variables aléatoires à valeurs réelles (ie.  $B = \mathbb{R}$ ).

## 2.3 Fonction de répartition

**Définition 2.3.1 (Répartition)** On appelle fonction de répartition d'une variable aléatoire réelle  $X : \Omega \rightarrow \mathbb{R}$  la fonction  $F_X$  définie sur  $\mathbb{R}$  par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]).$$

**Proposition 2.3.1 (Propriétés de la fonction de répartition)**

1.  $F_X$  est croissante.
2.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .
3.  $F_X$  est continue à droite et admet une limite à gauche :

$$\lim_{x \rightarrow x_0, x \leq x_0} F_X(x) = \mathbb{P}(X < x_0).$$

On dit que la fonction  $F_X$  est càdlàg (continue à droite avec une limite à gauche).

4. En fait si  $x$  n'est pas un atome de  $X$ , alors  $F_X$  est continue à gauche (donc continue) en  $x$ . L'ensemble des points de discontinuité de  $F_X$  est l'ensemble des atomes de  $X$ .

**Démonstration :** 1) est dû à la croissance de la mesure de probabilité  $\mathbb{P}$  : pour  $x \leq y$ , on a :

$$F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$$

car  $\{X \leq x\} \subset \{X \leq y\}$ .

2) et 3) s'obtiennent en appliquant les propriétés de monotonie séquentielle des mesures. Dans ces arguments, on utilise systématiquement la monotonie de  $F_X$  pour justifier que les limites de  $F$  coïncident avec les limites le long de sous-suites spéciales.

- Pour 2), prendre d'abord  $A_n = ]-\infty, -n]$ , on a  $\bigcap_{n \geq 1} A_n = \emptyset$ , ensemble de mesure  $\mathbb{P}_X$  nulle, si bien que

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq -n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X\left(\bigcap_{n \geq 1} A_n\right) = 0.$$

Puis prendre  $B_n = ]-\infty, n]$  de réunion  $\bigcup_{n \geq 1} B_n = ]-\infty, +\infty[ = \mathbb{R}$  de mesure  $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$  si bien que

$$\lim_{x \rightarrow +\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X\left(\bigcup_{n \geq 1} B_n\right) = \mathbb{P}_X(\mathbb{R}) = 1.$$

- Pour le 3), prendre d'abord  $A_n = ]-\infty, x + 1/n]$  d'intersection  $\bigcap_{n \geq 1} A_n = ]-\infty, x]$ , ensemble de mesure  $\mathbb{P}_X\left(\bigcap_{n \geq 1} A_n\right) = F_X(x)$  si bien que

$$\lim_{y \rightarrow x^+} F_X(y) = \lim_{n \rightarrow +\infty} F_X(x + 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x + 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n)$$

$$= \mathbb{P}_X \left( \bigcap_{n \geq 1} A_n \right) = F_X(x).$$

Ensuite, prendre  $B_n = ]-\infty, x - 1/n]$  de réunion  $\bigcup_{n \geq 1} B_n = ]-\infty, x[$ , ensemble de mesure  $\mathbb{P}_X \left( \bigcup_{n \geq 1} B_n \right) = \mathbb{P}(X < x)$ .

**Attention :**  $\mathbb{P}(X < x)$  peut être distinct de  $\mathbb{P}(X \leq x)$  car

$$\mathbb{P}(X \leq x) - \mathbb{P}(X < x) = \mathbb{P}(\{X \leq x\} \setminus \{X < x\}) = \mathbb{P}(X = x)$$

qui peut être non nul si la loi de  $X$  a un atome en  $x$ . On a alors

$$\begin{aligned} \lim_{y \rightarrow x^-} F_X(y) &= \lim_{n \rightarrow +\infty} F_X(x - 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x - 1/n) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X \left( \bigcup_n B_n \right) = \mathbb{P}(X < x). \end{aligned}$$

- Pour le 4), on constate que si  $\mathbb{P}(X = x) = 0$  alors  $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$  et la continuité à gauche manquante vient. Réciproquement si  $x$  est un atome alors  $\mathbb{P}(X < x) \neq \mathbb{P}(X \leq x)$  et  $\lim_{y \nearrow x} F_X(y) < F_X(x)$  et il y a une discontinuité de  $F_X$  en  $x$ .  $\square$

**Théorème 2.3.1** Une fonction  $F : \mathbb{R} \rightarrow [0, 1]$  croissante càdlàg avec  $\lim_{t \rightarrow -\infty} F(t) = 0$  et  $\lim_{t \rightarrow +\infty} F(t) = 1$  est la fonction de répartition d'une variable aléatoire  $X$ . De plus l'ensemble des points où la fonction  $F$  a un saut est l'ensemble des atomes de  $X$ .

**Démonstration :** Soit  $F$  une fonction croissante et continue à droite de limites 0 et 1 respectivement en  $-\infty$  et en  $+\infty$ . On définit la fonction d'ensemble  $m$  par  $m(A) = \sum_{i=1}^n (F(b_i) - F(a_i))$  pour une union disjointe  $A = \bigcup_{i=1}^n ]a_i, b_i]$ . Notons que la fonction  $m$  est bien définie car si on a aussi  $A = \bigcup_{j=1}^p ]c_j, d_j]$ , alors on vérifie que  $\sum_{j=1}^p (F(d_j) - F(c_j)) = \sum_{i=1}^n (F(b_i) - F(a_i))$ .

On note  $\mathcal{C}$  la classe des ensembles qui sont unions disjointes finies d'intervalles du type  $]a, b]$ . Comme  $\mathcal{C}$  est stable par union (finie) disjointe et comme si  $]a, b] \cap ]c, d] \neq \emptyset$ , on a  $]a, b] \cup ]c, d] = ]a \wedge c, b \vee d] \in \mathcal{C}$ , on déduit que  $\mathcal{C}$  est stable par union finie. Puis

$$\left( \bigcup_{i=1}^n ]a_i, b_i] \right) \cap \left( \bigcup_{j=1}^p ]c_j, d_j] \right) = \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} (]a_i, b_i] \cap ]c_j, d_j]) = \bigcup_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} ]a_i \vee c_j, b_i \wedge d_j] \in \mathcal{C},$$

$\mathcal{C}$  est donc stable par intersection finie. Enfin, notons que

$$]a, b] \setminus ]c, d] = ]a, b] \cap (]-\infty, c] \cup ]d, +\infty[) = ]a, b \wedge c] \cup ]a \vee d, b] \in \mathcal{C}$$

et de même pour des réunions finies d'intervalles  $A = \bigcup_{i=1}^n ]a_i, b_i]$  si bien que  $\mathcal{C}$  est stable par différence. Notons que la restriction  $\mathcal{C}_n$  de  $\mathcal{C}$  à chaque intervalle  $]-n, n]$  est une algébre de  $]-n, n]$ .

On montre maintenant que  $m$  est une mesure d'algébre sur  $\mathcal{C}_n$ , cf. Définition 1.4.3. On pourra alors prolonger  $m$  en une mesure sur  $] -n, n]$ .

Pour cela, on voit d'abord aisément que  $m$  est additif : si  $A, B \in \mathcal{C}$  sont disjoints on a  $m(A \cup B) = m(A) + m(B)$ .

Soit  $A = ]a, b]$  et  $\varepsilon > 0$  fixé. Comme  $F$  est continue à droite, il existe  $0 < \delta < b - a$  tel que  $m(]a, a + \delta]) = F(a + \delta) - F(a) \leq \varepsilon$ . Alors  $[a + \delta, b] \subset ]a, b]$  et

$$m(]a, b]) = m(]a, a + \delta]) + m(]a + \delta, b]) \leq m(]a + \delta, b]) + \varepsilon.$$

On procéde de la même façon si  $A = \bigcup_{i=1}^n ]a_i, b_i] \in \mathcal{C}$  est une réunion finie et on montre qu'il existe  $B \in \mathcal{C}$  tel que  $\bar{B} \subset A$  et  $m(A) \leq m(B) + \varepsilon$ .

Soit  $(A_n)_{n \geq 1}$  une suite décroissante de  $\mathcal{C}$  telle que  $\bigcap_{n \geq 1} A_n = \emptyset$ . Comme précédemment on associe  $B_n$  à  $A_n$  pour  $\varepsilon/2^n$ . Alors  $\bigcap_{n \geq 1} \bar{B}_n \subset \bigcap_{n \geq 1} A_n = \emptyset$ . Les ensembles  $B_n$  sont des réunions finies d'intervalles fermés bornés donc compacts. Par compacité, il existe  $n_0$  tel que  $\bigcap_{n \leq n_0} B_n = \emptyset$ . On a

$$A_{n_0} = A_{n_0} \setminus \bigcap_{n=1}^{n_0} \bar{B}_n = \bigcup_{n=1}^{n_0} (A_{n_0} \setminus \bar{B}_n) \subset \bigcup_{n=1}^{n_0} (A_n \setminus B_n).$$

Ainsi

$$m(A_{n_0}) \leq \sum_{n=1}^{n_0} m(A_n \setminus B_n) \leq \sum_{n=1}^{n_0} \varepsilon 2^{-n} = \varepsilon.$$

On a donc  $\lim_{n \rightarrow +\infty} m(A_n) = 0$ , autrement dit  $m$  vérifie la propriété de Carathéodory et finalement  $m$  est une mesure d'algébre sur l'algébre donnée par la restriction de  $\mathcal{C}$  à chaque  $] -n, n]$ .

Par le théorème de Carathéodory (Théorème 1.4.3),  $m_{]-n,n]}$  s'étend de manière unique en une mesure  $\mu_n$  sur  $] -n, n]$ . Par unicité, on déduit une mesure  $\mu$  sur  $\mathcal{B}(\mathbb{R})$  telle que

$$\mu(]a, b]) = F(b) - F(a).$$

Par convergence monotone pour la mesure  $\mu$  et d'après la limite de  $F$  en  $-\infty$ , avec  $a \rightarrow -\infty$  et  $b = x$ , on a  $\mu(]-\infty, x]) = F(x)$ . Puis, encore avec convergence monotone pour la mesure  $\mu$  avec  $x \rightarrow +\infty$  et d'après la limite de  $F$  en  $+\infty$ , on a  $\mu(\mathbb{R}) = \lim_{x \rightarrow +\infty} F(x) = 1$  et  $\mu$  est une probabilité. L'expression  $\mu(]-\infty, x]) = F(x)$  justifie alors que  $F$  est la fonction de répartition de  $\mu$ .

Pour finir, on déduit que les atomes de la loi  $\mu$  sont les sauts de  $F$  comme on l'a déjà fait précédemment en partant d'un loi  $\mathbb{P}_X$ .

**Alternativement**,  $\mathcal{C}$  est non vide, stable par différence et union finie. Il s'agit donc d'un anneau d'ensemble au sens de la Définition 1.4.1. Puis  $m$  est additive et vérifie  $m(\bigcap_{n \geq 1} A_n) = 0$  pour toute suite décroissante  $(A_n)_{n \geq 1}$  de  $\mathcal{C}$  telle que  $\bigcap_{n \geq 1} A_n = \emptyset$ . D'après le Lemme 2.3.1 suivant,  $m$  est  $\sigma$ -additive sur  $\mathcal{C}$  : en fait, on applique la Remarque 2.3.1 : si  $A_n \in \mathcal{C}$  sont

deux à deux disjoints avec  $A_\infty = \bigcup_{n \geq 1} A_n \in \mathcal{C}$  alors nécessairement  $A_\infty$  s'écrit sous la forme d'une réunion finie  $\bigcup_{i=1}^n [a_i, b_i]$  et  $m(A_\infty) = \sum_{i=1}^n F(b_i) - F(a_i)$  est finie.

Comme  $m$  est  $\sigma$ -finie ( $\mathbb{R} = \bigcup_{n \geq 1} [-n, n]$  et  $m([-n, n]) = F(n) - F(-n) < +\infty$ , la version Théorème 1.4.1 du Théorème de Carathéodory s'applique : il existe une unique extension  $\mu$  à  $\sigma(\mathcal{C})$ . On montre ensuite comme précédemment que  $\mu$  est une probabilité et que  $F$  est sa fonction de répartition.

**Lemme 2.3.1** Soient  $E$  un ensemble muni d'une famille  $\mathcal{E} \subset \mathcal{P}(E)$  et  $m : \mathcal{E} \rightarrow [0, +\infty]$  une application additive (c'est-à-dire  $m(A \cup B) = m(A) + m(B)$  pour  $A, B \in \mathcal{E}$  et  $A \cap B = \emptyset$ ) et telle que  $m(E) < +\infty$ . Alors les assertions suivantes sont équivalentes :

1.  $m$  est  $\sigma$ -additive ;
2.  $m$  est continue sur les suites croissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{E}, \quad A_n \subset A_{n+1} \Rightarrow m\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} m(A_n);$$

3.  $m$  est continue sur les suites décroissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{E}, \quad A_n \supset A_{n+1} \Rightarrow m\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} m(A_n);$$

4.  $m$  est continue sur les suites décroissantes vers  $\emptyset$  :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{E}, \quad A_n \supset A_{n+1} \text{ et } \bigcap_{n \in \mathbb{N}} A_n = \emptyset \Rightarrow \lim_{n \rightarrow +\infty} m(A_n) = 0.$$

**Démonstration :** Il est clair que (1)  $\Rightarrow$  (2)  $\Leftrightarrow$  (3)  $\Rightarrow$  (4). (L'équivalence (2)  $\Leftrightarrow$  (3) est due au fait que  $m(E) < +\infty$  permet de passer au complémentaire). On a (4)  $\Rightarrow$  (3) car si on a  $A_n \supset A_{n+1}$ , en notant  $A_\infty = \bigcap_{n \in \mathbb{N}} A_n$ , la suite  $A_n \cap A_\infty^c$  est décroissante vers  $\bigcap_{n \geq 1} A_n \cap A_\infty^c = A_\infty \cap A_\infty^c = \emptyset$ . D'après le (4), on a  $m(A_n \cap A_\infty^c) \rightarrow 0$  et on a alors

$$\begin{aligned} m(A_\infty) &= m(A_n \cap A_\infty) = \lim_{n \rightarrow +\infty} m(A_n \cap A_\infty) + \lim_{n \rightarrow +\infty} m(A_n \cap A_\infty^c) \\ &= \lim_{n \rightarrow +\infty} m(A_n \cap (A_\infty \cup A_\infty^c)) = \lim_{n \rightarrow +\infty} m(A_n). \end{aligned}$$

Puis (2)  $\Rightarrow$  (1) car pour une suite de  $A_i$  disjoints, en notant  $C_n = \bigcup_{i=1}^n A_i$  on a une suite  $(C_n)_{n \geq 1}$  croissante à laquelle (2) s'applique et  $\bigcup_{n \geq 1} C_n = \bigcup_{i \geq 1} A_i$ . Ainsi :

$$m\left(\bigcup_{i \geq 1} A_i\right) = m\left(\bigcup_{n \geq 1} C_n\right) = \lim_{n \rightarrow +\infty} m(C_n).$$

Mais par additivité,  $m(C_n) = m\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n m(A_i)$  de limite  $\sum_{i=1}^{+\infty} m(A_i)$ . □

**Remarque 2.3.1** Dans les équivalences précédentes, l'hypothèse  $m(E) < +\infty$  ne sert que pour voir  $(2) \Leftrightarrow (3)$ . Si ce n'est pas le cas, on a quand même : Soit  $(A_n)_{n \in \mathbb{N}}$  une suite croissante de  $\mathcal{E}$  de réunion  $A_\infty = \bigcup_{n=1}^{+\infty} A_n$  alors on déduit (2) pour ces suites de (4) sans supposer  $m(E)$  finie : on pose  $B_n = A_\infty \setminus A_n$ ;  $(B_n)_{n \geq 1}$  est alors une suite décroissante de limite  $\bigcap_{n \geq 1} B_n = A_\infty \setminus \bigcup_{n=1}^{+\infty} A_n = \emptyset$ . D'après (4),  $\lim_{n \rightarrow +\infty} m(B_n) = 0$  et donc de

$$m(A_\infty) = m(A_\infty \setminus A_n) + m(A_n) = m(A_n) + m(B_n)$$

on déduit  $\lim_{n \rightarrow +\infty} m(A_n) = m(A_\infty)$ . On déduit encore comme dans la preuve précédente que pour des  $A_n$  disjoints de réunion  $A_\infty = \bigcup_{n=1}^{+\infty} A_n$  telle que  $m(A_\infty) < +\infty$ , on a  $m(A_\infty) = \sum_{n=1}^{+\infty} m(A_n)$ .

**Théorème 2.3.2** La fonction de répartition caractérise la loi, c'est à dire  $F_X = F_Y$  si et seulement si  $\mathbb{P}_X = \mathbb{P}_Y$ .

**Démonstration :** Si  $F_X = F_Y$  alors  $\mathbb{P}_X(]-\infty, t]) = \mathbb{P}_Y(]-\infty, t])$  pour tout  $t \in \mathbb{R}$ . La famille  $\mathcal{C}$  de ces intervalles  $]-\infty, t]$ ,  $t \in \mathbb{R}$ , est stable par intersection finie (c'est un  $\pi$ -système) et engendre  $\mathcal{B}(\mathbb{R})$ . Par le Théorème de Dynkin (Théorème 1.3.2, conséquence du Théorème 1.3.1 de classes monotones), l'égalité  $\mathbb{P}_X(A) = \mathbb{P}_Y(A)$  s'étend à  $A \in \sigma(\mathcal{C})$ , ie.  $\mathbb{P}_X = \mathbb{P}_Y$ .  $\square$

**Proposition 2.3.2** La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité. Une variable aléatoire réelle a donc un nombre au plus dénombrable d'atome.

**Démonstration :** Désignons par  $D_n$  l'ensemble des points de discontinuité avec un saut d'amplitude plus grande que  $1/n$  :

$$D_n = \left\{ t \in \mathbb{R} : F(t) - F(t^-) \geq \frac{1}{n} \right\}.$$

Comme  $0 \leq F(t) \leq 1$ , on a nécessairement  $\text{card}(D_n) \leq n$ . L'ensemble des points de discontinuité  $D = \bigcup_{n \in \mathbb{N}^*} D_n$  est donc lui aussi dénombrable.  $\square$

**Définition 2.3.2** Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . On appelle fonction de répartition de  $X$  ou de sa loi  $\mathbb{P}_X$ , la fonction définie sur  $\mathbb{R}^d$  par

$$F_X(t_1, \dots, t_d) = \mathbb{P}_X(]-\infty, t_1] \times \dots \times ]-\infty, t_d]) = \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d).$$

La loi de la  $j$ -ème marginale  $X_j$  est donnée par

$$F_{X_j}(t_j) = \lim_{t_1, \dots, t_{j-1}, t_{j+1}, t_d \rightarrow +\infty} F_X(t).$$

La loi d'un vecteur aléatoire détermine chacune de ses marginales mais la réciproque est fausse en général comme le montre le contre-exemple suivant : Soit  $(X, Y)$  de loi

$$\begin{aligned}\mathbb{P}_{X,Y} &= \frac{1}{6}\delta_{(0,0)} + \frac{1}{3}\delta_{(0,1)} + \frac{1}{12}\delta_{(1,0)} + \frac{5}{12}\delta_{(1,1)} \\ \mathbb{P}_{U,V} &= \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,1)}.\end{aligned}$$

Les lois marginales sont alors

$$\begin{aligned}\mathbb{P}_X = \mathbb{P}_U &= \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \\ \mathbb{P}_Y = \mathbb{P}_V &= \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1.\end{aligned}$$

Les lois jointes sont distinctes mais elles partagent donc les mêmes marginales. À partir des marginales, on ne connaît donc pas la loi jointe. Il manque la connaissance de la façon dont les lois marginales s'imbriquent pour connaître la loi jointe. Un cas particulier d'imbrication est l'indépendance. Ce cas très fréquent est étudié dans la suite.

## 2.4 Fonction quantile et simulation par inversion

**Définition 2.4.1 (Fonction quantile)** Soit  $F$  une fonction de répartition. On appelle fonction quantile la fonction définie sur  $]0, 1[$  par

$$G(u) = \inf(t \in \mathbb{R} : F(t) > u), \quad u \in ]0, 1[.$$

On considère dans le résultat suivant la loi uniforme sur  $]0, 1[$  : une variable aléatoire  $U$  suit cette loi si  $\mathbb{P}(U \in A) = \lambda(A)$  pour tout  $A \in \mathcal{B}(]0, 1[)$ . En particulier, on peut noter que cette loi n'a pas d'atome : pour tout  $x \in \mathbb{R}$ , on a  $\mathbb{P}(U = x) = \lambda(\{x\}) = 0$ .

**Proposition 2.4.1** Soit  $U$  une variable aléatoire uniforme sur  $]0, 1[$ . Alors  $G(U)$  a pour fonction de répartition  $F$ .

**Démonstration :** Notons que pour tout  $t \in \mathbb{R}$ , si  $F(t) > U$  alors  $G(U) \leq t$ , ie.

$$\{U < F(t)\} \subset \{G(U) \leq t\}. \quad (2.1)$$

Puis si  $G(U) = \inf(s \in \mathbb{R} : F(s) > U) \leq t$  alors par croissance de la fonction  $F$ ,  $F(s) > U$  pour tout  $s > t$ , ie.

$$\{G(U) \leq t\} \subset \{U < F(s)\} \quad \forall s > t. \quad (2.2)$$

Comme  $U$  suit la loi uniforme, on obtient pour tout  $s > t$  :

$$F(t) = \mathbb{P}(U \leq F(t)) = \mathbb{P}(U < F(t)) \leq \mathbb{P}(G(U) \leq t) \leq \mathbb{P}(U < F(s)) = \mathbb{P}(U \leq F(s)) = F(s).$$

en utilisant respectivement la définition de la loi uniforme, le fait qu'elle n'a pas d'atome et les inclusions (2.1) et (2.2). Comme  $F$  est continue à droite  $\lim_{s \searrow t} F(s) = F(t)$  et les inégalités précédentes sont toutes des égalités. En particulier,  $\mathbb{P}(G(U) \leq t) = F(t)$  pour tout  $t \in \mathbb{R}$ , ce qui est le résultat annoncé.  $\square$

**Remarque 2.4.1** La fonction quantile  $G$  est croissante et continue à droite.

**Remarque 2.4.2 (Simulation par inversion)** Cette Proposition 2.4.1 permet de générer des lois de probabilité à partir de la loi uniforme en inversant les fonctions de répartitions. Ainsi — si  $X \sim \mathcal{E}(\lambda)$  alors  $F_X(t) = (1 - e^{-\lambda t})\mathbf{1}_{t \geq 0}$  et  $G_X(u) = \frac{1}{\lambda} \ln \frac{1}{1-u}$ ,  $u \in ]0, 1[$ . On a donc

$$\frac{1}{\lambda} \ln \frac{1}{1-U} \sim \frac{1}{\lambda} \ln \frac{1}{U} \sim \mathcal{E}(\lambda) \quad \text{quand } U \sim \mathcal{U}(]0, 1[).$$

— si  $X \sim b(p)$  alors  $F_X(t) = 0$  si  $t < 0$ ,  $(1-p)$  si  $t \in [0, 1[$  et  $1$  si  $t \geq 1$ . On a alors  $G_X(u) = 0$  si  $0 < u < 1-p$  et  $1$  si  $1-p \leq u < 1$ , ie.  $G_X(u) = \mathbf{1}_{[1-p, 1]}$ . On a donc

$$\mathbf{1}_{[1-p, 1]}(U) \sim b(p) \quad \text{quand } U \sim \mathcal{U}(]0, 1[).$$

Le résultat suivant complète la Proposition 2.4.1.

**Proposition 2.4.2** Si la loi de  $X$  est sans atome alors  $F(X) \sim \mathcal{U}(]0, 1[)$ .

**Démonstration :** Par définition de  $F$ , pour  $s > t$  dans  $]0, 1[$ , on a

$$\{F(X) < t\} \subset \{G(t) > X\} \subset \{F(X) \leq s\}.$$

Soit, par convergence monotone des probabilités, avec  $s \searrow t$  :

$$\mathbb{P}(F(X) < t) \leq \mathbb{P}(G(t) > X) \leq \mathbb{P}(F(X) \leq s).$$

Comme la variable aléatoire  $F(X)$  a un nombre dénombrable de discontinuités, on a égalité presque partout précédemment :

$$\mathbb{P}(F(X) \leq t) = \mathbb{P}(X < G(t)) \quad \text{pp.} \tag{2.3}$$

Ensuite, comme avec  $U \sim \mathcal{U}(]0, 1[)$ , on a d'après la Proposition 2.4.1  $G(U) \sim X$ , il vient  $\mathbb{P}(X < G(t)) = \mathbb{P}(G(U) < G(t))$ . Et par croissance de  $G$  :

$$\{G(U) < G(t)\} \subset \{U < t\} \subset \{U \leq t\} \subset \{G(U) \leq G(t)\}. \tag{2.4}$$

Comme  $G(U) \sim X$  est sans atome, on a  $\mathbb{P}(G(U) < G(t)) = \mathbb{P}(G(U) \leq G(t))$  et donc d'après (2.4)  $\mathbb{P}(G(U) < G(t)) = \mathbb{P}(U \leq t) = t$  et (2.3) donne

$$\mathbb{P}(F(X) \leq t) = \mathbb{P}(X < G(t)) = \mathbb{P}(G(U) < G(t)) = \mathbb{P}(U \leq t) = t \quad \text{pp.}$$

Pour conclure, on note que les fonctions  $\mathbb{P}(F(X) \leq t)$  et  $t$  sont càdlàg et coïncident presque partout, elles coïncident donc partout.  $\square$

**Proposition 2.4.3** Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F$ . On a  $\mathbb{E}[F(X)] = \frac{1}{2} + \sum_{\alpha} \frac{\Delta F(\alpha)^2}{2}$  où  $\Delta F(\alpha) = F(\alpha) - F(\alpha^-)$  avec  $F(\alpha^-) = \lim_{x \nearrow \alpha} F(x)$ .

La somme précédente est dénombrable car  $F$  n'a qu'un nombre dénombrable de discontinuité d'après la Proposition 2.3.2.

**Démonstration :** Notons que (2.4) montre que

- si  $G(t)$  n'est pas un saut de  $X \sim G(U)$  :

$$\mathbb{P}(X < G(t)) = \mathbb{P}(G(U) < G(t)) = \mathbb{P}(U \leq t) = t.$$

- Si  $G(t)$  est un saut  $\alpha$  de  $X$ , on a  $G(t) = \alpha$  et

$$\mathbb{P}(X < G(t)) = \mathbb{P}(X < \alpha) = F(\alpha^-).$$

De plus,  $G(t)$  est un saut  $\alpha$  de  $X$  si  $t \in [F(\alpha^-), F(\alpha)[$ , ie.  $t$  est sauté par  $X$ .

En général, en notant  $\alpha$  les dates de sauts de  $F$  et en utilisant pour une variable aléatoire  $Z$  positive :

$$\mathbb{E}[Z] = \int_0^{+\infty} \mathbb{P}(Z > t) dt$$

on a avec (2.3) :

$$\begin{aligned} \mathbb{E}[F(X)] &= \int_0^1 (1 - \mathbb{P}(F(X) \leq t)) dt = \int_0^1 (1 - \mathbb{P}(X < G(t))) dt \\ &= \sum_{\alpha} \int_{F(\alpha^-)}^{F(\alpha)} (1 - \mathbb{P}(X < \alpha)) dt + \int_{\bigcup_{\alpha} [F(\alpha^-), F(\alpha)[^c} (1 - \mathbb{P}(U \leq t)) dt \\ &= \sum_{\alpha} \Delta F(\alpha)(1 - F(\alpha^-)) - \sum_{\text{saut } \alpha} \int_{F(\alpha^-)}^{F(\alpha)} (1-t) dt + \int_0^1 (1-t) dt \\ &= \sum_{\alpha} \Delta F(\alpha)(1 - F(\alpha^-)) - \sum_{\text{saut } \alpha} \left( \Delta F(\alpha) - \left[ \frac{t^2}{2} \right]_{F(\alpha^-)}^{F(\alpha)} \right) + \frac{1}{2} \\ &= \frac{1}{2} + \sum_{\text{saut } \alpha} \frac{\Delta F(\alpha)^2}{2}. \end{aligned}$$

□

## 2.5 Exemples de variable aléatoire

### 2.5.1 Variables aléatoires discrètes

**Définition 2.5.1 (Variable aléatoire discrète)** Une variable aléatoire  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow B, \mathcal{B}$  est discrète si son support (tel que défini dans la Définition 2.2.3) est au plus dénombrable. Autrement dit,  $X$  est à valeurs dans un ensemble au plus dénombrable.

Notons que  $\mathcal{S}(X)$  est l'adhérence de l'ensemble des atomes de  $X$  (les atomes et leurs points d'accumulation).

En effet si on note  $\mathcal{A}$  l'ensemble de ses atomes, on a  $\mathcal{A} \subset \mathcal{S}(X)$  car si  $x \in \mathcal{A}$  n'est pas dans  $\mathcal{S}(X)$  alors  $\mathcal{S}(X) \subset B \setminus \{x\}$  serait de probabilité  $\mathbb{P}_X(B \setminus \{x\}) = 1 - \mathbb{P}(X = x) < 1$ , ce qui est absurde. Comme  $\mathcal{S}(X)$  est fermé, on a aussi  $\overline{\mathcal{A}} \subset \mathcal{S}(X)$  et il y a donc l'égalité  $\overline{\mathcal{A}} = \mathcal{S}(X)$  car  $\overline{\mathcal{A}}$  est fermé et de mesure  $\mathbb{P}_X$  égale à 1.

**Rappel (Mesure de Dirac).** La mesure de Dirac  $\delta_a$  en  $a$  est définie par

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

Si  $X$  est une variable aléatoire discrète alors sa loi est une somme de mesures de Dirac en ses atomes :

$$\mathbb{P}_X = \sum_{x \in \mathcal{S}(X)} \mathbb{P}(X = x) \delta_x.$$

En effet pour  $A \in \mathcal{B}$ , comme  $\mathcal{S}(X) = \bigcup_{x \in \mathcal{S}(X)} \{x\}$  est une union dénombrable, on a

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}(X \in A \cap \mathcal{S}(X)) = \mathbb{P}\left(X \in \bigcup_{x \in \mathcal{S}(X)} (A \cap \{x\})\right) \\ &= \mathbb{P}\left(\bigcup_{x \in \mathcal{S}(X)} \{X \in A \cap \{x\}\}\right) = \sum_{x \in \mathcal{S}(X)} \mathbb{P}(X \in A \cap \{x\}) = \sum_{x \in \mathcal{S}(X)} \mathbb{P}(X = x, x \in A) \\ &= \sum_{x \in \mathcal{S}(X)} \mathbb{P}(X = x) \delta_x(A) \end{aligned}$$

car  $\{X \in A \cap \{x\}\} = \{X = x\} \cap \{x \in A\}$ .

Le cas générique est une variable aléatoire  $X$  qui prend les valeurs  $\{x_i : i \in I\}$ ,  $i \in I$ , avec les probabilités respectives  $p_i$ ,  $i \in I$ , avec  $\sum_{i \in I} p_i = 1$  (où  $I$  est fini ou dénombrable). On a

$$\mathcal{S}(X) = \{x_i : i \in I\}, \quad \mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = p_i$$

et

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}.$$

La loi est donc donnée par l'ensemble des atomes (le support, aux points d'accumulation des atomes prés) et leur probabilité ponctuelle. Pour une variable aléatoire réelle on précise la forme de la fonction de répartition :

**Proposition 2.5.1** Soit  $X$  une variable aléatoire réelle discrète de support  $\mathcal{S}(X) = \{x_i : i \in I\}$  supposé ordonné ( $x_i < x_{i+1}$ ). La fonction de répartition  $F_X$  de  $X$  est croissante de 0 en  $-\infty$  à 1 en  $+\infty$ , constante sur chaque intervalle  $[x_k, x_{k+1}[$  avec un saut  $p_k$  en chaque atome  $x_k$ .

Il est clair que la fonction  $F_X$  détermine complètement la loi de  $X$  : les points du support sont les points de sauts de  $F_X$  et la probabilité associée est donnée par

$$p_i = F_X(x_i) - F_X(x_{i-1}).$$

Autrement dit  $\mathbb{P}_X([a, b]) = F_X(b) - \lim_{x \rightarrow a^-} F_X(x) = F_X(b) - F_X(a^-)$ . On retrouve donc la loi à partir de  $F_X$ .

**Démonstration :** D'abord  $F_X$  est à valeurs positives car une probabilité est toujours positive. Si  $s < t$ ,

$$\begin{aligned} F(t) - F(s) &= \mathbb{P}(X \leq t) - \mathbb{P}(X \leq s) \\ &= \mathbb{P}(X \leq s) + \mathbb{P}(s < X \leq t) - \mathbb{P}(X \leq s) \\ &= \mathbb{P}(s < X \leq t) = \sum_{i : s < x_i \leq t} p_i \geq 0 \end{aligned}$$

donc  $F_X$  est croissante. Puis si  $s < t$  sont dans  $[x_k, x_{k+1}]$  alors

$$F(t) - F(s) = \sum_{i : s < x_i \leq t} p_i = 0$$

car la somme est vide : il n'y a pas d'atome  $x_i$  entre  $s$  et  $t$ . S'il y en avait un, il serait *a fortiori* entre  $x_k$  et  $x_{k+1}$ , ce qui est exclu, car par l'indexation, les atomes  $x_k$  et  $x_{k+1}$  sont consécutifs.

Puis avec  $s = x_k$  et  $t = x_{k+1}$ , on a

$$F(x_{k+1}) - F(x_k) = \sum_{i : x_k < x_i \leq x_{k+1}} p_i = \sum_{i : x_i \in ]x_k, x_{k+1}]} p_i = p_{k+1}$$

car  $x_{k+1}$  est le seul atome dans  $]x_k, x_{k+1}]$ . Il y a donc un saut  $p_{k+1}$  en  $x_{k+1}$ . Enfin,

$$\lim_{t \rightarrow -\infty} F_X(t) = \lim_{t \rightarrow -\infty} \sum_{i : x_i \leq t} p_i = 0$$

car pour  $t \leq \inf_k(x_k)$ , la somme est vide donc –par convention– nulle. Et

$$\lim_{t \rightarrow +\infty} F_X(t) = \lim_{t \rightarrow +\infty} \sum_{i : x_i \leq t} p_i = \sum_i p_i = 1$$

car pour  $t \geq \sup_k(x_k)$ , la somme devient  $\sum_{\{i : x_i \in \mathbb{R}\}} p_i = 1$ . □

**Exemples :** [Variables aléatoires réelles discrètes usuelles]

- Si  $X = c$  est une variable aléatoire constante, alors sa loi est  $\mathbb{P}_X = \delta_c$  et  $\mathcal{S}(X) = \{c\}$ . En effet

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(c \in A) = \delta_c(A).$$

- Soit  $X$  une variable aléatoire qui prend les valeurs  $x_1, \dots, x_n$  avec les probabilités respectives  $p_1, \dots, p_n$  et  $p_1 + \dots + p_n = 1$ . Alors son support est  $\mathcal{S}(X) = \{x_1, \dots, x_n\}$  et sa loi est donnée par :

$$\mathbb{P}_X = p_1\delta_{x_1} + \dots + p_n\delta_{x_n}.$$

Ici  $p_i = \mathbb{P}(X = x_i)$ . C'est le cas général d'une loi discrète à support fini.

- Lorsque tous les atomes ont même probabilité on obtient une loi équirépartie :  $X$  est équirépartie sur un ensemble fini  $\{x_1, \dots, x_n\}$  si  $\mathcal{S}(X) = \{x_1, \dots, x_n\}$  et

$$\mathbb{P}_X = \sum_{i=1}^n \frac{1}{n}\delta_{x_i}.$$

Exemple : la variable aléatoire  $X$  qui indique la face  $1, \dots, 6$  obtenue par le lancer d'un dé suit la loi équirépartie (ou uniforme) sur  $\{1, \dots, 6\}$ .

- Si  $A \in \mathcal{F}$  est un événement, alors  $X = \mathbf{1}_A$  est une variable aléatoire. Elle vaut 1 si l'événement  $A$  est réalisé 0 sinon, son support est donc  $\mathcal{S}(X) = \{0, 1\}$  et sa loi :

$$\mathbb{P}_X = p\delta_1 + (1-p)\delta_0.$$

Il s'agit de la loi de Bernoulli  $b(p)$  où  $p = \mathbb{P}(A)$ .

De façon générale, une variable aléatoire  $X$  suit la loi de Bernoulli de paramètre  $p \in [0, 1]$  si elle ne prend que deux valeurs, la plupart du temps 0 et 1 avec :

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p := q.$$

Exemple : pile ou face avec  $p = 1/2$  si la pièce est équilibrée,  $p \neq 1/2$  si elle est truquée.

- Une variable aléatoire  $X$  binomiale  $\mathcal{B}(n, p)$  si  $\mathcal{S}(X) = \{0, \dots, n\}$  et

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

Exemple : la loi qui indique le nombre de succès dans une suite de  $n$  épreuves chacune ayant une probabilité  $p$  de succès est la loi  $\mathcal{B}(n, p)$ .

On rappelle que  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  est le coefficient binomial. Il s'agit bien d'une loi de probabilité car la formule du binôme de Newton (d'où le nom de la loi) donne :

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

- Une variable aléatoire  $X$  suit la loi géométrique de paramètre  $p \in ]0, 1[$  notée  $\mathcal{G}(p)$  si  $\mathcal{S}(X) = \mathbb{N}^*$  et

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

Exemple : soit  $X$  la variable aléatoire qui indique le numéro du premier lancer où on obtient un 6 lors d'une suite infinie de lancer de dé. La variable  $X$  est discrète suivant la loi géométrique  $\mathcal{G}(1/6)$ .

De façon générale, dans une suite infinie d'épreuves indépendantes avec probabilité  $p$  de succès à chacune, elle modélise le rang du premier succès.

- Une variable aléatoire  $X$  suit la loi de Poisson  $\mathcal{P}(\lambda)$  donnée par  $\mathcal{S}(X) = \mathbb{N}$  et

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

Il s'agit donc d'une variable aléatoire discrète puisque  $\mathcal{S}(X) = \mathbb{N}$ . Cette loi sert à modéliser le temps d'attente dans les files d'attente. Elle sert aussi souvent à modéliser des effectifs de petite taille.

**Proposition 2.5.2** Soit  $X$  une variable aléatoire réelle discrète dont la loi est donnée par  $\mathbb{P}(X = x_j) = p_j$ ,  $j \in J$  et soit  $g : \mathbb{R} \rightarrow \mathbb{R}$  une application mesurable (dont le domaine contient au moins  $\mathcal{S}(X)$ ). Alors  $Y = g(X)$  est une variable aléatoire réelle discrète dont la loi est donnée par le support  $\{y_i : i \in I\} = g(\mathcal{S}(X))$  et de probabilités ponctuelles

$$q_i = \mathbb{P}(Y = y_i) = \sum_{j \in J : g(x_j) = y_i} p_j.$$

**Démonstration :** Les valeurs prises par  $Y$  sont les  $g(x_j)$  pour  $x_j \in \mathcal{S}(X)$ . Puis si  $y_i$  est l'une de ces valeurs alors  $g^{-1}(y_i)$  est l'ensemble des valeurs que  $X$  doit prendre pour avoir  $Y = y_i$  par  $g$  et

$$\mathbb{P}(Y = y_i) = \mathbb{P}(Y^{-1}(y_i)) = \mathbb{P}(X^{-1}(g^{-1}(\{y_i\}))) = \mathbb{P}_X(g^{-1}(\{y_i\})) = \sum_{x_i \in g^{-1}(\{y_i\})} \mathbb{P}(X = x_i).$$

□

À partir de la loi d'un vecteur aléatoire discret, on récupère facilement la loi des marginales. Le résultat suivant l'explique pour un couple aléatoire, le cas d'un vecteur général est analogue.

**Proposition 2.5.3** Si  $(X, Y)$  est un couple aléatoire de variables aléatoires discrètes de support  $\mathcal{S}(X, Y) = \{(x_i, y_j), i \in I, j \in J\}$ , les domaines des marginales  $X, Y$  s'obtiennent par projection :

$$\mathcal{S}(X) = p_1(\mathcal{S}(X, Y)) = \{x_i, i \in I\}, \quad \mathcal{S}(Y) = p_2(\mathcal{S}(X, Y)) = \{y_i, i \in I\}$$

où  $p_1, p_2$  sont les première et seconde projections

$$p_1 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto x \end{cases}, \quad p_2 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto y \end{cases}.$$

Les lois marginales  $\mathbb{P}_X$ ,  $\mathbb{P}_Y$  (ie. les lois de  $X$  et de  $Y$ , ses marginales) sont données par :

$$\begin{aligned}\forall x_i \in \mathcal{S}(X), \quad \mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) &= \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j), \\ \forall y_i \in \mathcal{S}(Y), \quad \mathbb{P}_Y(y_j) = \mathbb{P}(Y = y_j) &= \sum_{i \in I} \mathbb{P}(X = x_i, Y = y_j).\end{aligned}$$

**Démonstration :** Il suffit de faire la preuve pour le support et les probabilités ponctuelles de  $X$ . Or pour  $i$  fixé  $\{X = x_i\}$  est la réunion de la famille dénombrable d'événements deux à deux disjoints  $\{X = x_i, Y = y_j\}$  pour tous les  $j$  tels que  $y_j \in \mathcal{S}(Y)$  car  $\{\omega \in \Omega : \mathcal{S}(Y) = y_j\}_j$  est une partition de  $\Omega$ . On conclut alors par  $\sigma$ -additivité de  $\mathbb{P}$  :

$$\begin{aligned}\mathbb{P}(X = x_i) &= \mathbb{P}\left(\{X = x_i\} \cap \bigcup_j \{Y = y_j\}\right) \\ &= \mathbb{P}\left(\bigcup_j \{X = x_i, Y = y_j\}\right) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j).\end{aligned}$$

Puis  $\{x_i : i \in I\}$  et  $\{y_j : j \in J\}$  sont bien d'une part les projections de  $\mathcal{S}(X, Y)$  sur les premier et second facteurs de  $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$  et d'autre part les domaines de  $X$  et de  $Y$ .  $\square$

**Exemples :** Pour des variables aléatoires discrètes.

On donne le tableau de  $\mathbb{P}(X = x_i, Y = y_j)$  :

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0, 1	0, 05	0, 15	0	0, 3
$x_2 = 2$	0, 05	0, 2	0, 05	0, 1	0, 4
$x_3 = 3$	0, 1	0	0, 1	0, 1	0, 3
	0, 25	0, 25	0, 3	0, 2	1

On en déduit la loi de  $X : \mathcal{S}(X) = \{0, 2, 3\}$  et

$$\mathbb{P}(X = 0) = 0, 3, \quad \mathbb{P}(X = 2) = 0, 4, \quad \mathbb{P}(X = 3) = 0, 3$$

et celle de  $Y : \mathcal{S}(Y) = \{-1, 2, 3, 5\}$  et

$$\mathbb{P}(Y = -1) = 0, 25, \quad \mathbb{P}(Y = 2) = 0, 25, \quad \mathbb{P}(Y = 3) = 0, 3, \quad \mathbb{P}(Y = 5) = 0, 2.$$

Notons qu'il n'y a pas unicité des couples  $(X, Y)$  donnant les mêmes marginales. Ainsi, le couple suivant différent du précédent partage les mêmes marginales.

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0, 1	0, 1	0	0, 1	0, 3
$x_2 = 2$	0, 1	0, 1	0, 1	0, 1	0, 4
$x_3 = 3$	0, 05	0, 05	0, 2	0	0, 3
	0, 25	0, 25	0, 3	0, 2	1

## 2.5.2 Variables aléatoires à densité

35

### Rappels de théorie de la mesure

**Définition 2.5.2 (Absolue continuité)** Soit  $(X, \mathcal{A})$  un espace mesurable muni de deux mesures  $\mu$  et  $\nu$ . On dit que  $\mu$  est absolument continue par rapport à  $\nu$  si pour tout  $A \in \mathcal{A}$ , on a

$$\nu(A) = 0 \implies \mu(A) = 0.$$

On le note  $\mu \ll \nu$ .

**Théorème 2.5.1 (Radon-Nikodym)** Soit  $\mu$  et  $\nu$  des mesures  $\sigma$ -finies. Si  $\mu \ll \nu$ , alors il existe  $f : (X, \mathcal{A}) \rightarrow \mathbb{R}$  mesurable telle que pour tout  $A \in \mathcal{A}$

$$\mu(A) = \int_A f \, d\nu.$$

La fonction  $f$  s'appelle la densité de  $\mu$  par rapport à  $\nu$ . Elle est positive si les mesures sont positives. On la note  $f = \frac{d\mu}{d\nu}$ . De plus

- Si  $\mu$  est une mesure finie alors  $f \in L^1(\nu)$  ;
- On a le lien suivant entre les intégrales par rapport à  $\nu$  et celles par rapport à  $\mu$  :

$$\int g \, d\mu = \int gf \, d\nu.$$

Les lois des variables aléatoires réelles sont des mesures sur l'espace  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Cet espace a pour mesure de référence la mesure de Lebesgue  $\lambda$ . On peut donc se demander s'il y a une relation d'absolue continuité entre la loi  $\mathbb{P}_X$  d'une variable aléatoire  $X$  et la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}$ .

Ce n'est évidemment pas toujours vrai : aucune loi discrète n'est absolument continue par rapport à  $\lambda$  puisque qu'une telle loi  $\mathbb{P}_X$  a, par définition, des atomes :

$$\mathbb{P}_X(x) = \mathbb{P}(X = x) > 0, \quad \text{alors que} \quad \lambda(\{x\}) = 0.$$

Il suffit d'avoir un atome pour qu'une loi ne soit pas absolument continue par rapport à la mesure de Lebesgue. Par définition, les lois qui sont absolument continues par rapport à la mesure de Lebesgue sont les lois à densité :

**Définition 2.5.3 (Variable aléatoire à densité)** Une variable aléatoire  $X$  est une variable aléatoire de densité  $f$  si  $\mathbb{P}_X \ll \lambda$  et

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}(X \in A) = \int_A f d\lambda.$$

On a alors  $\mathbb{P}(X \in [a, b]) = \int_a^b f(x) dx$  pour tout  $a \leq b$  réels.

Dans les calculs d'intégration, on a alors l'écriture symbolique  $d\mathbb{P}_X(x) = f(x)dx$ . Noter qu'une variable aléatoire à densité n'a pas de nom

$$\mathbb{P}(X = x) = \int_{\{x\}} f(y)dy = 0.$$

On a aussi  $\mathbb{P}(X \leq x) = \mathbb{P}(X < x)$ .

Dans ce cas, la loi est donnée par la densité. C'est une généralisation diffuse du cas discret : le support de la densité  $\{x : f(x) > 0\}$  représente l'ensemble devenu continu des atomes et la valeur de la densité  $f(x)$  en  $x$  représente la probabilité infinitésimale autour de  $x$  dans le support de la loi :

$$\mathbb{P}(X \in (x, x + dx)) \approx f(x)dx.$$

**Remarque 2.5.1** On observe que la densité  $f$  doit vérifier  $f(x) \geq 0$  et  $\int_{\mathbb{R}} f(x)dx = 1$ . Réciproquement, toute fonction  $f$  mesurable positive d'intégrale  $\int_{-\infty}^{+\infty} f(x)dx = 1$  est une densité de probabilité. Par exemple,

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad \frac{1}{b-a} \mathbf{1}_{[a,b]}, \quad \alpha e^{-\alpha x} \mathbf{1}_{\mathbb{R}^+}(x), \quad \frac{1}{\pi(1+x^2)}$$

sont les densités respectivement des lois normale standard  $\mathcal{N}(0, 1)$ , uniforme  $\mathcal{U}([a, b])$ , exponentielle  $\mathcal{E}(\alpha)$  et de Cauchy  $\mathcal{C}(1)$ . Plus généralement, la loi normale (ou de Gauss)  $\mathcal{N}(m, \sigma^2)$  est de densité

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

**Proposition 2.5.4 (Support d'une variable aléatoire à densité)** Soit  $X$  une variable aléatoire réelle de densité  $f$ . Notons  $S_f = \overline{\{x \in \mathbb{R} : \forall V(x) \text{ voisinage de } x \int_{V(x)} f(y)dy > 0\}}$  le support de  $f$ . Alors  $\mathcal{S}(X) = S_f$ .

**Démonstration :** L'ensemble  $S_f^c = \{x \in \mathbb{R} : \exists V(x) \text{ voisinage de } x \int_{V(x)} f(y)dy = 0\}^\circ$  est un ouvert de  $\mathbb{R}$  donc réunion  $\bigcup_{i \in I} ]a_i, b_i[$  de ses composantes connexes. Si  $]a_i, b_i[$  est bornée, par compacité, on a  $[a_i, b_i] \subset \bigcup_{j=1}^p V(x_j)$  et  $0 \leq \int_{a_i}^{b_i} f(y)dy \leq \sum_{j=1}^p \int_{V(x_j)} f(y)dy = 0$ . Si  $]a_i, b_i[$  n'est pas bornée on procéde de la même façon sur tout borné de  $]a_i, b_i[$  pour avoir  $\int_{a_i}^{b_i} f(y)dy = 0$ . Finalement, on montre que  $\int_{S_f^c} f(y)dy = 0$ . On a alors  $\mathbb{P}(X \in S_f^c) = \int_{S_f^c} f(y)dy = 0$  et donc  $\mathcal{S}(X) \subset S_f$  car  $S_f$  est un fermé de probabilité 1.

Réciproquement, soit  $F$  fermé avec  $\mathbb{P}(X \in F) = 1$ . Alors  $U = F^c$  est ouvert et  $0 = \mathbb{P}(X \in U) = \int_U f(y)dy$ . Ainsi si  $x \in U$  alors

$$x \in \left\{ x \in \mathbb{R} : \exists V(x) \text{ voisinage de } x \int_{V(x)} f(y)dy = 0 \right\}.$$

On en déduit  $U \subset S_f^c$  et donc  $S_f \subset F$ . Par définition de  $\mathcal{S}(X)$ , cela exige  $S_f \subset \mathcal{S}(X)$  et l'égalité.  $\square$

Pour une variable aléatoire à densité, la densité  $f$  est reliée à la fonction de répartition  $F_X$  de la façon suivante.

**Proposition 2.5.5** *Si  $X$  est une variable aléatoire réelle de densité  $f$ , sa fonction de répartition  $F_X$  vérifie :*

1.  $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt.$
2.  $F_X$  est continue sur  $\mathbb{R}$ .
3. Si  $f$  est continue au point  $x_0$ , alors  $F_X$  est dérivable en  $x_0$  de dérivée  $F'_X(x_0) = f(x_0)$ .

**Remarque 2.5.1 (Escalier de Cantor)** D'après 2) de la Proposition 2.5.5, la fonction de répartition d'une loi à densité est toujours continue. Attention, cependant toute loi avec une fonction de répartition continue n'est pas à densité. L'**escalier de Cantor** offre un célèbre contre-exemple : il s'agit d'une fonction  $f : [0, 1] \rightarrow [0, 1]$  continue, dérivable presque partout et de dérivée presque partout nulle. On la construit à partir d'un ensemble (triadique) de Cantor  $K_3$  (on rappelle que  $K_3$  est sous-ensemble compact de  $[0, 1]$  d'intérieur vide donc totalement discontinu ; il est non dénombrable mais négligeable pour la mesure de Lebesgue ; il s'écrit  $K_3 = \left\{ \sum_{n=1}^{+\infty} \frac{x_n}{3^n} : x_n \in \{0, 2\} \right\}$  : il s'agit de l'ensemble des réels dont le développement en base 3 n'admet pas de 1). La construction de l'escalier de Cantor est la suivante :

- On considère  $f_0(x) = x$  ;
- La fonction  $f_1$  est la fonction continue affine par morceaux qui vaut 0 en 0, 1 en 1 et  $1/2$  sur  $[1/3, 2/3]$  ;
- On passe de  $f_n$  à  $f_{n+1}$  comme suit : sur chaque  $[u, v]$  où  $f_n$  n'est pas constante, on remplace  $f_n$  par la fonction linéaire par morceaux qui vaut  $\frac{f_n(u)+f_n(v)}{2}$  sur  $[2u/3 + v/3, 2v/3 + u/3]$  ;
- On vérifie que  $|f_{n+1}(x) - f_n(x)| \leq 2^{-n}$ , ce qui justifie que la série  $\sum_{n \geq 1} (f_{n+1} - f_n)$  converge uniformément et donc la suite  $(f_n)_n$  aussi.
- La limite  $f$  est donc continue, croissante avec  $f(0) = 0$ ,  $f(1) = 1$ . De plus  $f$  a une dérivée nulle sur le complémentaire du Cantor  $K_3$  puisque celui-ci est une réunion d'intervalle sur lesquels par construction  $f$  est constante.

En fait les fonctions de répartition des lois à densité sont plus que continues, elles sont absolument continues. On veillera donc à ce qu'une loi sans atome n'est pas nécessairement à densité.

**Démonstration :** Puisque  $X$  a pour densité  $f$ , et comme

$$F_X(b) = \mathbb{P}(X \in ]-\infty, b]) = \mathbb{P}(X \in ]-\infty, a] \cup ]a, b]) = F_X(a) + \mathbb{P}(X \in ]a, b]),$$

on a pour tous réels  $a < b$  :

38

$$F_X(b) - F_X(a) = \mathbb{P}(X \in ]a, b]) = \int_a^b f(t) dt. \quad (2.5)$$

1) Il suffit d'appliquer la monotonie séquentielle des probabilités avec  $b = x$  fixé et  $a = -n$  pour chaque  $n \in \mathbb{N}$  tel que  $n > -x$ . La suite d'événements

$$A_n = \{\omega : X(\omega) \in ]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et de réunion  $A = \{\omega : X(\omega) \in ]-\infty, x]\} = \{X \leq x\}$ . Par la propriété de continuité monotone séquentielle (ou par convergence dominée), on a  $\mathbb{P}(A_n) \nearrow \mathbb{P}(A)$ , d'où

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \int_{-n}^x f(t) dt = \int_{-\infty}^x f(t) dt.$$

2) On fixe  $x_0 \in \mathbb{R}$  quelconque. D'abord  $F_X$  est continue à droite en tout point car c'est une fonction de répartition.

Il reste à voir la continuité à gauche. Par monotonie, il suffit de vérifier, pour une  $x_n < x_0$  une suite croissante qui converge vers  $x_0$  :

$$\lim_{n \rightarrow +\infty} F_X(x_n) = F_X(x_0).$$

On a

$$F_X(x_0) - F_X(x_n) = \int_{x_n}^{x_0} f(t) dt = \int f(t) \mathbf{1}_{[x_n, x_0]}(t) dt.$$

Or  $|f(t) \mathbf{1}_{[x_n, x_0]}(t)| \leq f(t)$ , intégrable, puisque  $f$  est une densité, puis pour presque chaque  $t \in \mathbb{R}$ ,  $f(t) \mathbf{1}_{[x_n, x_0]}(t) \rightarrow 0$  puisque  $\lim_{n \rightarrow +\infty} \mathbf{1}_{[x_n, x_0]}(t) = \mathbf{1}_{[t_0, t_0]}(t)$ .

Le théorème de convergence dominée de Lebesgue s'applique et donne

$$\lim_{n \rightarrow +\infty} F_X(x_0) - F_X(x_n) = \int_{\mathbb{R}} 0 dt = 0,$$

ce qui est le résultat souhaité.

3) Comme par hypothèse  $f$  est continue en  $x_0$ , elle est définie sur tout un voisinage de  $x_0$  et donc sur un intervalle  $[a, b]$  qui contient  $x_0$ . La continuité de  $f$  en  $x_0$  s'écrit :  $\forall \varepsilon > 0$ ,  $\exists \delta > 0$  tel que  $]x_0 - \delta, x_0 + \delta[ \subset ]a, b[$  et

$$\forall t \in ]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon.$$

Pour tout  $h$  tel que  $0 < |h| < \delta$ , on a alors  $F_X(x_0 + h) - F_X(x_0) = \int_{x_0}^{x_0+h} f(t) dt$ . D'où

$$|F_X(x_0 + h) - F_X(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} (f(t) - f(x_0)) dt \right|$$

$$\leq \left| \int_{x_0}^{x_0+h} |(f(t) - f(x_0))| dt \right| \leq |h|\varepsilon.$$

En divisant par  $h$  puis en faisant  $h \rightarrow 0$ , on constate que  $F_X$  est dérivable en  $x_0$ , de dérivée  $F'_X(x_0) = f(x_0)$ .  $\square$

Réiproquement à la Proposition 2.5.5, on peut déterminer si une variable aléatoire  $X$  a une densité en fonction de sa fonction de répartition  $X$ .

**Proposition 2.5.6** *Si une variable aléatoire  $X$  admet pour une fonction de répartition  $F$  de la forme*

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2.6)$$

où  $f$  est mesurable positive alors  $X$  admet nécessairement  $f$  pour densité.

Une fonction satisfaisant (2.6) est dite absolument continue. En fait, la loi de  $X$  est absolument continue (en tant que mesure) si et seulement si sa fonction de répartition est absolument continue (en tant que fonction).

**Démonstration :** D'abord comme  $F$  est une fonction de répartition, on montre facilement que  $f$  est une densité puisque  $f \geq 0$  et

$$\int_{-\infty}^{+\infty} f(x) dx = \lim_{x \rightarrow +\infty} F(x) = 1.$$

Si on note  $\mu$  la mesure définie par

$$\mu(A) = \int_A f(x) dx, \quad A \in \mathcal{B}(\mathbb{R})$$

alors  $\mu$  est une mesure de probabilité de densité  $f$  et d'après la forme de  $F$  en (2.6), les mesures  $\mathbb{P}_X$  et  $\mu$  coïncident sur la famille  $\mathcal{C} = \{]-\infty, x] : x \in \mathbb{R}\}$ . Comme cette famille engendre  $\mathcal{B}(\mathbb{R})$  et est un  $\pi$ -système (stable par intersection), par un argument de classe monotone,  $\mathbb{P}_X = \mu$  sur  $\mathcal{B}(\mathbb{R})$ . Finalement,  $\mathbb{P}_X$  admet bien  $f$  pour densité.  $\square$

### Cas des vecteurs aléatoires à densité

**Définition 2.5.4 (Densité en dimension  $d$ )** *Une fonction  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  mesurable est appelée densité de probabilité (en dimension  $d$ ) si*

- $f$  est positive : en tout point où elle est définie,  $f(t_1, \dots, t_d) \geq 0$ ,
- $f$  est intégrable sur  $\mathbb{R}^d$  d'intégrale 1 :

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(t_1, \dots, t_d) dt_1 \dots dt_d = 1.$$

**Définition 2.5.5 (Vecteur aléatoire)** *Le vecteur aléatoire  $(X_1, \dots, X_d)$  suit la loi de densité  $f$  si pour tous les intervalles  $[a_i, b_i]$ ,  $i = 1, \dots, d$*

$$\mathbb{P}\left((X_1, \dots, X_d) \in [a_1, b_1] \times \dots \times [a_d, b_d]\right) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_d}^{b_d} f(t_1, t_2, \dots, t_d) dt_1 \dots dt_d.$$

À nouveau, le sens des bornes dans les intervalles (ouvertes ou fermées) n'est pas important. À nouveau encore, la densité caractérise la loi : si  $(Y_1, \dots, Y_d)$  a même loi que  $(X_1, \dots, X_d)$  alors ce vecteur a la même densité et réciproquement.

**Proposition 2.5.7** *Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{R}^d$  de densité  $f_X$  et soit  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  un difféomorphisme de  $\mathbb{R}^d$  (bijection continûment différentiable ainsi que son inverse). Alors  $Y = g(X)$  est une variable aléatoire à valeurs dans  $\mathbb{R}^d$  qui admet pour densité la fonction*

$$f_Y(y) = f_X(g^{-1}(y)) |Jac g^{-1}(y)|, \quad y \in \mathbb{R}^d$$

où  $Jac g(x) = \det\left(\frac{\partial g_i}{\partial x_j} : 1 \leq i, j \leq d\right)$  est le jacobien de  $g$ .

**Démonstration :** On calcule pour  $B \in \mathcal{B}(\mathbb{R}^d)$

$$\mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \int_{\mathbb{R}^d} \mathbf{1}_{\{x \in g^{-1}(B)\}} f_X(x) dx = \int_{\mathbb{R}^d} \mathbf{1}_B(g(x)) f_X(x) dx.$$

On pose  $y = g(x) \Leftrightarrow x = g^{-1}(y)$  et on applique le théorème de changement de variable

$$\mathbb{P}(g(X) \in B) = \int_{\mathbb{R}^d} \mathbf{1}_B(y) f_X(g^{-1}(y)) |Jac g^{-1}(y)| dy.$$

□

**Proposition 2.5.8 (Densité marginale)** *Si  $(X, Y)$  est un couple aléatoire de loi de densité  $f$ , ses lois marginales  $\mathbb{P}_X$ ,  $\mathbb{P}_Y$  sont données par :*

$$\begin{aligned} \forall A \in \mathcal{B}(\mathbb{R}), \mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \int_a^b \int_{-\infty}^{+\infty} f(x, y) dy dx, \\ \forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}_Y(B) &= \mathbb{P}(Y \in B) = \int_a^b \int_{-\infty}^{+\infty} f(x, y) dx dy. \end{aligned}$$

Autrement dit, la loi de  $X$  est de densité  $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ , celle de  $Y$  est de densité  $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$ .

**Démonstration :** La preuve est une application directe du théorème de Fubini-Tonelli sur les intégrales doubles une fois qu'on a remarqué que

$$\begin{aligned}\mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \mathbb{P}_{(X,Y)}(A \times \mathbb{R}) \\ &= \int_{A \times \mathbb{R}} f(x, y) dx dy = \int_A \left( \int_{\mathbb{R}} f(x, y) dy \right) dx = \int_A f_X(x) dx\end{aligned}$$

avec la densité annoncée  $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$ . Il s'applique sans problème car par définition d'une densité,  $f$  est positive (et même intégrable sur  $\mathbb{R}^2$ ). Idem pour  $Y$ .  $\square$

**Exemples :** • Considérons  $f(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$ . Il s'agit bien d'une densité car  $f$  est positive et

$$\begin{aligned}\int \int_{\mathbb{R}^2} f(x, y) dx dy &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) dx dy \\ &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) dx dy \\ &= \underbrace{\frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) dx}_{=1} \underbrace{\int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) dy}_{=2 - (-1) = 3} = 1.\end{aligned}$$

Considérons un couple  $(X, Y)$  de loi de densité  $f$ . La loi de  $X$  est alors de densité donnée par :

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) dy \\ &= \mathbf{1}_{[0,1]}(x) \times \underbrace{\frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) dy}_{=1} = \mathbf{1}_{[0,1]}(x).\end{aligned}$$

De la même façon,  $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$ .

• Montrer que  $f(x, y) = \lambda \mu e^{-\lambda x - \mu y} \mathbf{1}_{\mathbb{R}^+ \times \mathbb{R}^+}(x, y)$  est la densité d'un couple  $(X, Y)$  de  $\mathbb{R}^2$ . Montrer que  $X$  est de loi  $\mathcal{E}(\lambda)$  et  $Y$  de loi  $\mathcal{E}(\mu)$ .

• Montrer que

$$f(x, y) = \frac{e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2(\sigma')^2}}}{2\pi\sigma\sigma'}$$

est la densité d'un couple  $(X, Y)$  de  $\mathbb{R}^2$ . Montrer que  $X$  est de loi  $\mathcal{N}(0, \sigma^2)$  et  $Y$  de loi  $\mathcal{N}(0, (\sigma')^2)$ .

• Montrer que

$$f(x, y) = ye^{-xy} \mathbf{1}_{\mathbb{R}^+}(x) \mathbf{1}_{[0,1]}(y)$$

est la densité d'un couple  $(X, Y)$  de  $\mathbb{R}^2$ . Montrer que  $X$  est de loi donnée par la densité

$$f_X(x) = \frac{1 - e^{-x} - xe^{-x}}{x^2} \mathbf{1}_{\mathbb{R}^+}(x)$$

et  $Y$  de loi uniforme sur  $[0, 1]$ .

- Soit  $f(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$ . Il s'agit d'une densité car

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) dx dy &= \int \int_{\mathbb{R}^2} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx dy}{3\pi} \\ &= \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx dy}{3\pi} = \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2}{2\times 3}} e^{-\frac{4y^2}{2\times 3}} \frac{dx dy}{3\pi} \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2\times 3}} dx \right) e^{-\frac{4y^2}{2\times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{-\frac{z^2}{2\times 3}} dz \right) e^{-\frac{4y^2}{2\times 3}} \frac{dy}{3\pi} \\ &= \int_{\mathbb{R}} \sqrt{2\pi \times 3} e^{-\frac{4y^2}{2\times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{y^2}{2\times (3/4)}} \frac{dy}{\sqrt{2\pi \times 3/4}} = 1 \end{aligned}$$

en utilisant la normalisation de la loi normale  $\mathcal{N}(0, \sigma^2)$  :  $\int_{\mathbb{R}} e^{-\frac{t^2}{2\sigma^2}} dt = \sqrt{2\pi\sigma^2}$ .

Considérons un couple  $(X, Y)$  de densité  $f$ , alors  $X$  est de densité

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x, y) dy = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x+\sqrt{5}y)^2+4x^2}{6}} \frac{dy}{3\pi} \\ &= \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x+\sqrt{5}y)^2}{6}} e^{-\frac{4x^2}{30}} \frac{dy}{3\pi} = e^{-\frac{4x^2}{30}} \int_{\mathbb{R}} e^{-\frac{z^2}{2\times 3}} \frac{dz}{3\pi\sqrt{5}} = e^{-\frac{4x^2}{30}} \frac{\sqrt{2\pi \times 3}}{3\pi\sqrt{5}} \\ &= \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}. \end{aligned}$$

La marginale  $Y$  est de densité :

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f(x, y) dx = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2\times 3}} e^{-\frac{4y^2}{6}} \frac{dx}{3\pi} \\ &= e^{-\frac{4y^2}{6}} \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2\times 3}} \frac{dx}{3\pi} = e^{-\frac{4y^2}{6}} \frac{\sqrt{2\pi \times 3}}{3\pi} = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}. \end{aligned}$$

Les marginales  $X$  et  $Y$  sont donc de lois  $\mathcal{N}(0, 15/4)$  et  $\mathcal{N}(0, 3/4)$ .

### Généralisation au cas gaussien

Soit  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$  une forme quadratique positive. Notons

$$C = \int_{\mathbb{R}^n} e^{-\Phi(x_1, \dots, x_n)} dx_1 \dots dx_n.$$

Alors tout vecteur aléatoire de densité

43

$$f(x_1, \dots, x_n) = \frac{1}{C} e^{-\Phi(x_1, \dots, x_n)}$$

est dit gaussien. Ses marginales sont des variables aléatoires gaussiennes en dimension 1. On traitera ce type de vecteur en détail dans le Chapitre 9.

### 2.5.3 Lois de probabilité usuelles

Dans cette section, nous présentons les principales lois de probabilités usuelles. Nous donnons les atomes  $x_i$  et leur probabilité ponctuelle  $p_i$  dans le cas discret ou la densité avec son support dans le cas à densité. En anticipant sur les Chapitres 3 et 4, nous donnons également pour chaque loi son espérance, sa variance et sa fonction caractéristique. Pour chaque loi, on décrit son interprétation pour un modèle simple.

**1. Loi de Bernoulli**  $b(p)$ ,  $p \in [0, 1]$  :  $P = (1 - p)\delta_0 + p\delta_1$ .

Une variable aléatoire  $X$  suit la loi de Bernoulli  $b(p)$  si

$$\mathbb{P}(X = 0) = 1 - p, \quad \mathbb{P}(X = 1) = p.$$

Si  $p = 0$  (respectivement  $p = 1$ )  $X$  est presque sûrement égale à 0 (respectivement 1). Cette loi modélise une expérience à deux issues possibles (succès et échec) dont la probabilité de succès vaut  $p$ . Le résultat  $X$  de cette expérience suit la loi de Bernoulli de paramètre  $p$ .

$$\mathbb{E}[X] = p, \quad \text{Var}(X) = p(1 - p), \quad \varphi_X(t) = 1 - p + pe^{it}.$$

**2. Loi binomiale**  $\mathcal{B}(n, p)$ ,  $n \in \mathbb{N}^*$ ,  $p \in ]0, 1[$  :  $P = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k$ .

Une variable aléatoire  $X$  suit la loi binomiale  $\mathcal{B}(n, p)$  si

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{pour } k = 0, \dots, n$$

et  $\mathbb{P}(X = k) = 0$  sinon. Cette loi modélise le nombre de succès obtenus au cours de  $n$  expériences indépendantes, chacune ayant une probabilité  $p$  de succès. Le nombre  $X$  de succès obtenus suit la loi binomiale.

$$\mathbb{E}[X] = np, \quad \text{Var}(X) = np(1 - p), \quad \varphi_X(t) = (1 - p + pe^{it})^n.$$

**3. Loi uniforme discrète (ou équirépartie)**  $\mathcal{U}(x_1, \dots, x_r)$ ,  $r \in \mathbb{N}^*$  :  $P = \frac{1}{r} \sum_{j=0}^r \delta_{x_j}$ .

Une variable aléatoire  $X$  suit la loi uniforme  $\mathcal{U}(x_1, \dots, x_r)$  si

$$\mathbb{P}(X = x_j) = \frac{1}{r}, \quad j = 1, \dots, r.$$

Cette loi modélise une expérience à  $r$  issues possibles, chacune ayant la même probabilité  $1/r$ . Le résultat  $X$  de cette expérience suit la loi uniforme (ou équirépartie).

$$\mathbb{E}[X] = \frac{x_1 + \cdots + x_r}{r}$$

Lorsque la loi est  $\mathcal{U}(1, \dots, r)$ , on peut préciser espérance, variance et fonction caractéristique

$$\mathbb{E}[X] = \frac{r+1}{2}, \quad \text{Var}(X) = \frac{r^2 - 1}{12}, \quad \varphi_X(t) = \frac{1}{r} e^{it} \frac{1 - e^{irt}}{1 - e^{it}}.$$

**4. Loi géométrique**  $\mathcal{G}(p)$ ,  $p \in ]0, 1[$  :  $P = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k$ .

Une variable aléatoire  $X$  suit la loi géométrique  $\mathcal{G}(p)$  si

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

Cette loi modélise le rang du premier succès dans une suite d'expériences indépendantes, chacune ayant la probabilité  $p$  de succès. Ce rang  $X$  suit la loi géométrique. Par exemple  $X = 4$  signifie que les trois premières expériences ont échoué et la 4ème réussi.

$$\mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}, \quad \varphi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}.$$

**5. Loi de Poisson**  $\mathcal{P}(\lambda)$ ,  $\lambda > 0$  :  $Q = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k$ .

Une variable aléatoire  $X$  suit la loi de Poisson  $\mathcal{P}(\lambda)$  si

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

Cette loi intervient pour modéliser les occurrences d'événements peu probables.

$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda, \quad \varphi_X(t) = \exp(\lambda(e^{it} - 1)).$$

**6. Loi hypergéométrique\***  $\mathcal{H}(N, n, m)$ ,  $N \in \mathbb{N}^*$ ,  $1 \leq n, m \leq N$ ,  $p \in ]0, 1[$ .

Une variable aléatoire  $X$  suit la loi hypergéométrique  $\mathcal{H}(N, n, m)$  si

$$\mathbb{P}(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \quad \text{pour } \max\{0, n - (N - m)\} \leq k \leq \min\{n, m\}$$

et = 0 sinon. Cette loi modélise le tirage en une seule fois et sans remise de  $n$  boules dans une urne contenant  $N$  boules dont  $m$  sont blanches et  $N - m$  sont rouges. Le nombre  $X$  de boules blanches tirées suit la loi hypergéométrique.

$$\mathbb{E}[X] = n \frac{m}{N}, \quad \text{Var}(X) = \frac{N-n}{N-1} n \frac{m}{N} \frac{N-m}{N}.$$

**7. Loi multinomiale\***  $\mathcal{M}(n, p_1, \dots, p_n)$ ,  $n \in \mathbb{N}$ ,  $p_1 + \dots + p_d = 1$ ,  $p_1, \dots, p_d \in [0, 1]$ .  
Un vecteur aléatoire  $(X_1, \dots, X_d)$  à valeurs dans  $\mathbb{N}^d$ , suit la loi multinomiale  $\mathcal{M}(n, p_1, \dots, p_n)$  si

$$\mathbb{P}((X_1, \dots, X_d) = (n_1, \dots, n_d)) = \frac{n!}{n_1! \dots n_d!} p_1^{n_1} \dots p_d^{n_d} \quad \text{pour } n_1 + \dots + n_d = n$$

et = 0 sinon. Si on dispose de  $n$  boules que l'on jette une par une aléatoirement dans  $d$  boîtes différentes, chaque boule ayant la probabilité  $p_i$  d'être jetée dans la  $i$ -ème boîte, les nombres  $(X_1, \dots, X_d)$  de boules dans les boîtes suivent la loi multinomiale.

$$\begin{aligned} \mathbb{E}[(X_1, \dots, X_d)] &= (np_1, \dots, np_d), \quad \text{Var}(X_j) = np_j(1 - p_j), \quad \text{Cov}(X_j, X_k) = -np_j p_k, \\ \varphi_X(t_1, \dots, t_d) &= \left( \sum_{j=1}^d p_j e^{it_j p_j} \right)^n. \end{aligned}$$

### 8. Loi uniforme $\mathcal{U}([a, b])$ , $-\infty < a \leq b < \infty$ .

Une variable aléatoire réelle  $X$  suit une loi uniforme  $\mathcal{U}([a, b])$  si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

Cette loi modélise des phénomènes dont les valeurs sont uniformément réparties sur un intervalle.

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12}, \quad \varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

### 9. Loi exponentielle $\mathcal{E}(\lambda)$ , $\lambda > 0$ .

Une variable aléatoire réelle  $X$  suit une loi exponentielle  $\mathcal{E}(\lambda)$  si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty[}(x).$$

Cette loi modélise des temps d'attente ou des durées de vie.

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}, \quad \varphi_X(t) = \frac{\lambda}{\lambda - it}.$$

On montre que la loi exponentielle est caractérisée par la propriété d'absence de mémoire :

**Proposition 2.5.9** *Une variable aléatoire  $X$  positive est exponentielle si elle vérifie la propriété d'absence de mémoire :*

$$\mathbb{P}(X \geq t+s | X \geq s) = \mathbb{P}(X > t), \quad \text{pour tout } t, s \geq 0$$

**Démonstration :** D'abord si  $X \sim \mathcal{E}(\lambda)$  alors

$$\mathbb{P}(X > t+s | X > s) = \frac{\mathbb{P}(X > t+s, X > s)}{\mathbb{P}(X > s)} = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}$$

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

Ensuite, si on note  $f(t) = \mathbb{P}(X > t)$  alors la propriété d'absence de mémoire s'écrit

$$f(t+s) = f(t)f(s). \quad (2.7)$$

De plus,  $f(0) = \mathbb{P}(X \geq 0) = 1$ . Comme  $f(t) \in [0, 1]$ , la seule solution de ce type d'équation fonctionnelle (2.7) est  $f(t) = e^{-\lambda t}$ . En effet on a  $f(n) = f(n-1)f(1) = f(1)^n$ ; puis  $f(1) = f(\frac{n}{n}) = f(\frac{1}{n})^n$ , soit  $f(\frac{1}{n}) = f(1)^{1/n}$ . On en déduit que  $f(\frac{p}{q}) = f(1)^{p/q} = e^{-\lambda \frac{p}{q}}$  en écrivant  $f(1) = e^{-\lambda}$ . On a donc  $f(x) = e^{-\lambda x}$  pour tout  $x \in \mathbb{Q}$ . En effet, s'il existe  $x \in \mathbb{R}_+$  tel que  $f(x) \neq e^{-\lambda x}$  alors par exemple  $f(x) > e^{-\lambda x}$  et on peut choisir  $r \in \mathbb{Q}$  tel que  $f(x) > e^{-\lambda r} > e^{-\lambda x}$ , ce qui exige  $r < x$ ; mais alors par décroissance de  $f$ , on a

$$f(x) \leq f(r) = e^{-\lambda r} < f(x),$$

ce qui est aburde. Finalement,  $f(x) = e^{-\lambda x}$  pour tout  $x \in \mathbb{R}$ , ie  $\mathbb{P}(X \geq t) = e^{-\lambda t}$  et  $X \sim \mathcal{E}(\lambda)$ .  $\square$

### 10. Loi gamma\* $\gamma(p, \lambda)$ , $p > 0$ et $\lambda > 0$ .

Une variable aléatoire réelle  $X$  suit une loi gamma  $\gamma(p, \lambda)$  si sa densité par rapport à la mesure de Lebesgue est

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

où  $\Gamma(p) := \int_0^{+\infty} x^{p-1} e^{-x} dx$  est la fonction d'Euler de seconde espéce.

$$\mathbb{E}[X] = \frac{p}{\lambda}, \quad \text{Var}(X) = \frac{p}{\lambda^2}, \quad \varphi_X(t) = \left( \frac{\lambda}{\lambda - it} \right)^p.$$

### 11. Loi du chi-deux\* $\chi^2(n)$ , $n \in \mathbb{N}^*$ .

Une variable aléatoire réelle  $X$  suit une loi de chi-deux  $\chi^2(n)$  si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{]0, +\infty[}(x).$$

$$\mathbb{E}[X] = n, \quad \text{Var}(X) = 2n, \quad \varphi_X(t) = \frac{1}{(1 - 2it)^{n/2}}.$$

On peut noter que  $\chi^2(n) = \gamma(\frac{n}{2}, \frac{1}{2})$ .

### 12. Loi beta\* $\beta(p, q)$ , $p > 0, q > 0$ .

Une variable aléatoire réelle  $X$  suit une loi beta  $\beta(p, q)$  si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)} \mathbf{1}_{[0,1]}(x).$$

où  $B(p, q) := \int_0^1 x^{p-1}(1-x)^{q-1}dx$  est la fonction d'Euler de première espéce.

$$\mathbb{E}[X] = \frac{B(p+1, q)}{B(p, q)}, \quad \text{Var}(X) = \frac{pq}{(p+q)^2(p+q+1)}.$$

### 13. Loi de Laplace\*.

Une variable aléatoire réelle  $X$  suit une loi de Laplace si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{2}e^{-|x|}.$$

$$\mathbb{E}[X] = 0, \quad \text{Var}(X) = 1, \quad \varphi_X(t) = \frac{1}{1+t^2}.$$

### 14. Loi de Cauchy $\mathcal{C}(a)$ , $a > 0$ .

Une variable aléatoire réelle  $X$  suit une loi de Cauchy  $\mathcal{C}(a)$  si sa densité par rapport à la mesure de Lebesgue est

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

$$\text{pas d'espérance, pas de variance finie, } \varphi_X(t) = e^{-a|t|}.$$

### 15. Loi de Gauss (ou normale) $\mathcal{N}(m, \sigma^2)$ , $m \in \mathbb{R}$ , $\sigma^2 > 0$ .

Une variable aléatoire réelle  $X$  suit une loi de Gauss (ou normale)  $\mathcal{N}(m, \sigma^2)$  si sa densité par rapport à la mesure de Lebesgue est

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

$$\mathbb{E}[X] = m, \quad \text{Var}(X) = \sigma^2, \quad \varphi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

### 16. Loi de Gauss (ou normale) multidimensionnelle $\mathcal{N}(m, \Sigma)$ , $m \in \mathbb{R}^d$ , $\Sigma$ matrice symétrique définie positive.

Un vecteur aléatoire  $(X_1, \dots, X_d)$  à valeurs dans  $\mathbb{R}^d$  suit une loi de Gauss (ou normale)  $\mathcal{N}(m, \Sigma)$  si sa densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  est

$$f(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-m)^* \Sigma^{-1} (x-m)\right), \quad x \in \mathbb{R}^d.$$

$$\mathbb{E}[(X_1, \dots, X_d)] = m, \quad (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d} = \Sigma, \quad \varphi_X(t_1, \dots, t_d) = \exp\left(i\langle m, t \rangle - \frac{1}{2}t^* \Sigma t\right).$$

Plus de détails sont donnés sur ces lois multidimensionnelles au Chapitre 9.

# Chapitre 3

## Espérance d'une variable aléatoire

### 3.1 Rappels d'intégration

On commence dans cette section par des rappels d'intégration en théorie de la mesure. On renvoie à [JCB-Leb] pour des détails ou des preuves.

**Définition 3.1.1 (Espace produit)** Soient  $(X, \mathcal{A}, \mu)$  et  $(Y, \mathcal{B}, \nu)$  deux espaces mesurés.

- Sur l'espace produit  $X \times Y = \{(x, y) : x \in X, y \in Y\}$ , on appelle tribu produit  $\mathcal{A} \otimes \mathcal{B}$  la tribu de  $X \times Y$  engendrée par les produits  $A \times B$  pour  $A \in \mathcal{A}$ ,  $B \in \mathcal{B}$ . Il s'agit de la tribu la plus naturelle sur  $X \times Y$ .
- Sur l'espace produit  $X \times Y$  muni de la tribu produit  $\mathcal{A} \otimes \mathcal{B}$ , on définit la mesure produit  $\mu \otimes \nu$  de la façon suivante

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

En définissant  $\mu \otimes \nu$  sur la classe des pavés  $A \times B$ ,  $A \in \mathcal{A}$ ,  $B \in \mathcal{B}$ , comme elle engendre la tribu produit  $\mathcal{A} \otimes \mathcal{B}$ , par un argument de classe monotone,  $\mu \otimes \nu$  est bien définie sur tout  $\mathcal{A} \otimes \mathcal{B}$ .

**Exemple.** Sur  $\mathbb{R}^n$ , la tribu borélienne de  $\mathbb{R}^n$  coïncide avec la tribu produit des tribus boréliennes des espaces facteurs  $\mathbb{R}$  (cf. Remarque 1.1.1) :

$$\mathcal{B}(\mathbb{R}^n) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})}_{n \text{ fois}}.$$

Un borélien typique de  $\mathbb{R}^n$  est le pavé  $[a_1, b_1] \times \cdots \times [a_n, b_n]$ .

En général en dimension  $n$ , on utilise la mesure de Lebesgue de dimension  $n$ , il s'agit de la mesure produit des mesures de Lebesgue sur  $\mathbb{R}$  :  $\lambda_n = \underbrace{\lambda \otimes \cdots \otimes \lambda}_{n \text{ fois}}$ . Par exemple,

$$\lambda_n([a_1, b_1] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \times \cdots \times (b_n - a_n).$$

### 3.1.1 Théorèmes de Fubini

49

Sous de bonnes conditions, le théorème de Fubini permet de permuter les intégrations dans des intégrales multiples. De cette façon, les intégrales multiples, ou par rapport à des mesures produits, se ramènent à des intégrales simples emboitées.

Dans la suite, on autorise les valeurs infinies et les fonctions positives sont à valeurs dans  $[0, +\infty] = [0, +\infty] \cup \{+\infty\}$ . On métrise facilement l'ensemble  $[0, +\infty]$  et on considère la tribu borélienne associée.

**Théorème 3.1.1 (Fubini-Tonelli)** Soit  $f : (X \times Y, \mathcal{A} \otimes \mathcal{B}) \rightarrow [0, +\infty]$  une fonction mesurable, avec  $\mu$  et  $\nu$  des mesures  **$\sigma$ -finies** sur  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$ . Alors

1. La fonction  $x \mapsto \int_Y f(x, \cdot) d\nu$  est  $\mathcal{A}$ -mesurable et la fonction  $y \mapsto \int_X f(\cdot, y) d\mu$  est  $\mathcal{B}$ -mesurable.
2. Puis, on a

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X \left( \int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_Y \left( \int_X f(x, y) d\mu(x) \right) d\nu(y).$$

**Démonstration :** Cf. notes de cours d' "Intégration et mesures" [JCB-Leb]. □

Le théorème de Fubini-Tonelli ne s'applique qu'à des fonctions mesurables positives (sans conditions supplémentaires). Pour des fonctions de signes quelconques intégrables, on a le résultat suivant :

**Théorème 3.1.2 (Fubini)** Soient  $(X, \mathcal{A}, \mu)$  et  $(Y, \mathcal{B}, \nu)$  deux espaces  **$\sigma$ -finis** et  $f : (X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu) \rightarrow \mathbb{R}$  ou  $\mathbb{C}$  une fonction  $(\mu \otimes \nu)$ -intégrable, c'est à dire

$$\int_{X \times Y} |f| d(\mu \otimes \nu) < +\infty.$$

Alors

1. Pour  $\mu$ -presque chaque  $x$ , la fonction  $f(x, \cdot)$  est  $\nu$ -intégrable et pour  $\nu$ -presque chaque  $y$ , la fonction  $f(\cdot, y)$  est  $\mu$ -intégrable.

Posons  $I(x) = \int_Y f(x, \cdot) d\nu$  et  $J(y) = \int_X f(\cdot, y) d\mu$ , alors

2. Les fonctions  $I$  et  $J$  sont respectivement  $\mu$ - et  $\nu$ -intégrables.
3. On a

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X I d\mu = \int_Y J d\nu.$$

On a donc en écrivant les variables d'intégration

$$\int_{X \times Y} f(x, y) d(\mu \otimes \nu)(x, y) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y).$$

**Démonstration :** Cf. notes de cours d' "Intégration et mesures" [JCB-Leb]. □

**Remarque 3.1.1** • En pratique pour utiliser le théorème de Fubini, on raisonne de la façon suivante :

1. On montre que  $f$  est mesurable (arguments généraux).
2. Pour montrer que  $f$  est intégrable, on calcule  $\int_{X \times Y} |f| d(\mu \otimes \nu)$  en appliquant le théorème de Fubini-Tonelli à la fonction positive  $|f|$  :

$$\int |f| d(\mu \otimes \nu) = \int_Y \left( \int_X |f(x, y)| d\mu(x) \right) d\nu(y) = \int_X \left( \int_Y |f(x, y)| d\nu(y) \right) d\mu(x)$$

en choisissant la forme la plus convenable (intégrer d'abord en  $x$  ou en  $y$ ) pour faire le calcul.

3. Si le résultat est fini, on peut alors appliquer le théorème de Fubini.
- Si  $F$  est positive, on peut intervertir directement les intégrations (par la version Fubini-Tonelli du résultat). Si  $f$  ne l'est pas, il faut vérifier l'intégrabilité en calculant l'intégrale de  $|f|$  en appliquant par exemple la version Fubini-Tonelli à  $|f| > 0$  pour se ramener à des intégrales simples.
  - L'utilisation du théorème de Fubini permet de ramener de nombreux calculs d'intégrales doubles (ou triples ou plus généralement multiples) à des calculs successifs d'intégrales simples (aussi bien pour des calculs effectifs que pour montrer des convergences d'intégrales).

### 3.1.2 Changement de variable

#### Formule de transfert

Soient  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$  deux espaces mesurables et  $\varphi : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  une fonction mesurable.

Si on considère une mesure  $\mu$  sur  $(X, \mathcal{A})$  alors la mesure image  $\nu = \mu_\varphi$  est une mesure sur  $(Y, \mathcal{B})$ . On a un lien entre les intégrales par rapport à  $\mu$  sur  $X$  et celle par rapport à  $\nu = \mu \circ \varphi^{-1} = \mu_\varphi$  sur  $Y$  :

**Théorème 3.1.3 (Transfert)** Soit  $h : (Y, \mathcal{B}) \rightarrow K = \mathbb{R}, \bar{\mathbb{R}}, \mathbb{C}$  mesurable alors  $h$  est  $\nu$ -intégrable si et seulement si  $h \circ \varphi$  est  $\mu$ -intégrable et

$$\int_X h \circ \varphi d\mu = \int_Y h d\nu. \quad (3.1)$$

Il s'agit d'une formule de changement de variable abstraite, très générale puisque la seule condition pour le changement de variable  $y = \varphi(x)$  est que  $\varphi$  soit mesurable !

Malheureusement, la nouvelle mesure  $\nu = \mu_\varphi$  n'est pas explicite du tout. Ce résultat est donc essentiellement abstrait et difficile à utiliser pour des calculs explicites. Cependant dans le cas probabiliste, c'est ce résultat qui va permettre d'exprimer une espérance comme un calcul explicite d'intégrale contre la loi de la variable aléatoire.

On propose ci-dessous des résultats plus explicites, avec des conditions plus restrictives sur le changement de variables.

Soit  $I$  un intervalle de  $\mathbb{R}$  et  $\varphi : I \rightarrow \mathbb{R}$  strictement monotone et  $C^1$  tel que  $\varphi'$  ne s'annule pas sur  $I$ . Alors on a

$$\int_I f(x)dx = \int_{\varphi(I)} f(\varphi^{-1}(y))|(\varphi^{-1})'(y)| dy.$$

Pour cela, on pose  $y = \varphi(x)$  ou  $x = \varphi^{-1}(y)$  et en dérivant on a la relation entre  $dx$  et  $dy$

$$\frac{dx}{dy} = (\varphi^{-1})'(y) \quad \text{c'est à dire} \quad dx = (\varphi^{-1})'(y)dy.$$

### Cas de l'intégrale de Lebesgue

**Définition 3.1.2 (Difféomorphisme)** Soit  $F : D \subset \mathbb{R}^n \rightarrow D' \subset \mathbb{R}^n$  où  $D$  et  $D'$  sont des ouverts. L'application  $F$  est appelée un difféomorphisme si c'est une bijection de classe  $C^1$  dont la bijection réciproque est aussi de classe  $C^1$ .

**Définition 3.1.3 (Jacobien)** La matrice jacobienne d'un changement de variable

$$y = F(x) \iff (y_1, \dots, y_n) = (F_1(x_1, \dots, x_n), \dots, (F_n(x_1, \dots, x_n)))$$

est

$$J_F(x) = J_F(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1} & \dots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}.$$

Le jacobien est le déterminant de la matrice jacobienne  $\det J_F(x)$ .

La matrice jacobienne est la matrice des dérivées partielles.

**Rappel :** Calculs des déterminants d'ordre 2 et 3 :

- $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$
- Règle de Sarrus :  $\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + b_1c_2a_3 + c_1a_2b_3 - a_3b_2c_1 - b_3c_2a_1 - c_3a_2b_1.$
- Développements selon une ligne ou une colonne pour se ramener à des déterminants d'ordres inférieurs.

**Théorème 3.1.4 (Changement de variable)** Soit  $V$  un ouvert de  $\mathbb{R}^d$  et  $\varphi$  un  $C^1$ -difféomorphisme de  $V$  dans  $\varphi(V) \subset \mathbb{R}^d$ . Alors, on a les formules de changements de variables

$$\begin{aligned} \int_V f(x)\lambda(dx) &= \int_{\varphi(V)} f(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)|\lambda(dy) \\ \int_V h(\varphi(x))\lambda(dx) &= \int_{\varphi(V)} h(y) |J_{\varphi^{-1}}(y)|\lambda(dy) \end{aligned}$$

pour toutes fonctions  $f$  et  $h$  mesurables telles que  $f$  est  $\lambda$ -intégrable et  $h \circ \varphi$  est  $\lambda$ -intégrable.

**Démonstration :** Cf. cours d' « Intégration et mesures » [JCB-Leb]. □

### Changement de variables en coordonnées polaires dans $\mathbb{R}^2$

Un changement de variables utile dans le plan  $\mathbb{R}^2$  est le changement de variables en polaire qui consiste à passer de  $(x, y)$  représentant des coordonnées cartésiennes dans un repère orthonormé à  $(r, \theta)$  les coordonnées polaires correspondantes données par

$$(r, \theta) = \varphi(x, y) \iff \varphi^{-1} : \begin{cases} x &= r \cos \theta \\ y &= r \sin \theta \end{cases}, \quad r \in [0, +\infty[, \theta \in [0, 2\pi[.$$

Conformément au Théorème 3.1.4, on remplace alors  $dxdy$  par  $rdrd\theta$  car le jacobien du changement de variable est

$$J_{\varphi^{-1}}(r, \theta) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta + r \sin^2 \theta = r.$$

Ainsi :

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) \, dxdy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dxdy \\ &= \int \int_{[0, +\infty[ \times [0, 2\pi[} f(r \cos \theta, r \sin \theta) \, rdrd\theta. \end{aligned}$$

### Changement de variables en coordonnées sphériques dans $\mathbb{R}^3$

En dimension 3, le changement de variables utile est le changement en coordonnées sphériques donné par

$$(r, \theta, \varphi) = \phi(x, y, z) \iff \phi^{-1} : \begin{cases} x &= r \cos \theta \cos \varphi \\ y &= r \cos \theta \sin \varphi \\ z &= r \sin \theta \end{cases}$$

où  $\theta \in ]-\pi/2, \pi/2[$  est la latitude,  $\varphi \in [0, 2\pi[$  est la longitude et  $r \in [0, +\infty[$  la distance à l'origine. Le jacobien du changement de variable est

$$J_{\phi^{-1}}(r, \theta, \varphi) = \begin{vmatrix} \cos \theta \cos \varphi & -r \sin \theta \cos \varphi & -r \cos \theta \sin \varphi \\ \cos \theta \sin \varphi & -r \sin \theta \sin \varphi & r \sin \theta \cos \varphi \\ \sin \theta & r \cos \theta & 0 \end{vmatrix} = r^2 \cos \theta.$$

Ainsi le Théorème 3.1.4 s'écrit pour le changement de variable en sphériques :

$$\begin{aligned} &\int \int \int_{\mathbb{R}^3} f(x, y, z) \, dxdydz \\ &= \int \int \int_{[0, +\infty[ \times [0, 2\pi[ \times [-\frac{\pi}{2}, \frac{\pi}{2}[} f(r \cos \theta \cos \varphi, r \cos \theta \sin \varphi, r \sin \theta) r^2 \cos \theta \, drd\theta d\varphi. \end{aligned}$$

Ce type de changement de variable (polaires en dimension 2, sphériques en dimension 3) se généralise en dimension  $n$  pour passer de  $(x_1, \dots, x_n)$  à  $(r, \theta_1, \dots, \theta_{n-1})$  :

$$\begin{cases} x_1 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-2} \cos \theta_{n-1}, \\ x_2 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-2} \sin \theta_{n-1}, \\ x_3 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-3} \sin \theta_{n-2}, \\ x_4 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-4} \sin \theta_{n-3}, \\ \dots &= \dots \\ x_{n-1} &= r \cos \theta_1 \sin \theta_2, \\ x_n &= r \sin \theta_n. \end{cases}$$

## 3.2 Espérance d'une variable aléatoire réelle

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité et  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^+$  une variable aléatoire **positive**. L'intégrale de  $X$  par rapport à la mesure  $\mathbb{P}$  est appelée son espérance. Elle se définit comme l'intégrale par rapport à une mesure abstraite de proche en proche :

### Définition 3.2.1 (Espérance d'une variable aléatoire)

- *Indicatrice* : si  $X = \mathbf{1}_A$ ,  $\mathbb{E}[X] = \mathbb{P}(A)$ .
- *Variable étagée positive* : si  $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$  alors  $\mathbb{E}[X] = \sum_{i=1}^n a_i \mathbb{P}(A_i)$ . On vérifie que  $\mathbb{E}[X]$  ne dépend pas de l'expression de  $X$  : si  $X = \sum_{i=1}^p b_i \mathbf{1}_{B_i}$  alors  $\sum_{i=1}^p b_i \mathbb{P}(B_i) = \mathbb{E}[X]$ .
- *Variable positive* : si  $X$  est une variable aléatoire positive,

$$\mathbb{E}[X] = \sup (\mathbb{E}[Y] : Y \text{ variable aléatoire étagée} \leq X).$$

- *Variable quelconque* : Si  $X$  est une variable aléatoire quelconque telle que  $\mathbb{E}[|X|] < +\infty$  (au sens précédent) alors on définit l'espérance de  $X$  par

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

En effet comme  $0 \leq X^+ \leq |X|$  et  $0 \leq X^- \leq |X|$ , les espérances de variable aléatoires positives  $\mathbb{E}[X^+]$  et  $\mathbb{E}[X^-]$  sont bien définies. La quantité  $\mathbb{E}[|X|]$  s'appelle le moment d'ordre 1 de  $X$ .

L'espérance d'une variable aléatoire  $X$  est définie si son moment d'ordre 1 est fini :  $\mathbb{E}[|X|] < +\infty$ .

De la même façon si  $X$  est une variable aléatoire complexe intégrable, on peut définir son espérance en définissant comme précédemment l'espérance de ses parties réelle et imaginaire.

De façon générale, si  $X = (X_1, \dots, X_d)$  est un vecteur aléatoire, on définit son espérance comme étant le vecteur des espérances de ses marginales  $X_i$ ,  $1 \leq i \leq d$  :

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]).$$

**Définition 3.2.2 (Variable centrée)** Une variable aléatoire  $X$  intégrable est dite centrée si  $\mathbb{E}[X] = 0$ .

**Conséquence :** Comme l'espérance  $\mathbb{E}[X]$  est rien d'autre que l'intégrale (au sens de Lebesgue) de la fonction mesurable  $X$  par rapport à la mesure de probabilité  $\mathbb{P}$ , des propriétés de l'intégration, on déduit pour des variables aléatoires intégrables  $X, Y$  et des réels  $a, b$  :

**Proposition 3.2.1** 1. Si  $0 \leq X \leq Y$  alors  $0 \leq \mathbb{E}[X] \leq \mathbb{E}[Y]$ .

2. Si  $A \subset B$  et  $X \geq 0$  alors  $\mathbb{E}[X\mathbf{1}_A] \leq \mathbb{E}[X\mathbf{1}_B]$ .

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

4.  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$  (linéarité de  $\mathbb{E}$ ).

5. Si  $X \geq 0$ ,  $\mathbb{E}[X] = 0$  si et seulement si  $X = 0$  presque sûrement.

6. **Inégalité de Markov** : Si  $X$  est une variable aléatoire positive, on a :  $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$ , pour tout  $t > 0$ .

**Démonstration :** On renvoie à [JCB-Leb] pour les preuves relevant d'« intégration et de mesure » mais on rappelle la preuve de l'inégalité de Markov. Si  $X \geq 0$ , pour tout  $t > 0$  on a  $t\mathbf{1}_{\{X \geq t\}} \leq X$ . En prenant l'espérance, on a  $t\mathbb{P}(X \geq t) = \mathbb{E}[t\mathbf{1}_{\{X \geq t\}}] \leq \mathbb{E}[X]$ , ce qui permet de conclure en divisant par  $t > 0$ .  $\square$

Pour calculer des espérances, on se ramène à des intégrales sur  $\mathbb{R}$  avec la loi de la variable aléatoire comme mesure d'intégration grâce au théorème de transfert.

**Proposition 3.2.2** Soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  une variable aléatoire de loi  $\mathbb{P}_X$ . Alors  $X$  est  $\mathbb{P}$ -intégrable si et seulement si

$$\int_{\mathbb{R}} |x| \mathbb{P}_X(dx) < +\infty$$

et son espérance est alors

$$\mathbb{E}[X] = \int X d\mathbb{P} = \int_{\mathbb{R}} x \mathbb{P}_X(dx).$$

**Démonstration :** On applique le théorème de transfert avec la fonction mesurable  $\varphi = X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ , la mesure image  $\mathbb{P}_X$  et la fonction  $h(x) = x$  :

$$\int_{\Omega} h \circ X d\mathbb{P} = \int_{\mathbb{R}} h d\mathbb{P}_X.$$

$\square$

Plus généralement, si  $h : \mathbb{R} \rightarrow \mathbb{R}$  est borélienne (ie. mesurable) et  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  est une variable aléatoire réelle alors  $h(X)$  est une variable aléatoire car c'est une application de  $\Omega$  dans  $\mathbb{R}$

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} \mathbb{R} \xrightarrow{h} \mathbb{R},$$

c'est à dire  $h \circ X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ . Puis la variable aléatoire  $h(X)$  est  $\mathbb{P}$ -intégrable si et seulement si  $h$  est  $\mathbb{P}_X$ -intégrable ( $\int_{\mathbb{R}} |h(x)| \mathbb{P}_X(dx) < +\infty$ ) et alors

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X) d\mathbb{P} = \int_{\mathbb{R}} h(x) \mathbb{P}_X(dx).$$

Pour cela, on applique le théorème de transfert avec  $(X, \mathcal{A}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$ ,  $(Y, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et  $\varphi = X$ .

Parfois, il est plus facile de calculer directement l'espérance à l'aide de la fonction de répartition à l'aide du résultat suivant :

**Proposition 3.2.3** *Si  $X$  est une variable aléatoire réelle **positive** de fonction de répartition  $F_X$ , alors*

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > t) dt = \int_0^{+\infty} (1 - F_X(t)) dt.$$

De plus,  $\mathbb{E}[X] < +\infty$  si et seulement si, pour un ou tout  $\varepsilon > 0$ ,

$$\sum_{n \geq 0} \mathbb{P}(X > \varepsilon n) < +\infty \quad \text{ou} \quad \sum_{n \geq 0} 2^n \mathbb{P}(X > \varepsilon 2^n) < +\infty.$$

**Démonstration :** D'après le théorème de Fubini, on peut écrire

$$\int_0^{+\infty} \mathbb{P}(X > t) dt = \int_0^{+\infty} \mathbb{E}[\mathbf{1}_{\{X > t\}}] dt = \mathbb{E} \left[ \int_0^{+\infty} \mathbf{1}_{\{X > t\}} dt \right] = \mathbb{E} \left[ \int_0^X dt \right] = \mathbb{E}[X].$$

Pour la deuxième partie, on voit que pour  $t \in [n, n+1]$ , on a  $\mathbb{P}(X > n+1) \leq \mathbb{P}(X > t) \leq \mathbb{P}(X > n)$  et donc

$$\mathbb{P}(X > n+1) \leq \int_n^{n+1} \mathbb{P}(X > t) dt \leq \mathbb{P}(X > n)$$

puis en sommant

$$\sum_{n \geq 0} \mathbb{P}(X > n+1) \leq \int_0^{+\infty} \mathbb{P}(X > t) dt \leq \sum_{n \geq 0} \mathbb{P}(X > n)$$

en décomposant  $[0, +\infty[$  en intervalles  $[n, n+1[$ . De la même façon avec  $[0, +\infty[ = \bigcup_{n=0}^{+\infty} [2^n, 2^{n+1}[$

$$\sum_{n \geq 0} 2^n \mathbb{P}(X > 2^{n+1}) \leq \int_1^{+\infty} \mathbb{P}(X > t) dt \leq \sum_{n \geq 0} 2^n \mathbb{P}(X > 2^n)$$

et comme  $0 \leq \int_0^1 \mathbb{P}(X > t) dt \leq 1$  on a

$$\sum_{n \geq 0} 2^n \mathbb{P}(X > 2^{n+1}) \leq \int_0^{+\infty} \mathbb{P}(X > t) dt \leq 1 + \sum_{n \geq 0} 2^n \mathbb{P}(X > 2^n)$$

On conclut en remplaçant  $X$  par  $X/\varepsilon$ .  $\square$

## Espérance d'une variable aléatoire discrète

Pour utiliser correctement les lois des variables aléatoires discrètes, il est essentiel de noter que, pour une fonction  $f$  mesurable  $\int f d\delta_c = f(c)$  : cela est clair pour  $f = \mathbf{1}_A$  puis par linéarité pour  $f$  étagée et par convergence monotone pour  $f$  positive et enfin par linéarité pour  $f$  quelconque.

Soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  avec  $\mathcal{S}(X) = \{x_i : i \in I\}$  dénombrable. La loi de  $X$  est donnée par la mesure de probabilité discrète

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$$

où  $p_i = \mathbb{P}(X = x_i)$ ,  $i \in I$ . La loi est une somme (dénombrable) de mesures de Dirac : en chaque atome  $x \in \mathcal{S}(X)$ , il y a la masse  $\mathbb{P}(X = x)$ . Alors  $X$  est intégrable si et seulement si

$$\mathbb{E}[|X|] = \sum_{i \in I} |x_i| p_i < +\infty \quad (3.2)$$

et dans ce cas  $\mathbb{E}[X] = \sum_{i \in I} x_i p_i$  où la somme est au plus dénombrable puisque  $\mathcal{S}(X) = \{x_i : i \in I\}$  est dénombrable (la variable aléatoire  $X$  est discrète).

Noter que la condition (3.2) est vérifiée si  $X$  variable aléatoire a un nombre fini d'atomes car alors la somme dans (3.2) est une somme finie. De même, dans de nombreux cas usuels, les variables aléatoires discrètes sont positives et leur espérance sont forcément définies (mais possiblement infinies) sans prendre la peine de vérifier (3.2).

Si  $h : \mathbb{R} \rightarrow \mathbb{R}$  est mesurable, alors  $h(X)$  est une variable aléatoire discrète (Proposition 2.5.2), elle est intégrable si et seulement si

$$\sum_{i \in I} |h(x_i)| p_i < +\infty.$$

Son espérance est alors

$$\mathbb{E}[h(X)] = \sum_{i \in I} h(x_i) p_i.$$

### Exemples : (Espaces des variables aléatoires discrètes classiques)

- Si  $X = c$  est une variable aléatoire constante, sa loi est  $\mathbb{P}_X = \delta_c$ . Son espérance est

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

On le retrouve aussi avec l'expression  $\mathbb{E}[X] = \int_{\mathbb{R}} x d\delta_c(x) = c$ .

- Soit  $X$  de loi de Bernoulli de paramètre  $p$  notée  $b(p)$ . Son espérance est

$$\mathbb{E}[X] = p \times 1 + (1 - p) \times 0 = p.$$

- Soit  $X$  de loi équirépartie sur l'ensemble fini  $\{x_1, \dots, x_n\}$ . Son espérance est

$$\mathbb{E}[X] = \sum_{k=1}^n x_i \times \frac{1}{n} = \frac{x_1 + \dots + x_n}{n}.$$

- Soit  $X$  de loi binomiale de paramètres  $n, p$  notée  $\mathcal{B}(n, p)$ . Son espérance est  $\mathbb{E}[X] = np$ . En effet

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{l=0}^n \binom{n-1}{l} p^l (1-p)^{n-1-l} = np(1-p+p)^{n-1} = np. \end{aligned}$$

- Soit  $X$  de loi géométrique de paramètre  $p \in ]0, 1[$ , notée  $\mathcal{G}(p)$ . Son espérance est  $\mathbb{E}[X] = 1/p$ . En effet

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=1}^{+\infty} k(1-p)^{k-1} p = p \sum_{k=1}^{+\infty} (x^k)'_{x=1-p} = p \left( \sum_{k=0}^{+\infty} x^k \right)'_{x=1-p} \\ &= p \left( \frac{1}{1-x} \right)'_{x=1-p} = p \left( \frac{1}{(1-x)^2} \right)'_{x=1-p} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}. \end{aligned}$$

- Soit  $X$  de loi de Poisson de paramètre  $\lambda$ . Son espérance est  $\mathbb{E}[X] = \lambda$ . En effet

$$\mathbb{E}[X] = \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{l=1}^{+\infty} e^{-\lambda} \frac{\lambda^l}{l!} = \lambda.$$

## Espérance d'une variable aléatoire à densité

Soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  une variable de densité  $f$ . La loi de  $X$  est donnée par la mesure de forme intégrale

$$\mathbb{P}_X(A) = \int_A f d\lambda = \int_a^b f(x) dx, \quad \forall A \in \mathcal{B}(\mathbb{R}), \text{ ou } [a, b] \text{ intervalle}$$

où la densité  $f$  est une fonction mesurable positive d'intégrale 1. Alors  $X$  est intégrable si et seulement si

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} |x|f(x) dx < +\infty \quad (3.3)$$

et dans ce cas  $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x) dx$ .

Si  $h : \mathbb{R} \rightarrow \mathbb{R}$  est mesurable, alors  $h(X)$  est une variable aléatoire, elle est intégrable si et seulement si  $\int_{\mathbb{R}} |h(x)|f(x) dx < +\infty$ . Son espérance est alors

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x) dx.$$

Lorsqu'on considère une variable aléatoire positive, l'espérance est forcément définie (mais possiblement infinie) sans avoir à vérifier (3.3). Si la densité de  $X$  admet un support borné et n'a pas de singularité, la condition (3.3) est alors directement vérifiée et l'espérance bien définie.

### Exemples : (Lois à densité classiques)

- Loi uniforme

La variable aléatoire  $X$  suit une loi uniforme sur l'intervalle  $[a, b]$  ( $-\infty < a < b < +\infty$ ) si elle a une densité  $f$  constante sur cet intervalle et nulle en dehors. Sa densité est alors

$$f(t) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(t) = \begin{cases} 1/(b-a) & \text{si } t \in [a, b], \\ 0 & \text{si } t \notin [a, b]. \end{cases}$$

Son espérance est

$$\mathbb{E}[X] = \int xf(x)dx = \frac{1}{b-a} \int_a^b xdx = \frac{1}{b-a} \left( \frac{b^2}{2} - \frac{a^2}{2} \right) = \frac{a+b}{2}.$$

En fait on peut définir une loi uniforme sur un ensemble borélien  $A \subset \mathbb{R}$  quelconque de mesure de Lebesgue finie (pas forcément un intervalle), c'est la loi de densité  $\frac{1}{\lambda(A)} \mathbf{1}_A$ . Dans ce cas général, il n'y a pas de forme explicite de l'espérance.

- Loi exponentielle

La variable aléatoire  $X$  suit une loi exponentielle de paramètre  $\alpha > 0$ , notée  $\mathcal{E}(\alpha)$  si elle admet pour densité :

$$f(t) = \alpha e^{-\alpha t} \mathbf{1}_{[0,+\infty[}(t).$$

Elle est utilisée pour modéliser un temps d'attente d'un phénomène aléatoire. Le temps d'attente moyen est alors

$$\begin{aligned} \mathbb{E}[X] &= \int tf(t)dt = \int_0^{+\infty} \alpha te^{-\alpha t} dt = \left[ -te^{-\alpha t} \right]_0^{+\infty} + \int_0^{+\infty} e^{-\alpha t} dt \\ &= 0 + \frac{1}{\alpha} \left[ -e^{-\alpha t} \right]_0^{+\infty} = \frac{1}{\alpha}. \end{aligned}$$

- Loi de Cauchy

Une variable aléatoire réelle suit une loi de Cauchy de paramètre  $a \in \mathbb{R}_+^*$  si elle admet pour densité :

$$f(t) = \frac{a}{\pi} \frac{1}{a^2 + t^2}.$$

Son espérance n'est pas définie car  $\mathbb{E}[|X|] = \int_{\mathbb{R}} \frac{a|x|}{\pi(a^2 + x^2)} dx = +\infty$ , la fonction  $x/(1 + x^2) \approx 1/x$  n'est pas intégrable en  $\pm\infty$ .

- Loi normale (standard)

Une variable aléatoire  $X_0$  suit la loi normale standard  $\mathcal{N}(0, 1)$  si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

L'espérance est bien définie car  $|t|e^{-t^2/2}$  est intégrable en  $\pm\infty$ . Par (im)parité de  $te^{-t^2/2}$ , il vient immédiatement que son espérance est  $\mathbb{E}[X_0] = 0$ .

Dans le cas général  $X \sim \mathcal{N}(m, \sigma^2)$ , le plus simple est de remarquer que  $X$  peut se définir comme une translatée et dilatée de  $X_0$  par

$$X = m + \sigma X_0.$$

Son espérance est alors  $\mathbb{E}[X] = m + \sigma\mathbb{E}[X_0] = m$ .

**Proposition 3.2.4 (Centrer-réduire)** *Si  $X \sim \mathcal{N}(m, \sigma^2)$ ,  $\sigma^2 > 0$ , alors  $\tilde{X} = \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$  (cette opération s'appelle centrer-réduire). De ce fait, la loi  $\mathcal{N}(0, 1)$  s'appelle la loi normale centrée, réduite.*

Réiproquement, si  $X \sim \mathcal{N}(m, \sigma^2)$  alors on peut écrire, avec  $\tilde{X} \sim \mathcal{N}(0, 1)$  :

$$X \stackrel{\mathcal{L}}{=} m + \sigma \tilde{X}.$$

**Démonstration :** Il suffit de faire un changement de variable dans

$$\begin{aligned} \mathbb{P}\left(\frac{X-m}{\sigma} \in A\right) &= \mathbb{P}(X \in m + \sigma A) = \int_{\mathbb{R}} \mathbf{1}_{m+\sigma A} \frac{\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} dx \\ &= \int_{\mathbb{R}} \mathbf{1}_A \frac{\exp\left(-\frac{y^2}{2}\right)}{\sqrt{2\pi}} dy \quad (y = \frac{x-m}{\sigma}) \\ &= \mathbb{P}(X \in A). \end{aligned}$$

Ainsi  $\frac{X-m}{\sigma}$  a pour densité  $\frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$ , ie.  $\tilde{X} \sim \mathcal{N}(0, 1)$ . □

Par conséquent, si  $X \sim \mathcal{N}(m, \sigma^2)$ , on a  $\mathbb{E}[X] = m$  et  $\text{Var}(X) = \sigma^2$ .

### 3.3 Convergences monotone et dominée

On rappelle dans un contexte probabiliste les théorèmes fondamentaux de convergence en théorie de la mesure : convergence monotone, lemme de Fatou et théorème de convergence dominée. On renvoie au cours d'« Intégration et mesures » (par exemple [JCB-Leb]) pour les versions générales.

**Définition 3.3.1 (Convergence presque sûre)** *On dit que  $X_n$  converge vers  $X$  presque sûrement (ps) si l'ensemble des  $\omega \in \Omega$  tel que  $X_n(\omega) \rightarrow X(\omega)$  est de probabilité 1 :*

$$\mathbb{P}(X_n \rightarrow X) = 1.$$

**Théorème 3.3.1 (Convergence monotone, Beppo Levi)** *Soit  $(X_n)_{n \geq 0}$  une suite croissante de variables aléatoires positives ( $0 \leq X_n \leq X_{n+1}$ ). Soit  $X = \lim_{n \rightarrow +\infty} X_n$  la limite ps de  $X_n$  dans  $[0, +\infty]$ . Alors*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]. \quad (3.4)$$

**Démonstration :** Comme  $(X_n)_{n \geq 1}$ , la suite  $\mathbb{E}[X_n]$ ,  $n \geq 1$ , croît et a donc une limite  $\alpha$ . Comme  $X_n \leq X$  alors  $\mathbb{E}[X_n] \leq \mathbb{E}[X]$  et nécessairement  $\alpha \leq \mathbb{E}[X]$ .

Soit  $Y = \sum_{i=1}^p a_i \mathbf{1}_{B_i}$  une variable aléatoire étagée telle que  $Y \leq X$ . Pour  $c \in ]0, 1[$ , on note  $A_n = \{X_n \geq cY\}$ . Comme  $X_n \leq X_{n+1}$ , on a  $A_n \subset A_{n+1}$  et  $\bigcup_{n \geq 1} A_n = \{X \geq cY\} = \Omega$ . On a

$$\mathbb{E}[X_n] \geq \mathbb{E}[X_n \mathbf{1}_{A_n}] \geq c \mathbb{E}[Y \mathbf{1}_{A_n}] = c \sum_{i=1}^p a_i \mathbb{E}[\mathbf{1}_{B_i \cap A_n}] = \sum_{i=1}^p a_i \mathbb{P}(B_i \cap A_n). \quad (3.5)$$

Mais par comme  $(B_i \cap A_n)_{n \geq 1}$  forme une suite croissante de réunion  $B_i \cap (\bigcup_{n \geq 1} A_n) = B_i$ , la convergence monotone des probabilités assure  $\mathbb{P}(B_i \cap A_n) \rightarrow \mathbb{P}(B_i)$  quand  $n \rightarrow +\infty$ . En passant à la limite dans (3.5), on a  $\alpha \geq \sum_{i=1}^p a_i \mathbb{P}(B_i) = \mathbb{E}[Y]$ . On a donc

$$\alpha \geq \sup_{Y \text{ étagée} \leq X} \mathbb{E}[Y] = \mathbb{E}[X]$$

par définition de  $\mathbb{E}[X]$ . Cela achève de montrer  $\alpha = \mathbb{E}[X]$ . □

**Lemme 3.3.1 (Fatou)** *Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires positives. Alors*

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

**Démonstration :** On a  $\liminf_{n \rightarrow +\infty} X_n = \sup_{k \geq 0} \inf_{n \geq k} X_n = \lim_{k \rightarrow +\infty} \inf_{n \geq k} X_n$  où  $Y_k := \inf_{n \geq k} X_n$  est une suite croissante de variables aléatoires positives. En appliquant le théorème de convergence monotone à  $(Y_k)_{k \geq 0}$ , on a

$$\mathbb{E}\left[\liminf_{n \rightarrow +\infty} X_n\right] = \mathbb{E}\left[\lim_{k \rightarrow +\infty} Y_k\right] = \lim_{k \rightarrow +\infty} \mathbb{E}[Y_k].$$

Mais comme  $Y_k \leq X_n$  pour tout  $n \geq k$ , on a  $\mathbb{E}[Y_k] \leq \mathbb{E}[X_n]$  pour tout  $n \geq k$  et donc  $\mathbb{E}[Y_k] \leq \inf_{n \geq k} \mathbb{E}[X_n]$ . On a alors

$$\mathbb{E}\left[\liminf_{n \rightarrow +\infty} X_n\right] \leq \liminf_{n \rightarrow +\infty} \inf_{n \geq k} \mathbb{E}[X_n] = \sup_{n \geq 0} \inf_{n \geq k} \mathbb{E}[X_n] = \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n].$$

□

**Proposition 3.3.1** Soit  $Y$  une variable aléatoire intégrable et soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires intégrables.

1. Si  $Y \leq X_n$  alors  $\mathbb{E}\left[\liminf_{n \rightarrow +\infty} X_n\right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n]$ .
2. Si  $X_n \leq Y$  alors  $\limsup_{n \rightarrow +\infty} \mathbb{E}[X_n] \leq \mathbb{E}\left[\limsup_{n \rightarrow +\infty} X_n\right]$ .

**Démonstration :** Par le lemme de Fatou, on a

$$\mathbb{E}\left[\liminf_{n \rightarrow +\infty} (X_n - Y)\right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n - Y]$$

ce qui prouve 1) en rajoutant  $\mathbb{E}[Y]$  qui est finie par hypothèse. Pour le 2), on a de la même façon

$$\mathbb{E}\left[\liminf_{n \rightarrow +\infty} (Y - X_n)\right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[Y - X_n]$$

on retranche  $\mathbb{E}[Y] < +\infty$  et on change le signe pour récupérer l'inégalité sur la  $\limsup$ . □

**Théorème 3.3.2 (Convergence dominée)** Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires telle que  $X_n \rightarrow X$  ps quand  $n \rightarrow +\infty$ . S'il existe une variable aléatoire  $Y$  intégrable ( $\mathbb{E}[|Y|] < +\infty$ ) telle que pour tout  $n$ ,  $|X_n| \leq Y$  ps alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]. \quad (3.6)$$

**Conséquence.** Si la convergence est dominée, on peut intervertir limite et espérance.

**Démonstration :** Comme  $\lim_{n \rightarrow +\infty} X_n = X$  et  $-Y \leq X_n \leq Y$ , la Proposition 3.3.1 assure que

$$\limsup_{n \rightarrow +\infty} \mathbb{E}[X_n] \leq \mathbb{E}\left[\limsup_{n \rightarrow +\infty} X_n\right] = \mathbb{E}[X] = \mathbb{E}\left[\liminf_{n \rightarrow +\infty} X_n\right] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n]$$

ce qui prouve le résultat. □

Le résultat suivant n'est valable que pour des intégrales par rapport à une mesure de probabilité 1, c'est à dire pour des espérances.

**Théorème 3.3.3 (Inégalité de Jensen)** Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction convexe et  $X : \Omega \rightarrow \mathbb{R}$  une variable aléatoire telle que  $X$  et  $\varphi(X)$  soient intégrables. Alors

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

**Démonstration :** La convexité de  $\varphi$  assure qu'en tout point de son graphe, la courbe de  $\varphi$  est au dessus de sa tangente, ie. pour tout  $y \in \mathbb{R}$ , il existe  $d_y \in [\varphi'_g(y), \varphi'_d(y)]$  tel que

$$\forall y \in \mathbb{R}, \quad \varphi(x) \geq \varphi(y) + d_y(x - y).$$

En appliquant cette inégalité avec  $y = \mathbb{E}[X]$  et  $x = X(\omega)$  puis en prenant l'espérance, on a

$$\int \varphi(X(\omega)) d\mathbb{P} \geq \int \varphi(\mathbb{E}[X]) d\mathbb{P} + d_{\mathbb{E}[X]} \int (X(\omega) - \mathbb{E}[X]) d\mathbb{P}.$$

La conclusion en découle car  $\int (X(\omega) - \mathbb{E}[X]) d\mathbb{P} = 0$ .  $\square$

**Remarque 3.3.1** — S'il y a égalité dans l'inégalité de Jensen, alors il y a égalité dans  $\varphi(X) \geq \varphi(\mathbb{E}[X]) + d_{\mathbb{E}[X]}(X - \mathbb{E}[X])$ . Si  $\varphi$  est strictement convexe, cela exige  $X = \mathbb{E}[X]$  ps, ie.  $X$  est presque sûrement constante.

— De plus si l'égalité  $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$  a lieu pour toute variable aléatoire réelle  $X$  alors  $\varphi$  est convexe : en l'écrivant pour  $X$  de loi  $p\delta_x + (1-p)\delta_y$ , on a

$$\varphi(px + (1-p)y) \leq p\varphi(x) + (1-p)\varphi(y).$$

— Dans la pratique, l'inégalité de Jensen est le plus souvent utilisée pour les fonctions de la forme  $\varphi(x) = x, |x|, x^2$  et  $1/x$  lorsque  $x \geq 0$ . En particulier, pour une variable aléatoire positive dont l'inverse est intégrable, on a

$$\mathbb{E}[1/X] \geq 1/\mathbb{E}[X].$$

## 3.4 Moments des variables aléatoires

**Définition 3.4.1 (Moment)** Une variable aléatoire  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$  ou  $\mathbb{C}$  a un moment d'ordre  $p > 0$  si et seulement si

$$\mathbb{E}[|X|^p] = \int_{\Omega} |X|^p d\mathbb{P} < +\infty.$$

**Définition 3.4.2 (Espace  $L^p$ )** Si  $p \in (0, +\infty)$  alors  $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} : \mathbb{E}[|X|^p] < +\infty\}$ .

Si  $p = +\infty$ , alors  $L^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} : il existe c tel que \mathbb{P}(|X| > c) = 0\}$ . C'est l'ensemble des variables aléatoires presque sûrement bornées.

Si  $p = 0$ , alors  $L^0(\Omega, \mathcal{F}, \mathbb{P})$  est simplement l'ensemble des variables aléatoires. Si  $p \in [1, +\infty]$ ,  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  est un espace vectoriel normé avec pour norme

$$\begin{aligned}\|X\|_p &= (\mathbb{E}[|X|^p])^{1/p} \text{ (norme } L^p), \\ \|X\|_\infty &= \inf(c > 0 : \mathbb{P}(|X| > c) = 0) \text{ (supremum essentiel).}\end{aligned}$$

Il s'agit en effet d'un espace vectoriel car si  $X, Y \in L^p(\Omega, \mathcal{F}, \mathbb{P})$  alors par convexité de  $x \mapsto x^p$ , on a :

$$\begin{aligned}\frac{1}{2^p}|X+Y|^p &\leq \frac{1}{2}|X|^p + \frac{1}{2}|Y|^p \\ \mathbb{E}[|X+Y|^p] &\leq 2^{p-1}(\mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p])\end{aligned}\tag{3.7}$$

et donc  $X+Y \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ .

**Remarque 3.4.1** Formellement, on définit d'abord  $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} : \mathbb{E}[|X|^p] < +\infty\}$  pour lequel  $\|\cdot\|_p$  est seulement une semi-norme car  $\|X\|_p = 0$  entraîne seulement  $X = 0$  ps. Cela justifie l'introduction de l'espace quotient  $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) / \sim_{ps}$  où  $\sim_{ps}$  désigne la relation d'équivalence « être égale presque sûrement », ie.  $X \sim_{ps} Y$  si et seulement si  $\mathbb{P}(X \neq Y) = 0$ . Dans la suite, on identifie les variables aléatoires ps égales et on considère donc implicitement l'espace quotient  $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) / \sim_{ps}$ .

Comme  $\|\cdot\|_p$  est homogène ( $\|\lambda X\|_p = |\lambda| \|X\|_p$ ) et  $\|X\|_p = 0$  si et seulement si  $X = 0$ , il ne reste plus qu'à vérifier l'inégalité triangulaire pour s'assurer que  $\|\cdot\|_p$  définit une norme, cf. l'inégalité de Minkowski (Th. 3.4.2). On parle de la norme  $L^p$ .

Les résultats généraux sur les espaces  $L^p(X, \mathcal{A}, \mu)$  se formulent dans le cadre probabiliste de la façon suivante. On rappelle que deux réels  $p, q \in [1, +\infty]$  sont conjugués si  $\frac{1}{p} + \frac{1}{q} = 1$ . Par exemple, 2 et 2 sont conjugués, 1 et  $+\infty$  sont conjugués, 3 et  $3/2$  sont conjugués.

**Théorème 3.4.1 (Inégalité de Hölder)** Soient  $p, q$  exposants conjugués avec  $1 \leq p \leq +\infty$  et  $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ ,  $Y \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ . Alors  $XY \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  et

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

**Démonstration :** On utilise la convexité de la fonction exp : Si  $\|X\|_p \|Y\|_q = 0$  alors par exemple  $\|X\|_p = 0$  et donc  $X = 0$  ps ; on a alors  $XY = 0$  ps et  $\mathbb{E}[|XY|] = 0$ . Supposons donc que  $\|X\|_p \|Y\|_q \neq 0$  et soient  $a, b > 0$ . Il existe  $s, t \in \mathbb{R}$  tels que  $a = e^{s/p}$  et  $b = e^{t/q}$ . D'après la convexité de exp, on a

$$\exp\left(\frac{s}{p} + \frac{t}{q}\right) \leq \frac{1}{p} \exp(s) + \frac{1}{q} \exp(t)$$

c'est à dire  $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ . En utilisant  $a = |X|/\|X\|_p$  et  $b = |Y|/\|Y\|_q$ , on a

$$\frac{XY}{\|X\|_p\|Y\|_q} \leq \frac{1}{p} \frac{|X|^p}{\mathbb{E}[|X|^p]} + \frac{1}{q} \frac{|Y|^q}{\mathbb{E}[|Y|^q]}.$$

On conclut en prenant l'espérance dans l'inégalité précédente.  $\square$

Dans le cas particulier où  $p = q = 2$ , l'inégalité de Hölder devient celle de Cauchy-Schwarz :

**Corollaire 3.4.1 (Inégalité de Cauchy-Schwarz)** *Soient  $X, Y$  des variables aléatoires carées intégrables. Alors on a :*

$$\|XY\|_1 \leq \|X\|_2\|Y\|_2.$$

**Démonstration :** Une autre façon –classique– de prouver l'inégalité de Cauchy-Schwarz s'obtient en considérant le polynôme en  $c$  :

$$\mathbb{E}[(X + cY)^2] = \mathbb{E}[Y^2]c^2 + 2\mathbb{E}[XY]c + \mathbb{E}[X^2] \quad (3.8)$$

Comme il est de signe positif constant, son discriminant (réduit)  $\mathbb{E}[XY]^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2]$  est négatif.  $\square$

**Remarque 3.4.2** S'il y a égalité dans l'inégalité de Cauchy-Schwarz alors le discriminant du polynôme (3.8) est nul et ce polynôme a une racine double  $c_0$  pour laquelle on a alors  $\mathbb{E}[(X + c_0Y)^2] = 0$ , soit  $X = -c_0Y$  ps.

**Théorème 3.4.2 (Inégalité de Minkowski)** *Soient  $X, Y$  des variables aléatoires, alors pour  $1 \leq p \leq +\infty$ , on a :*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Cette inégalité justifie que  $\|\cdot\|_p$  vérifie l'inégalité triangulaire et définit bien une norme.

**Démonstration :** Pour  $p = +\infty$ , on considère  $c > \|X\|_\infty$  et  $d > \|Y\|_\infty$ . Alors presque sûrement,  $|X + Y| \leq |X| + |Y| \leq c + d$ , ce qui assure  $\|X + Y\|_\infty \leq c + d$ . En faisant  $c \searrow \|X\|_\infty$  et  $d \searrow \|Y\|_\infty$ , on a

$$\|X + Y\|_\infty \leq \|X\|_\infty + \|Y\|_\infty.$$

Dans la suite, on considère  $p \in [1, +\infty[$ . Si  $\|X + Y\|_p = 0$ , alors l'inégalité est triviale puis si  $\|X + Y\|_p = +\infty$  alors l'inégalité de convexité (3.7) montre que  $\|X + Y\|_p = +\infty$  exige  $\|X\|_p^p = +\infty$  ou  $\|Y\|_p^p = +\infty$ , en tout cas, l'inégalité de Minkowski est assurée.

On considère le cas  $\|X+Y\|_p \in (0, +\infty)$  et on utilise l'inégalité de Hölder avec les exposants conjugués  $p$  et  $q = \frac{p}{p-1}$  :

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

ce qui donne le résultat en simplifiant par  $\|X + Y\|_p \neq 0, +\infty$ .  $\square$

**Proposition 3.4.1** Si  $X$  possède un moment d'ordre  $q \in [1, +\infty]$ , pour tout  $1 \leq p \leq q$ ,  $X$  en possède un d'ordre  $p$ , ie.  $L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \subset L^p(\Omega, \mathcal{F}, \mathbb{P}) \subset L^q(\Omega, \mathcal{F}, \mathbb{P})$ . De plus, l'application  $p \mapsto \|\cdot\|_p$  est croissante. Puis

$$\lim_{p \rightarrow +\infty} \|X\|_p = \|X\|_\infty.$$

En fait, il s'agit d'une inclusion topologique (l'inclusion est continue pour la norme  $L^p$ ). En particulier, noter qu'une variable aléatoire bornée admet des moments de tous les ordres (c'est immédiat puisque si  $|X| \leq M$  alors  $\mathbb{E}[|X|^p] \leq \mathbb{E}[M^p] = M^p < +\infty$ ).

**Démonstration :** Si  $X \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ , on applique l'inégalité de Hölder avec  $\alpha = \frac{q}{p} > 1$  :

$$\mathbb{E}[|X|^p] = \mathbb{E}[|X|^p \times 1] \leq \mathbb{E}[|X^p|^{q/p}]^{p/q} \times \mathbb{E}[1^{q/(q-p)}]^{(q-p)/q} = \mathbb{E}[|X^q|]^{p/q},$$

ce qui prouve aussi  $\|X\|_p \leq \|X\|_q$  et la croissance de la norme  $L^p$  en  $p$ .

Pour la dernière assertion, on voit d'abord que  $|X| \leq \|X\|_\infty$  ps d'où  $\|X\|_p \leq \|X\|_\infty$ . Puis par croissance de  $p \mapsto \|\cdot\|_p$ , on déduit que  $\lim_{p \rightarrow +\infty} \|X\|_p \leq \|X\|_\infty$ .

Soit ensuite  $a$  arbitraire tel que  $a < \|X\|_\infty$ . Alors l'événement  $A := \{|X| > a\}$  est de probabilité strictement positive. En particulier  $\mathbb{P}(A)^{1/p} = \exp(\frac{1}{p} \ln \mathbb{P}(A)) \rightarrow 1$ , lorsque  $p \rightarrow +\infty$ .

On a  $|X|^p > a^p \mathbf{1}_A$ , d'où  $\|X\|_p \geq a \mathbb{P}(A)^{1/p}$ , donc  $\lim_{p \rightarrow +\infty} \|X\|_p \geq a$ . Comme ceci est valable pour tout  $a < \|X\|_\infty$ , on déduit, avec  $a \nearrow \|X\|_\infty$ ,  $\lim_{p \rightarrow +\infty} \|X\|_p \geq \|X\|_\infty$ , ce qui donne l'égalité cherchée.  $\square$

**Théorème 3.4.3 (Riesz-Fischer)** L'espace  $(L^p(\Omega, \mathcal{F}, \mathbb{P}), \|\cdot\|_p)$  est un espace vectoriel normé complet, c'est à dire un espace de Banach.

**Démonstration :** Cf. les Théorème 7.3.1 et Corollaire 7.3.1 avec la notion d'uniforme intégrabilité et la Remarque 7.3.2.  $\square$

**Remarque 3.4.3 (Dual de  $L^p$ )** Pour  $p, q$  conjugués et  $1 \leq p < +\infty$ , le dual topologique de  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  est  $L^q(\Omega, \mathcal{F}, \mathbb{P})$ . Autrement dit, les formes linéaires continues sur  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  sont les variables aléatoires de la forme  $X \in L^p(\Omega, \mathcal{F}, \mathbb{P}) \mapsto \mathbb{E}[XY] \in \mathbb{R}$  pour  $Y \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ . De plus,  $\|X\|_p = \sup(\mathbb{E}[XY] : \|Y\|_q \leq 1)$ . Attention, par contre le dual de  $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$  n'est pas  $L^1(\Omega, \mathcal{F}, \mathbb{P})$ .

Comme d'habitude la structure  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  est la plus riche puisque

**Proposition 3.4.2** L'espace  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  est un espace de Hilbert pour le produit scalaire

$$\langle X, Y \rangle = \mathbb{E}[XY].$$

Cette structure hilbertienne est intensivement utilisée pour les calculs gaussiens, cf. Chapitre 9.

**Exemple de calcul de moments.** Pour  $X \sim \mathcal{N}(0, 1)$ , on a (avec des intégrations par parties) :

$$\mathbb{E}[X^{2k+1}] = 0, \quad \mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k k!}.$$

Noter que tous les moments de la loi  $\mathcal{N}(0, 1)$  sont définis car

$$\mathbb{E}[|X|^k] = \int_{-\infty}^{+\infty} |x|^k \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^k \exp(-x^2/2) dx < +\infty.$$

## 3.5 Variance, covariance

En plus du moment d'ordre 1, lié à l'espérance, le plus important est le moment d'ordre 2, lié à la variance.

**Définition 3.5.1 (Variance)** Si  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ , on définit la variance de  $X$  par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (3.9)$$

On définit aussi l'**écart-type** de  $X$  par  $\sigma_X = \sqrt{\text{Var}(X)}$ .

**Remarque 3.5.1** L'espérance d'une variable aléatoire donne la valeur moyenne (au sens probabiliste) de la variable aléatoire. Sa variance (ou son écart-type) mesure la dispersion des valeurs de la variable aléatoire autour de sa moyenne.

Il est équivalent de dire que la variance de  $X$  est finie et que  $X$  admet un moment d'ordre 2 fini ou encore que son écart-type est fini.

Dans les cas discrets et à densité, la variance s'exprime comme une somme ou comme une intégrale :

- Si  $X$  est discrète de domaine  $\mathcal{S}(X) = \{x_i : i \in I\}$ , avec  $I$  dénombrable, la loi de  $X$  est  $\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$  et la variance en (3.9) devient

$$\text{Var}(X) = \sum_{i \in I} (x_i - \mathbb{E}[X])^2 \mathbb{P}_X(\{x_i\}) = \sum_{i \in I} (x_i - \mathbb{E}[X])^2 p_i.$$

- Si  $X$  est une variable aléatoire de densité  $f$  alors la loi de  $X$  est la mesure de densité  $f$ ,  $\mathbb{P}_X(dx) = f(x)dx$  et la variance en (3.9) devient

$$\text{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x) dx.$$

**Proposition 3.5.1 (Propriétés de la variance)**

1.  $\text{Var}(X) \geq 0$ .
2.  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$  (*Formule de Koenig*).
3.  $\text{Var}(aX) = a^2 \text{Var}(X)$ .
4.  $\text{Var}(X + b) = \text{Var}(X)$  pour toute constante  $b \in \mathbb{R}$ .
5.  $\text{Var}(X) = 0$  si et seulement si  $X$  est constante ps (et vaut alors  $\mathbb{E}[X]$ ).

La variance est un opérateur quadratique non linéaire.

**Démonstration :** • Le premier point est clair.

- On développe  $\text{Var}(X)$ , en notant  $m = \mathbb{E}[X]$  :

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - m)^2] = \mathbb{E}[X^2 - 2Xm + m^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[Xm] + m^2 \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]m + m^2 \\ &= \mathbb{E}[X^2] - 2m^2 + m^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned}$$

- Pour les troisième et quatrième points :

$$\begin{aligned} \text{Var}(aX) &= \mathbb{E}[(aX - \mathbb{E}[aX])^2] = \mathbb{E}[(a(X - \mathbb{E}[X]))^2] = a^2 \mathbb{E}[(X - \mathbb{E}[X])^2] = a^2 \text{Var}(X) \\ \text{Var}(X + b) &= \mathbb{E}[(X + b - \mathbb{E}[X + b])^2] = \mathbb{E}[(X + b - \mathbb{E}[X] - b)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \text{Var}(X). \end{aligned}$$

- Si  $X = c$  une constante ps alors  $\mathbb{E}[X] = \mathbb{E}[c] = c$  et  $\mathbb{E}[X^2] = \mathbb{E}[c^2] = c^2$  si bien que  $\text{Var}(X) = c^2 - c^2 = 0$ . Réciproquement, si  $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = 0$  alors la variable aléatoire  $(X - \mathbb{E}[X])^2$ , positive d'espérance nulle, est elle-même nulle ps, c'est à dire  $X = \mathbb{E}[X]$  ps.  $\square$

**Définition 3.5.2 (Covariance)** Soient  $X, Y$  deux variables aléatoires avec des variances finies, on définit la covariance de  $X$  et de  $Y$  par

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

**Remarque 3.5.1** —  $(X, Y) \mapsto \text{Cov}(X, Y)$  est une application bilinéaire.

- Si  $X$  ou  $Y$  est centrée alors  $\text{Cov}(X, Y) = \mathbb{E}[XY]$ .
- $\text{Cov}(X, X) = \text{Var}(X)$ .
- $\text{Cov}(X, a) = 0$  puisque  $\text{Cov}(X, a) = \mathbb{E}[(X - \mathbb{E}[X])(a - \mathbb{E}[a])] = \mathbb{E}[(X - \mathbb{E}[X])(a - a)] = 0$ .
- On a une identité de Koenig pour la covariance :  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ . En effet,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[Y]] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0. \end{aligned}$$

La variance  $\text{Var}$  est une forme quadratique sur  $L^2(\Omega)$ , d'application bilinéaire associée la covariance  $\text{Cov}$ .

**Proposition 3.5.2** *Si  $X$  et  $Y$  sont deux variables aléatoires avec des moments d'ordre 2 alors*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y).$$

**Démonstration :** Il suffit de développer les carrés dans le calcul de la variance par l'identité de Koenig :

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \end{aligned}$$

□

Attention, en général, on n'a pas  $\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$ . Prendre par exemple  $X = Y$ . Par contre on verra que c'est vrai si  $X$  et  $Y$  sont des variables aléatoires indépendantes, cf. Proposition 9.2.5.

**Proposition 3.5.3** *Soient  $X, Y$  des variables aléatoires de variances finies. On a*

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

**Démonstration :** On applique l'inégalité de Cauchy-Schwarz (Corollaire 3.4.1) :

$$\begin{aligned} |\text{Cov}(X, Y)| &= |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \\ &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\text{Var}(X) \text{Var}(Y)}. \end{aligned}$$

□

**Définition 3.5.3 (Coefficient de corrélation)** *Soient  $X, Y$  deux variables aléatoires non constantes et de variances finies, leur coefficient de corrélation est*

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Comme  $X, Y$  sont non constantes, on a  $\text{Var}(X), \text{Var}(Y) \neq 0$ . On constate facilement avec la Proposition 3.5.3 que  $\rho(X, Y) \in [-1, 1]$ .

**Proposition 3.5.4** Si  $\rho(X, Y) = \pm 1$  alors il y a un lien linéaire entre  $X$  et  $Y$  :  $Y = aX + b$ , pour  $a, b \in \mathbb{R}$ . En plus, si  $X$  n'est pas constante, on montre que

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad b = \mathbb{E}[Y] - a\mathbb{E}[X].$$

**Démonstration :** En effet,  $\rho(X, Y) = \pm 1$  s'il y a égalité dans l'inégalité de Cauchy-Schwarz de la preuve de la Prop 3.5.3. D'après la Remarque 3.4.2, dans ce cas il y a une relation linéaire entre  $X$  et  $Y$  :  $(X - \mathbb{E}[X]) = -c_0(Y - \mathbb{E}[Y])$ , ie.  $X = \mathbb{E}[X] + c_0\mathbb{E}[Y] - c_0Y$ . Puis,

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(X, aX + b) \\ &= a \text{Cov}(X, X) + \text{Cov}(X, b) \\ &= a \text{Var}(X) + 0 \end{aligned}$$

car  $\text{Cov}(X, b) = \mathbb{E}[Xb] - \mathbb{E}[X]\mathbb{E}[b] = 0$ . On en déduit  $a$ . Puis comme  $Y = aX + b$ , on a  $\mathbb{E}[Y] = a\mathbb{E}[X] + b$ , d'où vient  $b$  aussi.  $\square$

Heuristiquement, si  $\rho(X, Y)$  est proche de 1 ou  $-1$ , alors c'est que  $X$  et  $Y$  prennent des valeurs « peu » dispersées par rapport à une relation linéaire affine et on pourra supposer que c'est le cas. On a donc à peu près  $Y \approx aX + b$ . La droite  $y = ax + b$  est appelée droite de régression linéaire.

De plus, si  $\rho$  est proche de 1 alors  $a > 0$  et si  $\rho$  est proche de  $-1$  alors  $a < 0$ .

**Théorème 3.5.1 (Inégalité de Tchebychev)** Si  $\text{Var}(X) < +\infty$ , on a pour tout  $t > 0$  :

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

**Démonstration :** Par l'inégalité de Markov (point 5 dans la Proposition 3.2.1), on a

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

$\square$

**Application :** On jette 3600 fois un dé. Minorer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 480 et 720.

Notons  $S$  le nombre d'apparitions du 1. On peut voir  $S$  comme la somme de 3600 variables aléatoires de Bernoulli indépendantes de paramètre  $p = 1/6$  (probabilité d'apparition du 1 au cours d'un lancer). Par un raisonnement classique,  $S$  suit une loi  $\mathcal{B}(3600, p)$ . On cherche ici

$$\mathbb{P}(480 < S < 720) = \sum_{k=481}^{719} \binom{3600}{k} p^k (1-p)^{3600-k}.$$

Ce résultat exact ne peut être calculé en pratique, même un ordinateur très puissant ne pouvant calculer tous ces coefficients binomiaux pour des chiffres aussi grands.

On peut penser à approximer la loi  $\mathcal{B}(3600, 1/6)$  par  $\mathcal{P}(600)$  mais il resterait à calculer

$$\sum_{k=481}^{719} e^{-600} \frac{600^k}{k!},$$

ce qui n'est pas évident non plus sans aide informatique.

On a alors recours à l'inégalité de Tchebychev pour obtenir une majoration de la probabilité cherchée : notons que  $\mathbb{E}[S] = np = 3600/6 = 600$  et  $\text{Var}(X) = npq = 3600 \times 5/6 \times 1/6 = 500$ . Remarquons de plus que

$$480 < S < 720 \iff -120 < S - 600 < 120 \iff |S - 600| < 120.$$

D'où

$$\begin{aligned} \mathbb{P}(\mathbb{P}(480 < S < 720)) &= \mathbb{P}(|S - 600| < 120) = 1 - \mathbb{P}(|S - 600| \geq 120) \\ &\geq 1 - \frac{500}{120^2} \geq 0,95833\dots \end{aligned}$$

**Remarque 3.5.2** Les valeurs 480 et 720 sont symétriques par rapport à la moyenne 600 de la variable aléatoire considérée, ce sont  $600 \pm 120$ . Ce n'est pas nécessaire : on peut aussi appliquer l'inégalité de Tchebychev sur un intervalle non centré autour de l'espérance. Il suffit pour cela d'utiliser le plus grand intervalle centré sur l'espérance qu'il contient (quitte à perdre de la précision de la majoration). Ainsi pour minorer  $\mathbb{P}(550 < S < 700)$ , il suffit de remarquer que

$$550 < S < 700 \iff \underbrace{550 < S < 650}_{\text{intervalle centré autour de } 600} \iff -50 < S - 600 < 50.$$

et

$$\begin{aligned} \mathbb{P}(550 < S < 700) \geq \mathbb{P}(550 < S < 650) &= \mathbb{P}(-50 < S - 600 < 50) = \mathbb{P}(|S - 600| < 50) \\ &= 1 - \mathbb{P}(|S - 600| \geq 50) \geq 1 - \frac{500}{50^2} = 0,8. \end{aligned}$$

## Cas des vecteurs aléatoires

La variance est définie pour les variables aléatoires réelles. Pour les vecteurs, l'équivalent est la matrice de covariance. Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  avec un moment d'ordre 2 ie.  $\mathbb{E}[\|X\|^2] < +\infty$  pour n'importe quel norme de  $\mathbb{R}^2$ , typiquement la norme euclidienne :

$$\mathbb{E}[\|X\|^2] = \mathbb{E}[X_1^2] + \dots + \mathbb{E}[X_d^2]$$

ainsi  $\mathbb{E}[\|X\|^2] < +\infty$  si et seulement si pour tout  $1 \leq i \leq d$ ,  $\mathbb{E}[X_i^2] < +\infty$ . Par l'inégalité de Cauchy-Schwarz, les moments joints  $\mathbb{E}[X_i X_j]$  sont tous bien définis.

**Définition 3.5.4 (Matrice de covariance)** La matrice de covariance d'un vecteur  $X = (X_1, \dots, X_d)$  avec un moment d'ordre 2 est la matrice  $d \times d$ , symétrique :

$$K_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}.$$

Comme  $\|X\|^2 = |X_1|^2 + \dots + |X_d|^2$ , on a  $\mathbb{E}[\|X\|^2] = \mathbb{E}[|X_1|^2] + \dots + \mathbb{E}[|X_d|^2]$  et le moment d'ordre deux de  $X$  est fini si et seulement si tous les moments d'ordre deux de ses marginales  $X_i$  le sont.

On peut constater que  $K_X = \mathbb{E}[XX^t] - \mathbb{E}[X]\mathbb{E}[X]^t$  où  $x^t$  désigne la transposée d'un vecteur  $x \in \mathbb{R}^d$ . En fait,  $K_X$  est aussi une matrice positive puisque si  $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ , on a

$$a^t K_X a = \sum_{i,j=1}^d (K_X)_{i,j} a_i a_j = \mathbb{E} \left[ \left( \sum_{i=1}^d a_i (X_i - \mathbb{E}[X_i]) \right)^2 \right] \geq 0.$$

La matrice de covariance est définie positive si aucune combinaison linéaire des marginales  $X_i$  n'est presque sûrement constante, ie. si le vecteur n'est pas concentré sur un sous-espace propre de  $\mathbb{R}^d$ . Dans ce cas, cela signifierait que la dimension  $d$  n'est pas la bonne dimension pour étudier  $X$  : le vecteur vivrait dans un espace de dimension inférieure.

### 3.6 Tableau comparatif des formules pour des variables aléatoires discrètes et continues à densité

Lorsque les intégrales et les séries concernées sont absolument convergentes, on a le tableau comparatif suivant entre le cas général et les déclinaisons discrètes et à densité (cas continu) :

$X$	Cas général	Variable discrète	Variable à densité $f$
$\mathcal{S}(X)$	quelconque	$\{x_i : i \in I\}$	$\mathbb{R}$ ou un borélien
$\mathbb{P}(a \leq X \leq b)$	$\int_{\Omega} \mathbf{1}_{[a,b]}(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{i : a \leq x_i \leq b} \mathbb{P}(X = x_i)$	$\int_a^b f(t) dt$
$F(x) = \mathbb{P}(X \leq x)$	$\int_{\Omega} \mathbf{1}_{]-\infty,x]}(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{i : x_i \leq x} \mathbb{P}(X = x_i)$	$\int_{-\infty}^x f(t) dt$
$\mathbb{E}[g(X)]$	$\int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{i \in I} g(x_i) \mathbb{P}(X = x_i)$	$\int_{-\infty}^{+\infty} g(t) f(t) dt$
$\mathbb{E}[X]$	$\int X(\omega) d\mathbb{P}(\omega)$	$\sum_{i \in I} x_i \mathbb{P}(X = x_i)$	$\int_{-\infty}^{+\infty} t f(t) dt$
$\mathbb{E}[X^2]$	$\int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega)$	$\sum_{i \in I} x_i^2 \mathbb{P}(X = x_i)$	$\int_{-\infty}^{+\infty} t^2 f(t) dt$
$\text{Var}(X)$	$\int_{\Omega} (X(\omega) - \mathbb{E}[X])^2 d\mathbb{P}(\omega)$	$\sum_{i \in I} (x_i - \mathbb{E}[X])^2 \mathbb{P}(X = x_i)$	$\int_{-\infty}^{+\infty} (t - \mathbb{E}[X])^2 f(t) dt$

# Chapitre 4

## Fonction caractéristique

### Introduction

On a déjà vu que la fonction de répartition d'une variable aléatoire  $X$  caractérise sa loi (Théorème 2.3.2). Autrement dit, dans le cas d'une variable aléatoire  $X$  réelle, la donnée de

$$F_X(t) = \mathbb{E}[\mathbf{1}_{]-\infty,t]}(X)], \quad t \in \mathbb{R}$$

détermine la loi de  $X$ . Mais comme les indicatrices sont des fonctions boréliennes bornées, la donnée de  $\mathbb{E}[\phi(X)]$  pour toute fonction borélienne bornée  $\phi$  caractérise la loi  $\mathbb{P}_X$ . En fait, comme par exemple  $\mathbf{1}_{]-\infty,t]}$  peut être approchée par

$$\varphi_n(x) = \begin{cases} 1 & \text{si } x \leq t \\ 1 + n(t - x) & \text{si } t \leq x \leq t + 1/n \\ 0 & \text{si } x > t + 1/n \end{cases}$$

il s'ensuit (par convergence dominée) que la donnée de  $\mathbb{E}[\phi(X)]$  pour toute fonction continue bornée  $\phi$  sur  $\mathbb{R}$  caractérise  $\mathbb{P}_X$ . Comme on peut aussi approcher les fonctions indicatrices par des fonctions  $C^\infty$  bornées, la donnée de  $\mathbb{E}[\phi(X)]$  pour toute fonction infiniment dérivable bornée caractérise encore  $\mathbb{P}_X$ . Dans ce chapitre, on va plus loin encore et on montre qu'il suffit de ne considérer que deux types de fonction : les sinus et les cosinus. Bien sûr, tout cela est lié à l'analyse de Fourier.

### 4.1 Définition et premières propriétés

**Définition 4.1.1 (Fonction caractéristique)** Soit  $X : \Omega \rightarrow \mathbb{R}^d$  un vecteur aléatoire sur l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Sa fonction caractéristique  $\varphi_X$  est définie comme la fonction de  $d$  variables, donc de  $\mathbb{R}^d$  dans  $\mathbb{C}$  :

$$\varphi_X(t) = \mathbb{E}[e^{i\langle t, X \rangle}] = \int_{\Omega} e^{i\langle t, X \rangle} d\mathbb{P} = \mathbb{E}\left[\exp\left(i \sum_{j=1}^d t_j X_j\right)\right] \quad (4.1)$$

où  $\langle t, s \rangle = \sum_{j=1}^d t_j s_j$  désigne le produit scalaire euclidien de  $\mathbb{R}^d$ .

Lorsque  $X$  est une variable aléatoire réelle, sa fonction caractéristique  $\varphi_X(t) = \mathbb{E}[e^{itX}]$  devient une fonction de  $\mathbb{R}$  dans  $\mathbb{C}$ .

#### Remarque 4.1.1 (Transformée de Fourier)

- $\varphi_X$  est toujours bien définie car  $|e^{itX}| = 1$  est  $\mathbb{P}$ -intégrable.
- Si  $X$  est de densité  $f$ , la fonction caractéristique s'écrit

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx = \mathcal{F}(f)\left(\frac{t}{2\pi}\right).$$

À quelques notations près, il s'agit de la transformée de Fourier  $\mathcal{F}(f)$  de la densité  $f$  de la variable aléatoire  $X$ .

- Si  $X$  est une variable aléatoire discrète, la fonction caractéristique peut se voir encore comme une transformée de Fourier mais il faut alors introduire une transformée de Fourier par rapport à une mesure discrète.
- De façon générale

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itx} \mathbb{P}_X(dx).$$

La fonction caractéristique  $\varphi_X$  est donc la transformée de Fourier de la mesure  $\mathbb{P}_X$ , loi de la variable aléatoire  $X$ .

Le résultat suivant justifie que le nom de ces fonctions n'est pas usurpé.

**Théorème 4.1.1** *La fonction caractéristique caractérise la loi :  $\varphi_X = \varphi_Y$  si et seulement si  $\mathbb{P}_X = \mathbb{P}_Y$ .*

La justification vient du résultat d'inversion des transformées de Fourier. Comme  $|\varphi_X(t)| \leq \mathbb{E}[|e^{itX}|] = 1$ ,  $\varphi_X$  est bornée donc intégrable par rapport à  $\mathbb{P}$ . En interprétant  $\varphi_X(t)$  comme la transformée de Fourier de la loi  $\mathbb{P}_X$ , le théorème d'inversion des transformées de Fourier s'applique et donne

$$\mathbb{P}_X = \mathcal{F}(\varphi_X).$$

Si  $\varphi_X = \varphi_Y$ , on en déduit donc l'égalité des lois  $\mathbb{P}_X = \mathbb{P}_Y$ .

On en donne maintenant une preuve complète :

**Démonstration :** On suppose toujours que  $X$  et  $Y$  sont deux variables aléatoires réelles. Par un calcul direct (ou alors utiliser le calcul de la fonction caractéristique de la loi de Laplace), on a

$$\int_{\mathbb{R}} e^{iyx} e^{-\lambda|x|} dx = \frac{1}{\lambda - iy} + \frac{1}{\lambda + iy} = \frac{2\lambda}{\lambda^2 + y^2}. \quad (4.2)$$

Soit  $f \in C_c^0(\mathbb{R}, \mathbb{R})$  (fonction continue à support compact). Par convergence dominée et comme  $\int_{\mathbb{R}} \frac{2}{1+u^2} du = 2\pi$ , on a facilement :

$$2\pi f(y) = \lim_{\lambda \searrow 0} \int_{\mathbb{R}} f(y - \lambda u) \frac{2}{1+u^2} du \quad (4.3)$$

$$= \lim_{\lambda \searrow 0} \int_{\mathbb{R}} f(s) \frac{2}{1 + \left(\frac{y-s}{\lambda}\right)^2} \frac{ds}{\lambda} \quad (\text{en posant } s = y - \lambda u) \quad (4.4)$$

$$= \lim_{\lambda \searrow 0} \int_{\mathbb{R}} f(s) \frac{2\lambda}{\lambda^2 + (y-s)^2} ds. \quad (4.5)$$

Comme de plus  $|\int_{\mathbb{R}} f(s) \frac{2\lambda}{\lambda^2 + (y-s)^2} ds| \leq 2\pi \|f\|_{\infty}$ , par convergence dominée, on a alors encore

$$\begin{aligned} \mathbb{E}[f(X)] &= \int_{\mathbb{R}} f(y) \mathbb{P}_X(dy) = \int_{\mathbb{R}} \left( \lim_{\lambda \searrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} f(s) \frac{2\lambda}{\lambda^2 + (y-s)^2} ds \right) \mathbb{P}_X(dy) \\ &\quad (\text{avec } y-s \text{ à la place de } y \text{ dans (4.5)}) \end{aligned} \quad (4.6)$$

$$\begin{aligned} &= \lim_{\lambda \searrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} f(s) \frac{2\lambda}{\lambda^2 + (y-s)^2} ds \mathbb{P}_X(dy) \quad (\text{convergence dominée}) \\ &= \lim_{\lambda \searrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} f(s) \int_{\mathbb{R}} e^{iyx} e^{-\lambda|x|} dx ds \mathbb{P}_X(dy) \quad (\text{avec (4.2)}) \\ &= \lim_{\lambda \searrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \left( \int_{\mathbb{R}} e^{iyx} \mathbb{P}_X(dy) \right) e^{-\lambda|x|} e^{-isx} dx \right) f(s) ds \quad (\text{par Fubini}) \\ &= \lim_{\lambda \searrow 0} \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi_X(x) e^{-\lambda|x|} e^{-isx} dx f(s) ds \end{aligned} \quad (4.7)$$

où dans l'avant dernière égalité on a utilisé le théorème de Fubini en notant que

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} |e^{iyx} e^{-\lambda|x|} e^{-isx} f(s)| \mathbb{P}_X(dy) ds &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} |f(s)| e^{-\lambda|x|} \mathbb{P}_X(dy) dx ds \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} |f(s)| e^{-\lambda|x|} dx ds < +\infty. \end{aligned}$$

avec  $f \in C_c^0(\mathbb{R}, \mathbb{R})$  et  $\lambda > 0$ . Le terme de droite de (4.7) ne dépend que de  $f$  et de  $\varphi_X$ . Si  $\varphi_X = \varphi_Y$  alors  $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$  pour toute fonction  $f$  continue à support compact.

Comme pour  $-\infty < a \leq b < +\infty$ , l'indicatrice  $\mathbf{1}_{[a,b]}$  s'approxime par une suite  $f_n$  de fonctions continues à support compact (on peut choisir la suite  $f_n$  bornée par exemple par 2), l'égalité s'étend par passage à la limite à

$$\begin{aligned} \mathbb{P}_X([a, b]) &= \mathbb{E}[\mathbf{1}_{[a,b]}(X)] = \mathbb{E}\left[\lim_{n \rightarrow +\infty} f_n(X)\right] \\ &= \lim_{n \rightarrow +\infty} \mathbb{E}[f_n(X)] \quad (\text{convergence dominée}) \\ &= \lim_{n \rightarrow +\infty} \mathbb{E}[f_n(Y)] \\ &= \mathbb{E}\left[\lim_{n \rightarrow +\infty} f_n(Y)\right] \quad (\text{convergence dominée}) \\ &= \mathbb{E}[\mathbf{1}_{[a,b]}(Y)] = \mathbb{P}_Y([a, b]). \end{aligned} \quad (4.8)$$

Par convergence monotone des probabilités, on étend encore (4.8) aux intervalles  $-\infty \leq a \leq b \leq +\infty$ . Finalement, par un argument de classe monotone, comme  $\{[a, b], -\infty \leq a \leq b \leq +\infty\}$  engendre  $\mathcal{B}(\mathbb{R})$ , l'égalité s'étend à tout  $\mathcal{B}(\mathbb{R})$ , ie.  $\mathbb{P}_X = \mathbb{P}_Y$  (Théorème 1.3.2 de Dynkin). Donc  $\varphi_X$  caractérise bien la loi  $\mathbb{P}_X$ .  $\square$

## 4.2 Propriétés et exemples

**Proposition 4.2.1** Soit  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . Alors sa fonction caractéristique vérifie :

- i)  $|\varphi_X(t)| \leq 1$ .
- ii)  $\varphi_X(-t) = \overline{\varphi_X(t)}$ .
- iii)  $\varphi_X(0) = 1$ .
- iv)  $\varphi_X$  est uniformément continue.
- v)  $\varphi_X$  est de type positif, ie.

$$\forall n \in \mathbb{N}^*, \quad \forall t_1, \dots, t_n \in \mathbb{R}^d, \quad \forall z_1, \dots, z_n \in \mathbb{C}, \quad \sum_{j,k=1}^n \varphi_X(t_k - t_j) z_k \bar{z}_j \geq 0.$$

**Démonstration :** i), ii), iii) sont faciles. Pour prouver iv), on calcule pour  $t, h \in \mathbb{R}^d$ ,

$$\begin{aligned} \varphi_X(t+h) - \varphi_X(t) &= \mathbb{E}[e^{i\langle t+h, X \rangle} - e^{i\langle t, X \rangle}] \\ &= \mathbb{E}\left[e^{i\langle t, X \rangle} e^{\frac{i}{2}\langle h, X \rangle} \left(e^{\frac{i}{2}\langle h, X \rangle} - e^{-\frac{i}{2}\langle h, X \rangle}\right)\right] \\ &= \mathbb{E}\left[e^{i\langle t, X \rangle} e^{\frac{i}{2}\langle h, X \rangle} 2i \sin\left(\frac{\langle h, X \rangle}{2}\right)\right]. \end{aligned}$$

D'où

$$|\varphi_X(t+h) - \varphi_X(t)| \leq \mathbb{E}\left[2 \left|\sin \frac{\langle h, X \rangle}{2}\right|\right] \leq \mathbb{E}[(|h| |X|) \wedge 2].$$

car  $|\sin(x)| \leq |x| \wedge 1$ . Par convergence dominée, le majorant ci-dessus tend vers 0 quand  $h \rightarrow 0$ , et ce indépendamment de  $t$ , ce qui garantit l'uniforme continuité.

Pour v), on a

$$\begin{aligned} \sum_{j,k=1}^n \varphi_X(t_k - t_j) z_k \bar{z}_j &= \sum_{j,k=1}^n \mathbb{E}[e^{i\langle t_k - t_j, X \rangle}] z_k \bar{z}_j = \mathbb{E}\left[\left(\sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k\right) \left(\sum_{j=1}^n e^{i\langle -t_j, X \rangle} \bar{z}_j\right)\right] \\ &= \mathbb{E}\left[\left(\sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k\right) \overline{\left(\sum_{j=1}^n e^{i\langle t_j, X \rangle} z_j\right)}\right] = \mathbb{E}\left[\left|\sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k\right|^2\right] \geq 0. \end{aligned}$$

En fait les propriétés de la Proposition 4.2.1 caractérisent les fonctions caractéristiques :

**Théorème 4.2.1 (Bochner-Herglotz)** Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  une fonction continue en 0 et telle que  $\varphi(0) = 1$ . On suppose que  $\varphi$  est de type positif, ie.

$$\forall n \geq 1, \forall z_1, \dots, z_n \in \mathbb{R}, \forall t_1, \dots, t_n \in \mathbb{R}, \quad \sum_{j,k=1}^n \varphi(t_j - t_k) z_j \bar{z}_k \geq 0.$$

Alors  $\varphi$  est la fonction caractéristique d'une variable aléatoire sur  $\mathbb{R}$ .

On définit la fonction suivante :

77

$$G_N(x) = \begin{cases} 0 & \text{pour } x < -\pi \\ \frac{1}{2\pi} \int_{-\pi}^x g_N(y) dy & \text{pour } -\pi \leq x < \pi \\ 1 & \text{pour } x \geq \pi. \end{cases}$$

Alors

$$\left(1 - \frac{|s|}{N}\right) \theta(s) = \int_{-\pi}^{\pi} e^{isx} dG_N(x), \quad -N \leq s \leq N.$$

En particulier  $\int_{-\pi}^{\pi} dG_N(x) = \theta(0) = 1$  et la fonction  $G_N$  est donc une fonction de distribution concentrée sur  $[-\pi, \pi]$ . La suite  $(G_n)_{n \geq 1}$  est une suite de fonctions uniformément bornées, croissantes, continues à droite. D'après le Théorème de Helly (Th. 4.2.2), on peut extraire  $(N_k)_{k \geq 1}$  de façon que  $(G_{N_k})_{k \geq 1}$  converge en loi vers une fonction de distribution  $G$  bornée croissante continue à droite sur  $\mathbb{R}$ .

Comme pour tout  $\varepsilon > 0$ , on a  $G_N(-\pi - \varepsilon) = 0$  et  $G_N(\pi + \varepsilon) = 1$ , on en déduit par la convergence précédente que  $G(-\pi - \varepsilon) = 0$  et  $G(\pi + \varepsilon) = 1$ . Ainsi,  $G$  est une fonction de distribution concentrée sur  $[-\pi, \pi]$ . De plus comme

$$\left(1 - \frac{|s|}{N_k}\right) \theta(s) = \int_{-\pi}^{\pi} e^{isx} dG_{N_k}(x)$$

en passant à la limite avec le Théorème de Helly-Bray (Th. 4.2.3), on a

$$\theta(s) = \int_{-\pi}^{\pi} e^{isx} dG(x), \quad s = 0, \pm 1, \pm 2, \dots$$

ce qui prouve le Lemme de Herglotz. □

On revient à la preuve du Théorème 4.2.1 (Th. Bochner-Herglotz).

**Démonstration :** Par hypothèse, pour tout  $n \geq 1$ , la suite  $\varphi(s/n)$ ,  $s = 0, \pm 1, \pm 2, \dots$  satisfait les conditions du lemme. Il existe donc  $G_n$  concentrée sur  $[-\pi, \pi]$  tel que

$$\varphi\left(\frac{s}{n}\right) = \int_{-\pi}^{\pi} e^{isx} dG_n(x), \quad s = 0, \pm 1, \pm 2, \dots$$

On note  $F_n(x) = G_n(x/n)$ ,  $x \in \mathbb{R}$ , la distribution concentrée sur  $[-n\pi, n\pi]$  et  $\varphi_n$  sa fonction caractéristique. Par changement de variables, on a :

$$\varphi_n(t) = \int_{-n\pi}^{n\pi} e^{its} dF_n(x) = \int_{-\pi}^{\pi} e^{ity} dG_n(y)$$

c'est à dire

$$\varphi\left(\frac{s}{n}\right) = \varphi_n\left(\frac{s}{n}\right), \quad s = 0, \pm 1, \pm 2, \dots$$

Soit  $t \in \mathbb{R}$  et  $n \geq 1$  alors il existe un entier  $k = k(n, t)$  tel que  $0 \leq t - (k/n) < 1/n$ . Avec  $\theta = t - (k/n)$ , on a

$$\begin{aligned} \left| \varphi_n(t) - \varphi_n\left(\frac{k}{n}\right) \right| &= \left| \varphi_n\left(\theta + \frac{k}{n}\right) - \varphi_n\left(\frac{k}{n}\right) \right| \\ &= \left| \int_{-n\pi}^{n\pi} e^{ix(\theta + \frac{k}{n})} dF_n(x) - \int_{-n\pi}^{n\pi} e^{ix\frac{k}{n}} dF_n(x) \right| \\ &\leq \int_{-n\pi}^{n\pi} |e^{i\theta x} - 1| dF_n(x) \\ &\leq \left( \int_{-n\pi}^{n\pi} |e^{i\theta x} - 1|^2 dF_n(x) \right)^{1/2} \quad (\text{par l'inégalité de Cauchy-Schwarz}) \\ &= \left( 2 \int_{-n\pi}^{n\pi} (1 - \cos(\theta x)) dF_n(x) \right)^{1/2}. \end{aligned}$$

Mais on a

$$1 - \cos(\theta x) \leq 1 - \cos\left(\frac{x}{n}\right) \quad \text{pour } 0 \leq \theta < \frac{1}{n} \text{ et } -n\pi \leq x \leq n\pi.$$

Ainsi

$$\begin{aligned} \left| \varphi_n(t) - \varphi_n\left(\frac{k}{n}\right) \right| &\leq \left( 2 \int_{-n\pi}^{n\pi} \left( 1 - \cos\left(\frac{x}{n}\right) \right) dF_n(x) \right)^{1/2} \\ &= \left( 2 \left( 1 - \operatorname{Re} \varphi_n\left(\frac{1}{n}\right) \right) \right)^{1/2} \\ &= \left( 2 \left( 1 - \operatorname{Re} \varphi\left(\frac{1}{n}\right) \right) \right)^{1/2}. \end{aligned}$$

Comme  $\varphi$  est continue sur  $\mathbb{R}$  et  $\varphi(0) = 1$ , on déduit de la majoration précédente

$$\lim_{n \rightarrow +\infty} \left| \varphi_n(t) - \varphi_n\left(\frac{k}{n}\right) \right| = 0.$$

Finalement,

$$\begin{aligned} \varphi_n(t) &= \left( \varphi_n(t) - \varphi_n\left(\frac{k}{n}\right) \right) + \varphi_n\left(\frac{k}{n}\right) \\ &= \left( \varphi_n(t) - \varphi_n\left(\frac{k}{n}\right) \right) + \varphi\left(\frac{k}{n}\right) \\ &\rightarrow 0 + \varphi(t) = \varphi(t), \quad n \rightarrow +\infty. \end{aligned}$$

On a exprimé  $\varphi$  comme la limite simple de fonctions caractéristiques. D'après le Théorème 7.4.4 (Th. de Lévy),  $\varphi$  est une fonction caractéristique, ce qui achève la preuve du Théorème 4.2.1.  $\square$

## Exemples

79

- Pour des variables aléatoires discrètes  $X$  de domaine  $\mathcal{S}(X) = \{x_k : k \in I\}$  avec  $I$  dénombrable et de probabilités ponctuelles  $p_k = \mathbb{P}(X = x_k)$  en  $x_k$ ,  $k \in I$ , on a

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \sum_{k \in I} e^{itx_k} p_k.$$

Variables	Domaine $\mathcal{S}(X)$	Probabilités ponctuelles $p_k$	Fonction caractéristique $\varphi_X$
Loi Dirac $\delta_a$	$\{a\}$	1	$e^{iat}$
Loi de Bernoulli $b(p)$	$\{0, 1\}$	$1-p, p$	$1-p+pe^{it}$
Loi binomiale $\mathcal{B}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$(1-p+pe^{it})^n$
Loi géométrique $\mathcal{G}(p)$	$\mathbb{N} \setminus \{0\}$	$(1-p)^{k-1} p$	$(pe^{it}/(1-(1-p)e^{it}))$
Loi de Poisson $\mathcal{P}(\alpha)$	$\mathbb{N}$	$e^{-\alpha} \frac{\alpha^k}{k!}$	$e^{\alpha(e^{it}-1)}$

- Pour des variables à densité : si  $X$  est de densité  $f$  alors  $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$ .

Variables	Densité $f$	Intervalle	Fonction caractéristique $\varphi_X$
Loi normale standard	$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$-\infty < x < +\infty$	$e^{-t^2/2}$
Loi normale $\mathcal{N}(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$	$-\infty < x < +\infty$	$e^{imt - \sigma^2 t^2/2}$
Loi uniforme $\mathcal{U}([0, 1])$	1	$0 \leq x \leq 1$	$\frac{e^{it}-1}{it}$
Loi exponentielle $\mathcal{E}(\lambda)$	$\lambda e^{-\lambda x}$	$0 \leq x < +\infty$	$\frac{\lambda}{\lambda-it}$
Loi gamma $\gamma(p, \lambda)$		$0 \leq x < +\infty$	$\left(\frac{\lambda}{\lambda-it}\right)^p$
Loi de Cauchy $\mathcal{C}(a)$	$\frac{a}{\pi(a^2+x^2)}$	$-\infty < x < +\infty$	$e^{-a t }$

**Proposition 4.2.2** Soit  $X \sim \mathcal{N}(m, \sigma^2)$ . Alors sa fonction caractéristique est

$$\varphi_X(t) = \exp(imt - \sigma^2 t^2/2).$$

**Démonstration :** On sait que  $X$  a même loi que  $\sigma X_0 + m$  où  $X_0 \sim \mathcal{N}(0, 1)$ , elle a aussi même fonction caractéristique :

$$\begin{aligned} \varphi_X(t) &= \varphi_{\sigma X_0 + m}(t) = \mathbb{E}[e^{it(\sigma X_0 + m)}] \\ &= \mathbb{E}[e^{it\sigma X_0} e^{itm}] = e^{itm} \varphi_{X_0}(t\sigma). \end{aligned}$$

Il suffit donc de montrer que  $\varphi_{X_0}(t) = e^{-t^2/2}$ . Or

$$\varphi_{X_0}(t) = \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2 - 2itx}{2}\right) \frac{dx}{\sqrt{2\pi}} \\
&= \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-it)^2 - (it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \\
&= \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-it)^2}{2}\right) e^{-t^2/2} \frac{dx}{\sqrt{2\pi}} \\
&= e^{-t^2/2} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \\
&= e^{-t^2/2}
\end{aligned}$$

car avec le changement de variable (complexe)  $y = x - it$ , on a :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) \frac{dy}{\sqrt{2\pi}} = 1.$$

Une autre preuve consiste à voir que  $\varphi_{X_0}$  est solution de l'équation différentielle

$$\begin{cases} y'(t) + ty(t) &= 0 \\ y(0) &= 1 \end{cases}$$

ce qui exige  $\varphi_{X_0}(t) = \exp(-t^2/2)$ . □

Il existe plusieurs formules dites d'inversion permettant d'obtenir effectivement la loi à partir de la fonction caractéristique. En voici une possible :

**Théorème 4.2.4 (Inversion de Fourier)** *Soit  $\varphi$  une fonction caractéristique intégrable par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ . Alors  $\varphi$  est la fonction caractéristique d'un vecteur aléatoire  $X$  admettant la densité continue bornée donnée par :*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \varphi(t) dt, \quad x \in \mathbb{R}^d.$$

**Démonstration :** Nous utilisons dans cette preuve les idées de celle de la caractérisation de la loi par la fonction caractéristique, ie. du Théorème 4.1.1. Pour simplifier nous allons supposer encore une fois que  $d = 1$ . L'expression (4.7) pour  $g \in C_c^0(\mathbb{R}, \mathbb{R})$  s'écrit :

$$\begin{aligned}
\mathbb{E}[g(X)] &= \frac{1}{2\pi} \lim_{\lambda \searrow 0} \int_{\mathbb{R}} \int_{\mathbb{R}} g(s) \varphi(x) e^{-isx} e^{-\lambda|x|} dx ds \\
&= \int_{\mathbb{R}} g(s) \left( \frac{1}{2\pi} \int_{\mathbb{R}} e^{-isx} \varphi(x) dx \right) ds
\end{aligned}$$

par convergence dominée (et Fubini) car

$$|g(s)\varphi(x)e^{-isx}e^{-\lambda|x|}| \leq |g(s)| |\varphi(x)| \in L^1(dx \otimes ds)$$

la fonction  $g$  étant bornée à support compact et  $\varphi$  intégrable. Nécessairement, la fonction

$$f(s) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-isx} \varphi(x) dx$$

doit être la densité de  $X$  car la donnée de  $\mathbb{E}[g(X)]$  pour  $g$  continue à support compact caractérise la loi de  $X$ , cf. la discussion en introduction de ce chapitre.  $\square$

**Remarque 4.2.1** Nous verrons dans les chapitres suivants que la fonction caractéristique est bien adaptée pour

- étudier l'indépendance de variables aléatoires, cf. Théorème 5.2.1 ;
- étudier la convergence en loi de variables aléatoires, cf. Théorème 7.4.3 et le Théorème de Lévy.

### 4.3 Régularité de la fonction caractéristique

**Proposition 4.3.1** Soit  $X$  une variable aléatoire réelle de fonction caractéristique  $\varphi_X$ .

- i) Si  $X$  admet un moment d'ordre  $k \in \mathbb{N}^*$  alors  $\varphi_X$  est  $k$  fois dérivable, de dérivée  $l$ -ème ( $l \leq k$ )

$$\varphi_X^{(l)}(t) = i^l \mathbb{E}[X^l e^{itX}].$$

En particulier,  $\varphi_X^{(l)}(0) = i^l \mathbb{E}[X^l]$ .

- ii) Si  $\varphi_X$  est  $k$  fois dérivable en 0,  $k \in \mathbb{N}^*$ , alors  $X$  admet des moments d'ordres plus petits ou égaux à  $2[k/2]$  (ie. au plus grand entier pair inférieur à  $k$ ).

**Démonstration :** La première partie vient du théorème de convergence dominée avec la borne (vérifiée par récurrence)

$$\left| e^{iu} - 1 - \frac{iu}{1!} - \cdots - \frac{(iu)^{n-1}}{(n-1)!} \right| \leq \frac{|u|^n}{n!}.$$

Réciproquement si  $\varphi_X$  est  $k$  fois dérivable en 0,  $k \in \mathbb{N}^*$ , comme  $\|\cdot\|_p$  est une fonction croissante de  $p$  (Prop. 3.4.1), il suffit de voir l'existence des moments d'ordres pairs inférieurs à  $2[k/2]$  pour assurer celle de tous les moments d'ordres inférieurs à  $2[k/2]$ . On suppose montrée l'existence du moment d'ordre  $2r$ ,  $r < [k/2]$ . La première partie donne alors  $\varphi^{(2r)}(t) = (-1)^r \mathbb{E}[X^{2r} e^{itX}]$ . C'est vrai pour  $r = 0$ , ce qui initialise la récurrence. On a alors

$$\begin{aligned} \varphi_X^{(2r+2)}(0) &= \lim_{h \rightarrow 0} \frac{\varphi_X^{(2r)}(2h) - 2\varphi_X^{(2r)}(0) + \varphi_X^{(2r)}(-2h)}{4h^2} \quad (\text{formule de Taylor à l'ordre 2}) \\ &= \lim_{h \rightarrow 0} (-1)^r \mathbb{E} \left[ \frac{e^{2ihX} - 2 + e^{-2ihX}}{4h^2} X^{2r} \right] \quad (\text{hypothèse de récurrence}) \end{aligned}$$

$$\begin{aligned}
&= \lim_{h \rightarrow 0} (-1)^r \mathbb{E} \left[ \left( \frac{e^{ihX} - e^{-ihX}}{2h} \right)^2 X^{2r} \right] \\
&= \lim_{h \rightarrow 0} (-1)^{r+1} \mathbb{E} \left[ X^{2r+2} \left( \frac{\sin(hX)}{hX} \right)^2 \right].
\end{aligned}$$

Finalement, le lemme de Fatou donne

$$\begin{aligned}
\mathbb{E}[X^{2r+2}] &= \mathbb{E} \left[ \lim_{h \rightarrow 0} \left( X^{2r+2} \left( \frac{\sin(hX)}{hX} \right)^2 \right) \right] = \mathbb{E} \left[ \liminf_{h \rightarrow 0} \left( X^{2r+2} \left( \frac{\sin(hX)}{hX} \right)^2 \right) \right] \\
&\leq \liminf_{h \rightarrow 0} \mathbb{E} \left[ X^{2r+2} \left( \frac{\sin(hX)}{hX} \right)^2 \right] = (-1)^{r+1} \varphi^{(2r+2)}(0) < +\infty
\end{aligned}$$

ce qui établit la récurrence et prouve la Proposition 4.3.1.  $\square$

**Conséquence :** La régularité de la fonction caractéristique  $\varphi_X$  est liée à l'intégrabilité de la variable aléatoire  $X$ . En particulier,

- $\varphi_X(t)$  est définie pour tout  $t \in \mathbb{R}$ .
- Si  $X$  a un moment d'ordre 1, alors  $\varphi_X$  est dérivable (et même  $C^1$ ) avec

$$\varphi'_X(t) = i\mathbb{E}[X e^{itX}].$$

En particulier :  $i\mathbb{E}[X] = \varphi'_X(0)$ .

- Si  $X$  a un moment d'ordre 2, alors  $\varphi_X$  est dérivable (et même  $C^2$ ) avec

$$\varphi''_X(t) = -\mathbb{E}[X^2 e^{itX}].$$

En particulier :  $\mathbb{E}[X^2] = -\varphi'_X(0)$  et  $\text{Var}(X) = -\varphi''_X(0) + \varphi'_X(0)^2$ .

## 4.4 Autres transformées caractéristiques

D'autres fonctions que la fonction caractéristique caractérisent la loi d'une variable aléatoire  $X$ . Lorsque  $X$  est à valeurs entières, on utilise la fonction génératrice des moments  $M_X$ . Lorsque  $X \geq 0$ , on utilise la transformée de Laplace  $\phi_X$ . Ces fonctions partageront le même type de propriété vis à vis de l'indépendance que la fonction caractéristique, cf. Propositions 4.4.2 et 4.4.4.

### 4.4.1 Fonction génératrice

On considère une variable  $X$  entière  $\mathcal{S}(X) \subset \mathbb{N}$  et  $p_k = \mathbb{P}(X = k)$ .

**Définition 4.4.1** Soit  $X$  une variable aléatoire à valeurs entières. La fonction génératrice de  $X$  est

$$M_X(t) = \mathbb{E}[t^X] = \sum_{k \geq 0} p_k t^k.$$

Comme  $\sum_{k \geq 0} p_k = 1$ , la fonction  $M_X$  est bien définie et elle est continue sur  $[0, 1]$ . On l'appelle aussi la fonction génératrice des probabilités pour la distinguer de la fonction génératrice des moments (cf. transformée de Laplace).

**Proposition 4.4.1** *La fonction génératrice est  $C^\infty$  sur  $[0, 1[$  et dérivable à l'ordre  $p$  en 1 si et seulement si  $\mathbb{E}[X^p] < +\infty$ . De plus, les dérivées successives en 0 permettent de retrouver la loi de  $X$  :*

$$M_X^{(k)}(0) = k! p_k.$$

Cela explique la terminologie « fonction génératrice des probabilités ».

**Démonstration :** La preuve vient de ce que  $M_X$  est une série entière de rayon de convergence  $R = 1$  : les dérivées de  $M_X$  se calculent termes à termes.  $\square$

**Exemples.** Des calculs simples donnent :

- si  $X \sim b(p)$  :  $M_X(t) = (1 - p) + pt$  ;
- si  $X \sim \mathcal{B}(n, p)$  :  $M_X(t) = (1 - p + pt)^n$  ;
- si  $X \sim \mathcal{G}(p)$  :  $M_X = \frac{pt}{1 - (1-p)t}$  ;
- si  $X \sim \mathcal{P}(\lambda)$  :  $M_X(t) = \exp(\lambda(t - 1))$ .

Pour les vecteurs aléatoires à valeurs dans  $\mathbb{N}^d$ , on définit aussi une notion de fonction génératrice :

$$M_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = \mathbb{E}[t_1^{X_1} \cdots t_d^{X_d}].$$

La fonction génératrice se comporte bien vis à vis de l'indépendance :

**Proposition 4.4.2 (Fonction génératrice et indépendance)** *Soient  $X, Y$  deux variables aléatoires entières.*

1. *Si  $X \perp\!\!\!\perp Y$  alors  $M_{X+Y}(t) = M_X(t)M_Y(t)$ .*
2.  *$X \perp\!\!\!\perp Y$  si et seulement si  $M_{(X,Y)}(s, t) = M_X(s)M_Y(t)$ .*

#### 4.4.2 Transformée de Laplace

**Définition 4.4.2 (Transformée de Laplace)** *Soit  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{F}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . On appelle transformée de Laplace de  $X$  (ou fonction génératrice des moments de  $X$ ), et on note*

$$\phi_X(t) = \mathbb{E}[\exp(\langle t, X \rangle)]$$

*la fonction définie pour les valeurs de  $t \in \mathbb{R}^d$  pour lesquelles  $\exp(\langle t, X \rangle)$  est intégrable.*

En particulier, noter que si  $\mathbb{E}[e^{\langle t, X \rangle}] < +\infty$  alors tous les moments de  $X$  sont finis. On peut alors les déterminer par convergence dominée :

$$\frac{\partial^p \phi}{\partial t_{i_1} \cdots \partial t_{i_p}}(t) = \mathbb{E}[X_{i_1} \cdots X_{i_p} \exp(\langle t, X \rangle)]$$

D'où  $\mathbb{E}[X_{i_1} \dots X_{i_p}] = \frac{\partial^p \phi}{\partial t_{i_1} \dots \partial t_{i_p}}(0)$ , ce qui explique la terminologie « fonction génératrice des moments » pour la transformée de Laplace.

**Remarque 4.4.1** La transformée de Laplace, si elle est finie sur un voisinage de 0, caractérise la loi, tout comme la transformée de Fourier.

Pour une variable aléatoire réelle, on spécialise les résultats :

**Proposition 4.4.3** *Soit  $X$  une variable aléatoire réelle telle que  $\exp(tX)$  est intégrable pour  $t$  dans un intervalle contenant 0. Alors la transformée de Laplace  $\phi_X$  est définie sur un intervalle contenant 0. De plus elle est analytique sur un voisinage de 0 et sur ce voisinage on a*

$$\phi_X(t) = \sum_{n \geq 0} \frac{t^n}{n!} \mathbb{E}[X^n],$$

*pour tout  $t$  dans ce voisinage. En particulier,*

$$\mu_n(X) = \mathbb{E}[X^n] = \phi_X^{(n)}(0).$$

La transformée de Laplace se comporte bien vis à vis de l'indépendance :

**Proposition 4.4.4 (Transformée de Laplace et indépendance)** *Soient  $X, Y$  deux variables aléatoires.*

1. *Si  $X \perp\!\!\!\perp Y$  alors  $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ .*
2.  *$X \perp\!\!\!\perp Y$  si et seulement si  $\phi_{(X,Y)}(s,t) = \phi_X(s)\phi_Y(t)$ .*

Si  $X$  est une variable aléatoire positive alors on définit la transformée de Laplace par  $\mathbb{E}[\exp(-tX)]$  qui est alors définie pour tout  $t \geq 0$ .

# Chapitre 5

## Indépendance

Il s'agit d'une notion fondamentale en probabilité. Dans ce chapitre, on la définit et on donne différents critères d'indépendance puis on explore quelques conséquences de l'indépendance. On note  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité.

### 5.1 Concept d'indépendance

#### Définition 5.1.1 (Indépendance d'évènements)

- Deux évènements  $A, B \in \mathcal{F}$  d'un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  sont indépendants si

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

On note  $A \perp\!\!\!\perp B$ .

- Une famille quelconque d'évènements observables  $A_i, i \in I$  est dite (mutuellement) indépendante si pour toute sous-famille  $A_{i_1}, \dots, A_{i_p}$  avec  $i_k \in I$ , on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_p}).$$

#### Remarque 5.1.1

- En particulier,  $A$  et  $B$  incompatibles ne peuvent pas être indépendants à moins que l'un des deux ne soit de probabilité nulle. Sinon  $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$ , tandis que  $\mathbb{P}(A)\mathbb{P}(B) > 0$ . Il ne faut donc pas confondre les deux notions.
- Il ne faut pas confondre l'indépendance mutuelle de  $A_1, \dots, A_n$  avec l'indépendance deux à deux.

La notion d'indépendance se généralise aux tribus de la façon suivante :

#### Définition 5.1.2 (Indépendance de familles, de tribus)

- Deux familles de parties mesurables  $\mathcal{M}_1$  et  $\mathcal{M}_2$  sur un même espace  $(\Omega, \mathcal{F}, \mathbb{P})$  sont indépendantes si pour tout  $A \in \mathcal{M}_1$  et  $B \in \mathcal{M}_2$  on a

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

En particulier, on définit de cette façon l'indépendance de deux tribus  $\mathcal{G}_1, \mathcal{G}_2$ . On note alors  $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2$ .

- Une famille quelconque de tribus  $(\mathcal{F}_i)_{i \in I}$  est dite indépendante si pour tous  $A_i \in \mathcal{F}_i$ ,  $i \in I$ , la famille d'évènements  $(A_i)_{i \in I}$  est (mutuellement) indépendante.

On rappelle la Proposition 1.3.2 du Chapitre 1 dû au théorème de classe monotone (Th. 1.3.1) :

**Proposition 5.1.1** Soient  $\mathcal{A}_1$  et  $\mathcal{A}_2$  deux familles d'ensembles stables par intersection finie (deux algèbres ou  $\pi$ -systèmes) indépendantes dans  $(\Omega, \mathcal{F}, \mathbb{P})$ . Alors les tribus engendrées  $\sigma(\mathcal{A}_1)$  et  $\sigma(\mathcal{A}_2)$  sont indépendantes.

**Démonstration :** Soit  $A_1 \in \mathcal{A}_1$ . Considérons la famille

$$\mathcal{M}_1 = \{A_2 \in \sigma(\mathcal{A}_2) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}$$

des évènements indépendants de  $A_1$ . Il s'agit d'une classe monotone qui contient, par hypothèse,  $\mathcal{A}_2$ . Elle contient donc la classe monotone engendrée par  $\mathcal{A}_2$ , qui coïncide par le théorème de classe monotone (Th. 1.3.1) avec  $\sigma(\mathcal{A}_2)$ . Soit maintenant  $A_2 \in \sigma(\mathcal{A}_2)$ . Considérons la famille

$$\mathcal{M}_2 = \{A_1 \in \sigma(\mathcal{A}_1) : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)\}$$

des évènements indépendants de  $A_2$ . Il s'agit encore d'une classe monotone qui, d'après la première partie, contient  $\mathcal{A}_1$  et d'après ce qui précède  $\sigma(\mathcal{A}_1)$ . On a donc  $\sigma(\mathcal{A}_1) \perp\!\!\!\perp A_2$ . Comme  $A_2 \in \mathcal{A}_2$  est quelconque, on a  $\sigma(\mathcal{A}_1) \perp\!\!\!\perp \sigma(\mathcal{A}_2)$ .  $\square$

Plus généralement, on a :

**Proposition 5.1.2** Soient  $(C_i)_{i \in I}$  des  $\pi$ -systèmes (stables par intersection finie). Alors les  $(C_i)_{i \in I}$  sont indépendants si et seulement si  $(\sigma(C_i))_{i \in I}$  le sont.

Il suffit donc de vérifier l'indépendance des tribus sur des  $\pi$ -systèmes qui les engendent.

**Démonstration :** Le sens réciproque est immédiat. Pour le sens direct, on montre que toute sous-famille finie  $\sigma(C_{i_1}), \dots, \sigma(C_{i_p})$  est indépendante. Soit  $A_{i_k} \in C_{i_k}$  pour  $1 \leq k \leq p$ . On considère la famille

$$\mathcal{M}_1 = \{A_{i_1} \in \sigma(C_{i_1}) : \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_p})\}.$$

Il s'agit d'une classe monotone qui contient, par hypothèse,  $C_{i_1}$ . Elle contient donc la classe monotone engendrée par  $C_{i_1}$ , qui coïncide par le théorème de classe monotone (Th. 1.3.1) avec  $\sigma(C_{i_1})$ . On a donc pour tout  $A_{i_1} \in \sigma(C_{i_1})$  et  $A_{i_2} \in C_{i_2}, \dots, A_{i_p} \in C_{i_p}$  :

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_p}). \quad (5.1)$$

On fixe maintenant  $A_{i_1} \in \sigma(C_{i_1})$  et  $A_{i_k} \in C_{i_k}$  pour  $2 \leq k \leq p$ . On considère la famille

$$\mathcal{M}_2 = \{A_{i_2} \in \sigma(C_{i_2}) : \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_p})\}.$$

Il s'agit encore d'une classe monotone qui contient, d'après ce qui précède,  $C_{i_2}$ . Par le théorème des classes monotones,  $\mathcal{M}_2 = \sigma(C_{i_2})$  et (5.1) est valable maintenant pour  $A_{i_1} \in \sigma(C_{i_1})$ ,  $A_{i_2} \in \sigma(C_{i_2})$  et  $A_{i_3} \in C_{i_3}, \dots, A_{i_p} \in C_{i_p}$ . En continuant ce procédé par récurrence, on montre alors que (5.1) est valable pour  $A_{i_1} \in \sigma(C_{i_1})$ ,  $A_{i_2} \in \sigma(C_{i_2})$ ,  $A_{i_3} \in \sigma(C_{i_3}), \dots, A_{i_p} \in \sigma(C_{i_p})$ , ce qui est le résultat souhaité.  $\square$

**Théorème 5.1.1 (Coalitions)** Soient  $(\mathcal{F}_i)_{i \in I}$  des sous tribus de  $\mathcal{F}$ . On suppose que les  $\mathcal{F}_i$  sont indépendantes. Soit  $I = \bigcup_{a \in A} I_a$  une partition de l'ensemble d'indexation  $I$  ( $I_a \cap I_b = \emptyset$ ) et  $\mathcal{G}_a = \sigma(\bigcup_{i \in I_a} \mathcal{F}_i)$ . Alors les tribus  $(\mathcal{G}_a)_{a \in A}$  sont indépendantes.

Heuristiquement, le théorème des coalitions explique qu'en réunissant des tribus indépendantes, on construit de nouvelles tribus indépendantes.

**Démonstration :** Soit  $\mathcal{C}_a = \{B \in \mathcal{F} \text{ tel que } \exists J \text{ fini} \subset I_a, \exists A_i \in \mathcal{F}_i, i \in J, B = \bigcap_{i \in J} A_i\}$ .

Comme pour  $i \in I_a$ , on a  $\mathcal{F}_i \subset \mathcal{C}_a$  alors  $\bigcup_{i \in I_a} \mathcal{F}_i \subset \mathcal{C}_a \subset \mathcal{G}_a$ . On observe que

- Les  $(\mathcal{C}_a)_{a \in A}$  sont stables par intersection finie : si  $B \in \mathcal{C}_a$  et  $B' \in \mathcal{C}_a$  alors  $B = \bigcap_{i \in J} A_i$  et  $B' = \bigcap_{i \in J'} A'_i$  pour  $J, J'$  finis dans  $I_a$ . Alors  $B \cap B' = (\bigcap_{i \in J} A_i) \cap (\bigcap_{i \in J'} A'_i) = \bigcap_{i \in J \cup J'} A''_i$  où  $A''_i = A_i$  si  $i \in J$  et  $A''_i = A'_i$  si  $i \in J'$ .
- Les  $(\mathcal{C}_a)_{a \in A}$  sont clairement indépendants car  $B \in \mathcal{C}_a$  s'écrit comme une intersection finie d'événements de  $\mathcal{F}_i$ ,  $i \in I_a$  pour laquelle on peut utiliser la propriété d'indépendance. Ce sont des  $\pi$ -systèmes.

Par la Proposition 5.1.2 (conséquence du Théorème des classes monotones), l'indépendance est encore vraie pour les tribus engendrées  $\sigma(\mathcal{C}_a) = \mathcal{G}_a$ .  $\square$

Pour l'indépendance des variables aléatoires, on commence par rappeler qu'on associe à chaque variable aléatoire  $X$  sa tribu engendrée

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$$

Il s'agit de la plus petite tribu sur  $\Omega$  rendant l'application  $X : \Omega \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  mesurable, cf. Définition 1.1.4. Heuristiquement, la tribu  $\sigma(X)$  est la tribu qui contient toutes les informations liées à la variable aléatoire  $X$ .

**Définition 5.1.3 (Indépendance de variables aléatoires)**

- Deux variables aléatoires  $X$  et  $Y$  sont indépendantes si et seulement si leur tribu engendrée sont indépendantes. On note toujours  $X \perp\!\!\!\perp Y$ .
- Toute suite  $(X_i)_{i \in I}$  de variables aléatoires est indépendante si et seulement si leur tribu associée sont indépendantes.

## 5.2 Critères et exemples

Par définition des tribus engendrées par les variables aléatoires, il est immédiat de voir que des critères plus concrets d'indépendance sont :

**Proposition 5.2.1 (Indépendance de variables aléatoires)**

- **Indépendance de deux variables aléatoires.** Deux variables aléatoires  $X, Y$  sont dites indépendantes si pour  $A, B \in \mathcal{B}(\mathbb{R})$ , mesurables de  $\mathbb{R}$ , les évènements  $\{X \in A\}, \{Y \in B\}$  sont indépendants :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

- **Indépendance d'une famille finie de variables aléatoires.** Les  $m$  variables aléatoires  $X_1, \dots, X_m$  sont dites (mutuellement) indépendantes si pour tout boréliens  $A_1, \dots, A_m$ , les évènements  $\{X_1 \in A_1\}, \dots, \{X_m \in A_m\}$  sont mutuellement indépendants :

$$\mathbb{P}(X_1 \in A_1, \dots, X_m \in A_m) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_m \in A_m).$$

- **Indépendance d'une suite de variables aléatoires.** Une suite  $(X_i)_{i \in \mathbb{N}}$  de variables aléatoires est dite indépendante si toute sous-suite finie de  $(X_i)_{i \in \mathbb{N}}$ , la propriété précédente est vraie.

On a d'autres critères pour l'indépendance des variables aléatoires. Ils portent sur la structure de la loi du vecteur associé, considérée comme mesure dans l'espace produit  $\mathbb{R}^n$  (cf. la Définition 3.1.1).

**Proposition 5.2.2** Un vecteur aléatoire  $X = (X_1, \dots, X_d)$  est à composantes indépendantes si et seulement si sa loi  $\mathbb{P}_X$  est la loi produit (de ses lois marginales) :

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}. \quad (5.2)$$

**Démonstration :** Sens direct : soit  $B = B_1 \times \dots \times B_d$  pavé de  $\mathcal{B}(\mathbb{R}^d)$ , alors

$$\begin{aligned} \mathbb{P}_{(X_1, \dots, X_d)}(B) &= \mathbb{P}((X_1, \dots, X_d) \in B_1 \times \dots \times B_d) \\ &= \mathbb{P}(X_1 \in B_1, \dots, X_d \in B_d) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_d \in B_d) \\ &= \mathbb{P}_{X_1}(B_1) \times \dots \times \mathbb{P}_{X_d}(B_d) = (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d})(B_1 \times \dots \times B_d) \\ &= (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d})(B). \end{aligned}$$

Comme l'ensemble  $\mathcal{P}$  des pavés  $B = B_1 \times \dots \times B_d$  engendre  $\mathcal{B}(\mathbb{R})^{\otimes d} = \mathcal{B}(\mathbb{R}^d)$ , comme  $\mathcal{P}$  est stable par intersections finies puis comme  $\mathbb{P}_{(X_1, \dots, X_d)}$  et  $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$  coïncident sur  $\mathcal{P}$  alors ces mesures coïncident sur  $\mathcal{B}(\mathbb{R}^d)$  d'après le Théorème 1.3.2 (Th. de Dynkin).

Sens indirect : pour la réciproque, prendre  $B$  un pavé et remonter les étapes.  $\square$

**Corollaire 5.2.1** Les variables aléatoires  $X_1, \dots, X_d$  sont indépendantes si et seulement si pour tout  $x_1, \dots, x_d$ , on a

$$\mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbb{P}(X_1 \leq x_1) \times \dots \times \mathbb{P}(X_d \leq x_d).$$

**Démonstration :** Appliquer le résultat précédent avec la famille  $\mathcal{D}$  d'ensembles  $B = ]-\infty, x_1] \times \dots \times ]-\infty, x_d]$ . Comme cette famille  $\mathcal{D}$  est stable par intersections finies (ie. il s'agit d'un  $\pi$ -système) et engendre la tribu  $\mathcal{B}(\mathbb{R}^d)$ , le Théorème 1.3.2 (Th. de Dynkin) s'applique encore et conclut.  $\square$

**Remarque 5.2.1** Dans les cas discret et à densité, on peut préciser les critères d'indépendances :

- Les variables aléatoires discrètes  $X$  et  $Y$  sont indépendantes si et seulement si

$$\forall x_i \in \mathcal{S}(X), \forall y_j \in \mathcal{S}(Y), \quad \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j). \quad (5.3)$$

En effet, si  $X \perp\!\!\!\perp Y$  alors le choix  $A = \{x_i\}$  et  $B = \{y_j\}$  donne (5.3). Réciproquement si (5.3) est vraie alors pour tout  $A, B \in \mathcal{B}(\mathbb{R})$ , on a

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \sum_{\substack{x_i \in A \\ y_j \in B}} \mathbb{P}(X = x_i, Y = y_j) = \sum_{\substack{x_i \in A \\ y_j \in B}} \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j) \\ &= \left( \sum_{x_i \in A} \mathbb{P}(X = x_i) \right) \left( \sum_{y_j \in B} \mathbb{P}(Y = y_j) \right) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \end{aligned}$$

- Les variables aléatoires  $X, Y$  de densités respectives  $f$  et  $g$  sont indépendantes si et seulement si le couple  $(X, Y)$  est de densité

$$f \otimes g : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto f(x)g(y) \quad (5.4)$$

ie. on peut factoriser la densité  $f(x, y)$  en deux facteurs l'un en  $x$ , l'autre en  $y$ . En effet si (5.4) est vraie alors pour  $A, B \in \mathcal{B}(\mathbb{R})$ , on a

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}_{(X,Y)}(A \times B) = \int_{A \times B} f(x)g(y)dxdy \\ &= \left( \int_A f(x)dx \right) \left( \int_B g(y)dy \right) \quad (\text{Th. Fubini-Tonelli}) \\ &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \end{aligned}$$

Réciproquement si  $X \perp\!\!\!\perp Y$  alors de la même façon on a

$$\mathbb{P}_{(X,Y)}(A \times B) = \int_{A \times B} f(x)g(y)dxdy = \mu(A \times B)$$

où  $\mu$  est la mesure sur  $\mathcal{B}(\mathbb{R}^2)$  donnée par  $\mu(C) = \int_C f(x)g(y)dxdy$ , ie.  $\mu$  est la mesure de densité  $f(x)g(y)$ . On a alors  $\mathbb{P}_{(X,Y)}$  et  $\mu$  qui coïncident sur la famille  $\mathcal{P}$  des produits de mesurables  $A \times B$ ,  $A, B \in \mathcal{B}(\mathbb{R})$ . Comme  $\mathcal{P}$  est stable par intersections finies et  $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R}^2)$ , on a  $\mathbb{P}_{(X,Y)} = \mu$ , c'est à dire  $(X, Y)$  admet pour densité  $f(x)g(y)$  (Théorème 1.3.2 de Dynkin).

**Remarque 5.2.2** Une conséquence importante : si on connaît les lois de  $X$  et  $Y$ , des variables supposées **indépendantes**, on peut reconstruire la loi du couple  $(X, Y)$  à partir des marginales par (5.2) (dans le cas discret par (5.3) et dans le cas à densité par (5.4)). Rappelons encore que ce n'est pas vrai en général quand  $X$  et  $Y$  ne sont pas indépendantes.

Dans les deux exemples de la page 34,  $X$  et  $Y$  ne sont pas indépendantes car par exemple pour le premier :

$$\mathbb{P}(X = 2, Y = 2) = 0,2, \quad \text{tandis que} \quad \mathbb{P}(X = 2) \times \mathbb{P}(Y = 2) = 0,4 \times 0,25 = 0,1.$$

et pour le second :

$$\mathbb{P}(X = 3, Y = 5) = 0, \quad \text{tandis que} \quad \mathbb{P}(X = 3) \times \mathbb{P}(Y = 5) = 0,3 \times 0,2 = 0,06.$$

**Exemples :** • On donne le tableau de la loi d'un couple  $(X, Y)$  en donnant les probabilités ponctuelles  $\mathbb{P}(X = x_i, Y = y_j)$  :

$X \setminus Y$	$y_1$	$y_2$	$y_3$	
$x_1$	0,12	0,08	0,20	0,4
$x_2$	0,18	0,12	0,30	0,6
	0,3	0,2	0,5	= 1

On vérifie ici que  $X$  et  $Y$  sont indépendantes car pour tout  $i = 1, 2$  et  $j = 1, 2, 3$ , on a

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j).$$

• Considérons le couple  $(X, Y)$  de loi donnée par la densité  $f_{(X,Y)}(x, y) = \frac{1}{3}\mathbf{1}_{[0,1] \times [-1,2]}(x, y)$ . On a vu que  $X$  et  $Y$  avaient pour densité  $f_X(x) = \mathbf{1}_{[0,1]}(x)$  et  $f_Y(y) = \frac{1}{3}\mathbf{1}_{[-1,2]}(y)$ . On a alors

$$f_{(X,Y)}(x, y) = \frac{1}{3}\mathbf{1}_{[0,1] \times [-1,2]}(x, y) = \mathbf{1}_{[0,1]}(x) \times \frac{1}{3}\mathbf{1}_{[-1,2]}(y) = f_X(x)f_Y(y).$$

Les variables  $X$  et  $Y$  sont donc indépendantes.

• Soit  $(X, Y)$  le couple aléatoire de loi donnée par la densité  $f_{(X,Y)}(x, y) = \frac{1}{3\pi}e^{-\frac{x^2+2xy+5y^2}{6}}$ . On a vu que les densités marginales sont

$$f_X(x) = \frac{1}{\sqrt{15\pi/2}}e^{-\frac{4x^2}{30}}, \quad f_Y(y) = \frac{1}{\sqrt{3\pi/2}}e^{-\frac{4y^2}{6}}.$$

On a alors

91

$$f_X(x)f_Y(y) = \frac{1}{\sqrt{15\pi/2}}e^{-\frac{4x^2}{30}} \times \frac{1}{\sqrt{3\pi/2}}e^{-\frac{4y^2}{6}} \neq \frac{1}{3\pi}e^{-\frac{x^2+2xy+5y^2}{6}} = f_{(X,Y)}(x,y).$$

Dans ce cas,  $X$  et  $Y$  ne sont pas indépendantes.

**Proposition 5.2.3** Soient  $X_i, i \in I$ , des variables aléatoires. Elles sont (mutuellement) indépendantes si et seulement si pour toute famille finie  $J \subset I$  et toutes fonctions boréliennes  $h_j : \mathbb{R} \rightarrow \mathbb{C}$  telles que  $h_j(X_j) \in L^1$  pour tout  $j \in J$ , on a

$$\mathbb{E}\left[\prod_{j \in J} h_j(X_j)\right] = \prod_{j \in J} \mathbb{E}[h_j(X_j)]. \quad (5.5)$$

**Démonstration :** • Dans le sens direct, on utilise le théorème de transfert, l'indépendance et Fubini. Supposons que  $J$  est fini et (pour simplifier les notations) égale à  $\{1, \dots, n\}$  :

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^n h_j(X_j)\right] &= \int \prod_{j=1}^n h_j(x_j) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \quad (\text{théorème de transfert}) \\ &= \int \prod_{j=1}^n h_j(x_j) \prod_{j=1}^n d\mathbb{P}_{X_j}(dx_j) \quad (\text{indépendance : Prop. 5.2.2}) \\ &= \int \prod_{j=1}^n (h_j(x_j) d\mathbb{P}_{X_j}(dx_j)) = \prod_{j=1}^n \int h_j(x_j) d\mathbb{P}_{X_j}(dx_j) \quad (\text{théorème de Fubini}) \\ &= \prod_{j \in J} \mathbb{E}[h_j(X_j)] \quad (\text{théorème de transfert}). \end{aligned}$$

- La réciproque s'obtient en écrivant (5.5) pour  $h_j = \mathbf{1}_{A_j}$ ,  $j \in J$ . □

**Corollaire 5.2.2** Soient  $X_1, \dots, X_n$  des variables aléatoires réelles (ou vecteurs aléatoires) indépendantes avec des moments d'ordre 1 finis, alors on a :

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

**Démonstration :** Appliquer le résultat précédent avec les fonctions  $h_j(x) = x$ . □

**La réciproque est fausse :** soient  $X_1$  de loi uniforme sur  $[-1, 1]$  et  $X_2 = X_1^2$ , on a  $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1^3] = \int_{-1}^1 x_1^3 dx_1 = 0$  et  $\mathbb{E}[X_1] = 0$  si bien qu'on a  $\mathbb{E}[X_1 X_2] = 0 = \mathbb{E}[X_1]\mathbb{E}[X_2]$  mais  $X_1$  et  $X_2$  ne sont pas indépendantes car par exemple

$$\mathbb{P}(X_1 \in [0, 1/2], X_2 \in [0, 1/4]) = 1/4, \quad \mathbb{P}(X_1 \in [0, 1/2]) \times \mathbb{P}(X_2 \in [0, 1/4]) = 1/8.$$

On a aussi :

**Proposition 5.2.4** Si deux variables aléatoires  $X$  et  $Y$  sont indépendantes alors

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Plus généralement, pour  $n$  variables aléatoires  $X_1, \dots, X_n$  (mutuellement) indépendantes on a

$$\varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t).$$

**Démonstration :** Appliquer le résultat précédent avec les fonctions  $h_j(x) = e^{itx}$ . On a

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}]$$

d'après la Proposition 5.2.3. La généralisation au cas de  $n$  variables est immédiate.  $\square$

En fait, on a aussi un critère d'indépendance de deux variables aléatoires  $X, Y$  en terme de fonction caractéristique :

**Théorème 5.2.1** Deux variables aléatoires  $X$  et  $Y$  sont indépendantes si et seulement si pour tout  $t, s \in \mathbb{R}$

$$\varphi_{(X,Y)}(t, s) = \varphi_X(t)\varphi_Y(s).$$

Le même type de résultat est vrai pour la fonction génératrice (cf. Proposition 4.4.2) et pour la transformée de Laplace (cf. Proposition 4.4.4) de la Section 4.4.

**Démonstration :** On raisonne par équivalence :

$$X \perp\!\!\!\perp Y \iff \mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y \iff \varphi_{\mathbb{P}_{(X,Y)}}(t, s) = \varphi_{\mathbb{P}_X \otimes \mathbb{P}_Y}(t, s)$$

où on utilise que  $\mathbb{P}_{(X,Y)}$  et  $\mathbb{P}_X \otimes \mathbb{P}_Y$  coïncident si et seulement si ces mesures ont même fonction caractéristique. On observe alors que  $\varphi_{\mathbb{P}_{(X,Y)}}(t, s) = \varphi_{(X,Y)}(t, s)$  et

$$\begin{aligned} \varphi_{\mathbb{P}_X \otimes \mathbb{P}_Y}(t, s) &= \int_{\mathbb{R}^2} e^{i(tx+sy)} \mathbb{P}_X(dx) \mathbb{P}_Y(dy) \\ &= \int_{\mathbb{R}} e^{itx} \mathbb{P}_X(dx) \int_{\mathbb{R}} e^{isy} \mathbb{P}_Y(dy) \\ &= \int_{\Omega} e^{itX} d\mathbb{P} \int_{\Omega} e^{isY} d\mathbb{P} \\ &= \varphi_X(t)\varphi_Y(s) \end{aligned}$$

en utilisant les théorèmes de Fubini et de transfert. On a donc

$$X \perp\!\!\!\perp Y \iff \varphi_{(X,Y)}(t, s) = \varphi_X(t)\varphi_Y(s).$$

$\square$

**Remarque 5.2.1** 1. Le critère d'indépendance se généralise immédiatement au cas de  $n$  variables : les variables aléatoires  $X_1, \dots, X_n$  sont (mutuellement) indépendantes si et seulement si pour tout  $t_1, \dots, t_n \in \mathbb{R}$

$$\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \varphi_{X_1}(t_1) \dots \varphi_{X_n}(t_n).$$

2. Les mêmes résultats que la Proposition 5.2.4 et le Théorème 5.2.1 sont vrais pour la fonction génératrice ou la transformée de Laplace.

## 5.3 Non-corrélation et indépendance

**Définition 5.3.1 (Non-corrélation)** Deux variables aléatoires  $X, Y$  de variances finies sont dites non corrélées si  $\text{Cov}(X, Y) = 0$ .

**Proposition 5.3.1** Soient  $X$  et  $Y$  deux vecteurs aléatoires indépendants de variances finies. Alors  $\text{Cov}(X, Y) = 0$ . Autrement dit l'indépendance implique la non-corrélation.

**Démonstration :** Comme par indépendance  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ , le résultat vient de l'identité de Koenig pour la covariance  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$ .  $\square$

La réciproque est fausse : si  $X$  et  $Y$  sont de covariance nulle alors ils ne sont pas nécessairement indépendants. Le même contre-exemple que pour le Corollaire 5.2.2 s'applique. Cependant dans le cas de variables aléatoires  $X, Y$  gaussiennes, on verra que la réciproque est vraie, cf. Proposition 9.2.5 au Chapitre 9.

Pour la somme d'une variance, on déduit sous indépendance une propriété remarquable :

**Corollaire 5.3.1** Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes avec des moments d'ordre deux finis alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Bien entendu, il suffit de supposer  $X, Y$  non corrélées.

**Démonstration :** C'est immédiat puisque d'après la Proposition 3.5.2, on a

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

et  $\text{Cov}(X, Y) = 0$  quand  $X \perp\!\!\!\perp Y$ .  $\square$

De façon générale, il vient :

**Proposition 5.3.2 (Identité de Bienaymé)** Si  $X_1, \dots, X_d$  sont des variables aléatoires de variances finies et deux à deux non corrélées alors

$$\text{Var}\left(\sum_{j=1}^d X_j\right) = \sum_{j=1}^d \text{Var}(X_j).$$

**Démonstration :** La non-corrélation de  $X_j, X_k$  signifie que les variables aléatoires  $X_j - \mathbb{E}[X_j]$  et  $X_k - \mathbb{E}[X_k]$  sont orthogonales (pour le produit scalaire associé à la covariance). On en déduit que

$$\begin{aligned}\text{Var}\left(\sum_{j=1}^d X_j\right) &= \sum_{j,k=1}^d \text{Cov}(X_j, X_k) \\ &= \sum_{j=1}^d \text{Cov}(X_j, X_j) + \sum_{1 \leq j, k \leq d} \text{Cov}(X_j, X_k) \\ &= \sum_{j=1}^d \text{Var}(X_j).\end{aligned}$$

□

En combinant l'identité de Bienaymé avec l'inégalité de Tchebychev, on a de suite :

**Proposition 5.3.3 (Inégalité de Bienaymé-Tchebychev)** *Si  $X_1, \dots, X_d$  sont des variables aléatoires  $L^2$  deux à deux non corrélées alors*

$$\mathbb{P}\left(\left|\sum_{j=1}^d (X_j - \mathbb{E}[X_j])\right| \geq t\right) \leq \frac{1}{t^2} \sum_{j=1}^d \text{Var}(X_j), \quad t > 0.$$

**Démonstration :** En notant  $S_n = \sum_{j=1}^d X_j$ , pour  $t > 0$ , on a :

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{j=1}^d X_j - \mathbb{E}[X_j]\right| \geq t\right) &= \mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq t) \\ &\leq \frac{\text{Var}(S_n)}{t^2} \quad (\text{inégalité de Tchebychev : Th 3.5.1}) \\ &\leq \frac{\sum_{j=1}^n \text{Var}(X_j)}{t^2} \quad (\text{inégalité de Bienaymé : Prop. 5.3.3}).\end{aligned}$$

## 5.4 Évènements asymptotiques

### 5.4.1 Tribus du futur et tribu asymptotique

**Définition 5.4.1** Soit  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  une suite de tribus de  $\mathcal{F}$ .

- Si la suite est croissante, ie.  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ , on parle de filtration.
- On appelle tribus du futur :  $\mathcal{F}^n = \sigma(\bigcup_{k \geq n} \mathcal{F}_k)$ ,  $n \in \mathbb{N}$ .
- On appelle tribu asymptotique :  $\mathcal{F}^\infty = \bigcap_{n \in \mathbb{N}} \mathcal{F}^n$ .

D'habitude on a  $\mathcal{F}_n \subset \mathcal{F}_{n+1}$  et la famille de tribus croissantes  $(\mathcal{F}_n)_{n \in \mathbb{N}}$  s'appelle une filtration. Dans ce cas, heuristiquement, si  $\mathbb{N}$  désigne le temps et  $n \in \mathbb{N}$  la date  $n$ , la tribu  $\mathcal{F}_n$  représente la tribu des évènements du passé par rapport au temps  $n$  (elle comporte toutes les informations sur ce qu'il s'est passé jusqu'au temps  $n$ ), la tribu  $\mathcal{F}^n$  représente la tribu des évènements du futur par rapport au temps  $n$  (elle comporte toutes les informations sur ce qu'il se passera après le temps  $n$ ). La tribu  $\mathcal{F}^\infty$  représente alors celle des évènements asymptotiques.

Typiquement, quand on a une suite de variables aléatoires  $(X_n)_{n \geq 1}$ , on lui associe sa filtration naturelle  $(\mathcal{F}_n)_{n \geq 1}$  en considérant  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ .

**Exemples.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes. On considère la suite de tribus  $\mathcal{F}_n = \sigma(X_n)$ .

- Notons  $A$  l'évènement «  $X_n$  repasse par 0 une infinité de fois ». Alors  $A$  est dans la tribu asymptotique des  $\mathcal{F}_n$ . En effet,

$$A = \limsup_{n \rightarrow +\infty} \{X_n = 0\} = \bigcap_n \bigcup_{k \geq n} \{X_k = 0\}$$

et  $\bigcup_{k \geq n} \{X_k = 0\} \in \mathcal{F}^n$ .

- $B = \{\text{les } X_n \text{ sont bornées par } M\}$  n'est pas dans  $\mathcal{F}^\infty$  car par exemple l'évènement  $\{X_1 \text{ bornée par } M\}$  n'est pas dans  $\mathcal{F}^2$  ni dans aucun  $\mathcal{F}^n$  ( $n \geq 2$ ).
- $C = \{\text{la suite } (X_n)_{n \geq 0} \text{ converge}\} \in \mathcal{F}^\infty$ . Heuristiquement, la convergence de la suite  $X_n$  est un résultat asymptotique (il ne concerne que les  $X_n$  pour  $n$  arbitrairement grand mais il ne concerne aucun  $X_n$  pour  $n$  fini) donc il est dans la tribu asymptotique. Plus précisément, comme dans  $\mathbb{R}$  la convergence d'une suite est équivalente à dire que la suite est de Cauchy,  $\omega \in C$  s'écrit :

$$\forall r \in \mathbb{N}, \exists N \in \mathbb{N}, \forall p, q \geq N, |X_p(\omega) - X_q(\omega)| < 1/r.$$

On a donc l'écriture suivante de  $C$  :

$$C = \bigcap_{r \in \mathbb{N}} \bigcup_{N \in \mathbb{N}} \bigcap_{p, q \geq N} \{|X_p - X_q| < 1/r\}$$

avec  $\{|X_p - X_q| < 1/r\} \in \mathcal{F}^p$  (si  $p < q$ ).

Le résultat sur les tribus asymptotiques est le suivant :

**Théorème 5.4.1 (Loi du 0/1 de Kolmogorov)** *Soit  $(\mathcal{F}_n)_{n \geq 1}$  une suite de tribus indépendantes alors la tribu asymptotique est triviale : c'est à dire*

$$\forall A \in \mathcal{F}^\infty, \quad \mathbb{P}(A) = 0 \text{ ou } 1.$$

**Démonstration :** On montre que  $\mathcal{F}^\infty$  est indépendante de  $\mathcal{F}^\infty$ . Ainsi si  $A \in \mathcal{F}^\infty$ , on a  $A = A \cap A$  et par indépendance de  $A$  avec  $A$  :  $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ , d'où  $\mathbb{P}(A) = 0$  ou 1.

Comme la suite de tribu  $(\mathcal{F}_n)_{n \geq 1}$  est indépendante, on peut appliquer le théorème des coalitions,

$$\mathcal{F}^n = \sigma \left( \bigcup_{k \geq n} \mathcal{F}_k \right) \perp \sigma \left( \bigcup_{k < n} \mathcal{F}_k \right).$$

Comme  $\mathcal{F}^\infty \subset \mathcal{F}^n$ , on a donc  $\mathcal{F}^\infty \perp \sigma(\bigcup_{k < n} \mathcal{F}_k)$ .

Soit  $\mathcal{C} = \bigcup_{n \geq 0} \sigma(\bigcup_{k < n} \mathcal{F}_k)$ . Il s'agit d'un  $\pi$ -système car  $\mathcal{C}$  est stable par intersection finie : En effet, soient  $A, B \in \mathcal{C}$ , on a  $A \in \sigma(\bigcup_{k < n} \mathcal{F}_k)$  et  $B \in \sigma(\bigcup_{k < n'} \mathcal{F}_k)$ . En notant  $N = \max(n, n')$ , on a  $A, B \in \sigma(\bigcup_{k < N} \mathcal{F}_k)$ . Par stabilité d'une tribu par intersection, on a aussi  $A \cap B \in \sigma(\bigcup_{k < N} \mathcal{F}_k)$  et donc  $A \cap B \in \mathcal{C}$ .

Comme  $\mathcal{C}$  est un  $\pi$ -système et  $\mathcal{C} \perp \mathcal{F}^\infty$ , on déduit  $\sigma(\mathcal{C}) \perp \mathcal{F}^\infty$  par la Proposition 5.1.2.

Pour tout  $k$ , on a  $\mathcal{F}_k \subset \mathcal{C}$  donc aussi  $\bigcup_{k \geq 1} \mathcal{F}_k \subset \mathcal{C}$ . Ainsi la tribu  $\sigma(\mathcal{C})$  contient  $\sigma(\bigcup_{k \geq 1} \mathcal{F}_k) = \mathcal{F}^1$  donc aussi  $\mathcal{F}^\infty$ .

De  $\sigma(\mathcal{C}) \perp \mathcal{F}^\infty$ , on déduit alors que  $\mathcal{F}^\infty \perp \mathcal{F}^\infty$ , ce qui permet de conclure que pour tout  $A \in \mathcal{F}^\infty$ , on a  $A \perp A$  et donc  $\mathbb{P}(A) = 0$  ou 1.  $\square$

**Exemples.** Dans les exemples précédents, avec  $\mathcal{F}_n = \sigma(X_n)$ ,  $n \geq 1$ , tribus indépendantes, on déduit de la loi de Kolmogorov que  $\mathbb{P}(A) = 0$  ou 1, de même pour  $\mathbb{P}(C)$ . Par contre, on ne peut rien dire de  $\mathbb{P}(B)$ .

### 5.4.2 Lemmes de Borel-Cantelli

Les lemmes de Borel-Cantelli complètent d'une certaine façon la loi du 0/1 de Kolmogorov en donnant des cas où la probabilité vaut 1 et d'autres où elle vaut 0. Ces résultats concernent les liminf et limsup d'événements pour lesquelles on renvoie à la Section 1.6.

**Théorème 5.4.2 (Lemme de Borel-Cantelli n°1)** Soit  $(A_n)_{n \geq 0}$  une suite (infinie) d'événements. On suppose que  $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty$ . Alors

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 0.$$

C'est à dire presque sûrement, seuls un nombre fini d'événements  $A_n$  sont réalisés.

**Démonstration :** Par définition,  $\limsup_{n \rightarrow +\infty} A_n \subset \bigcup_{m \geq n} A_m$  pour tout  $n \geq 0$ . On a donc

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) \leq \mathbb{P}\left(\bigcup_{m \geq n} A_m\right) \leq \sum_{m \geq n} \mathbb{P}(A_m). \quad (5.6)$$

Or  $\sum_{m \geq n} \mathbb{P}(A_m)$  est le reste d'ordre  $n$  de la série convergente  $\sum_{m \geq 1} \mathbb{P}(A_m)$ . Comme le reste d'une série convergente tend vers 0, un passage à la limite  $n \rightarrow +\infty$  dans (5.6) conclut.  $\square$

Ce résultat admet un complément quand les événements  $A_n$ ,  $n \geq 0$ , sont indépendants :

**Théorème 5.4.3 (Lemme de Borel-Cantelli n°2)** Soit  $(A_n)_{n \geq 0}$  une suite (infinie) d'évènements indépendants. On suppose que  $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$ . Alors

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 1.$$

C'est à dire presque sûrement, une infinité d'évènements  $A_n$ ,  $n \geq 1$ , sont réalisés.

**Remarque 5.4.1** Le résultat reste vrai si l'indépendance des évènements  $A_n$  n'est vraie que deux à deux .

**Démonstration :** D'abord, on a

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_n\right) = 1 - \mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n^c\right).$$

Puis par monotonie séquentielle croissante en (5.7), décroissante en (5.8) et comme les évènements  $A_n^c$  sont aussi indépendants (en (5.9)), on a :

$$\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n^c\right) = \mathbb{P}\left(\bigcup_{n \geq 0} \bigcap_{k \geq n} A_k^c\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right) \quad (5.7)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \mathbb{P}\left(\bigcap_{m=n}^q A_m^c\right) \quad (5.8)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \prod_{m=n}^q \mathbb{P}(A_m^c) \quad (5.9)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \prod_{m=n}^q (1 - \mathbb{P}(A_m)).$$

Mais, pour tout  $x \in \mathbb{R}$ , on a  $1 - x \leq e^{-x}$ , il vient donc

$$0 \leq \prod_{m=n}^q (1 - \mathbb{P}(A_m)) \leq \exp\left(-\sum_{m=n}^q \mathbb{P}(A_m)\right).$$

Comme  $\sum_{m \geq 0} \mathbb{P}(A_m) = +\infty$ , le passage à la limite  $q \rightarrow +\infty$  donne  $\lim_{q \rightarrow +\infty} \prod_{m=n}^q (1 - \mathbb{P}(A_m)) = 0$ . On a donc  $\mathbb{P}\left(\liminf_{n \rightarrow +\infty} A_n^c\right) = 0$ , ce qui conclut.  $\square$

## Applications

Dans la suite, on note i.s. pour infiniment souvent, ainsi  $\{A_n\text{ i.s.}\} = \limsup_{n \rightarrow +\infty} A_n$ .

- Singe dactylographique de Borel : presque sûrement, un singe qui écrit frénétiquement sur une machine à écrire finira par écrire la Bible. Notons  $N$  le nombre de caractères de la Bible, en supposant que le singe frappe sur chaque touche de la machine à écrire avec une probabilité uniforme, chaque touche est frappée avec probabilité  $1/26$  (ou –en considérant les touches de ponctuation–  $1/M$ ,  $M$  étant le nombre total de touche). Taper la Bible est alors un évènement de probabilité  $(1/M)^N$ . Les évènements  $A_i = \{\text{taper la Bible}\}$  sont des évènements avec  $\sum_{n \geq 1} \mathbb{P}(A_n) = \sum_{n \geq 0} M^{-N} = +\infty$ , ils sont indépendants s'ils considèrent des frappes différentes. D'après le lemme de Borel-Cantelli,  $\limsup_n A_n$  est presque sûr, ie. ps  $A_n$  est réalisé infiniment souvent : le singe tape donc même presque sûrement une infinité de fois la Bible.
- Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles telles que, pour un  $M \in \mathbb{R}$ , on ait  $\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \geq M) < +\infty$ . Alors

$$\mathbb{P}(X_n \geq M \text{ i.s.}) = 0 \iff \mathbb{P}\left(\liminf_n \{X_n < M\}\right) = 1.$$

Donc  $\limsup_n X_n < M$  presque sûrement.

De la même façon si  $\sum_{n \in \mathbb{N}} \mathbb{P}(X_n \leq M) < +\infty$  alors  $\liminf_n X_n > M$  presque sûrement.

- On lance une pièce équilibrée une infinité de fois. Quelle est la probabilité d'obtenir une infinité de fois deux piles consécutifs ?
- Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes de Bernoulli de paramètre  $1/2$  et posons  $A_n = \{X_n = X_{n+1} = 1\}$ . On s'intéresse à  $\mathbb{P}(A_n \text{ i.s.})$ . La suite d'évènement  $(A_n)_{n \in \mathbb{N}}$  n'est pas une suite indépendante car  $X_{n+1}$  intervient à la fois dans l'évènement  $A_n$  et dans l'évènement  $A_{n+1}$ . Par contre  $(A_{2n})_{n \in \mathbb{N}}$  est une suite d'évènements indépendants avec  $\mathbb{P}(A_{2n}) = 1/4$  pour tout  $n$ .
- On peut alors appliquer la deuxième partie du lemme de Borel-Cantelli (Th. 5.4.3) pour avoir  $\mathbb{P}(A_{2n} \text{ i.s.}) = 1$ . Comme  $\{A_{2n} \text{ i.s.}\} \subset \{A_n \text{ i.s.}\}$ , on conclut que  $\mathbb{P}(A_n \text{ i.s.}) = 1$ .

# Chapitre 6

## Somme de deux variables aléatoires indépendantes

Les sommes de variables aléatoires jouent un rôle important en probabilité. Ainsi typiquement,  $X + Y$  modélise les effets cumulés de  $X$  et de  $Y$ . Souvent  $X, Y$  représentent des phénomènes indépendants et on considère alors que les variables aléatoires sont indépendantes. Quand  $X, Y$  sont identiquement distribuées et indépendantes (*iid*), la somme  $X + Y$  représente les effets cumulés d'un phénomène qui se répète indépendamment. Les sommes  $X_1 + \dots + X_n$  pour des variables *iid* représentent alors les effets cumulés de  $n$  répétitions indépendantes du phénomène.

Remarquons que pour des variables aléatoires indépendantes de même loi et de carré intégrable, on a, par l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P} \left( \left| \sum_{j=1}^n (X_j - \mathbb{E}[X_j]) \right| \geq t\sqrt{n} \right) \leq \frac{\text{Var}(X_1)}{t^2}.$$

Cela suggère que l'ordre de grandeur de la somme  $\sum_{j=1}^n (X_j - \mathbb{E}[X_j])$  est au plus en  $\sqrt{n}$ . On peut ainsi dire que la somme  $\sum_{j=1}^n X_j$  ressemble à un terme déterministe  $n\mathbb{E}[X_1]$  de l'ordre de  $n$  (si  $\mathbb{E}[X_1] \neq 0$ ) plus un terme aléatoire de l'ordre au plus  $\sqrt{n}$ . On va évaluer dans ce chapitre la loi de  $\sum_{j=1}^n X_j$ . La loi du terme aléatoire sera étudiée plus en détails dans les prochains chapitres à l'aide de la LGN et du TCL.

### 6.1 Convolution de mesures

On considère un espace vectoriel mesurable  $(E, \mathcal{A})$ . Il y a donc sur  $E$  une structure d'espace vectoriel et une tribu  $\mathcal{A}$ . L'exemple typique que l'on considérera dans la suite est  $\mathbb{R}$  ou  $\mathbb{R}^n$ .

**Définition 6.1.1 (Convolution)** Soient  $\mu$  et  $\nu$  deux mesures sur un espace vectoriel mesurable  $(E, \mathcal{A})$ . La convolée  $\mu * \nu$  de ces deux mesures est

$$\mu * \nu(A) = \int_E \mu(A - x) d\nu(x) \tag{6.1}$$

où  $A - x = \{a - x : a \in A\}$ .

100

On vérifie facilement que  $\mu * \nu$  est une mesure (la  $\sigma$ -additivité vient de celle de  $\mu$  et du théorème de convergence monotone appliquée à l'intégrale en  $\nu$ ). Dans la suite, pour pouvoir appliquer le théorème de Fubini, on considère des mesures  $\sigma$ -finies (ce sera bien le cas des lois de probabilité considérées dans la suite).

**Proposition 6.1.1** *La convolution est commutative :  $\mu * \nu = \nu * \mu$ .*

**Démonstration :** En effet par le théorème de Fubini, on a

$$\mu * \nu(A) = \int \int \mathbf{1}_A(x + y) d\nu(x)d\mu(y).$$

On a obtenu une expression symétrique en  $\mu$  et  $\nu$ , ce qui justifie la proposition.  $\square$

**Proposition 6.1.2** *La mesure  $\delta_0$  est l'élément neutre de la convolution des mesures :  $\mu * \delta_0 = \mu$ .*

**Démonstration :** En effet pour  $A \in \mathcal{A}$  :

$$\mu * \delta_0(A) = \int_E \delta_0(A - x)d\mu(x) = \int_E \mathbf{1}_A d\mu(x) = \mu(A)$$

car  $\delta_0(A - x) = 1$  si et seulement si  $0 \in A - x$  c'est à dire si et seulement si  $x \in A$ , cela vaut 0 sinon, ie.  $\delta_0(A - x) = \mathbf{1}_A(x)$ .  $\square$

**Proposition 6.1.3** *Si  $\mu$  et  $\nu$  sont des mesures de probabilité alors  $\mu * \nu$  l'est aussi.*

**Démonstration :**  $\mu * \nu$  est une mesure de poids

$$\mu * \nu(E) = \int_E \mu(E - x)d\nu(x) = \int_E d\nu(x) = \nu(E) = 1$$

car  $E - x = E$  et donc  $\mu(E - x) = \mu(E) = 1$ . La mesure  $\mu * \nu$  est donc de poids total égale à 1 : c'est une mesure de probabilité.  $\square$

De la même façon, on montre que si  $\mu$  et  $\nu$  sont des mesures finies alors  $\mu * \nu$  aussi mais avec  $\mu * \nu(E) = \mu(E)\nu(E)$ .

## 6.2 Loi d'une somme de variables aléatoires à densité indépendantes

Désormais, nous considérons un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  sur lequel sont définies des variables aléatoires  $X, Y, \dots$ . Leur lois  $\mathbb{P}_X, \mathbb{P}_Y, \dots$  sont des mesures sur l'espace vectoriel  $\mathbb{R}$  qu'on va convoler.

**Proposition 6.2.1** *Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de loi  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ . Alors la loi de  $X + Y$  est  $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$ .*

**Démonstration :** Soit  $A \in \mathcal{B}(\mathbb{R})$ , on a

$$\begin{aligned}\mathbb{P}_{X+Y}(A) &= \mathbb{P}(X + Y \in A) = \mathbb{E}[\mathbf{1}_A(X + Y)] = \int \int \mathbf{1}_A(x + y) d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int \int \mathbf{1}_A(x + y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int \left( \int \mathbf{1}_{A-y}(x) \mathbb{P}_X(dx) \right) \mathbb{P}_Y(y) \\ &= \int \mathbb{P}_X(A - y) \mathbb{P}_Y(dy) = \mathbb{P}_X * \mathbb{P}_Y(A).\end{aligned}$$

□

Par une récurrence immédiate, on connaît donc la loi d'une somme de variables aléatoires indépendantes  $X_1 + \dots + X_n$  lorsqu'on a celles des termes  $X_i$  de la somme. Lorsqu'il n'y a pas indépendance, ce n'est pas suffisant : il faut connaître l'imbrication des lois c'est à dire la loi du vecteur associé  $(X_1, \dots, X_n)$ .

Ainsi de façon générale, la loi d'une somme  $X_1 + \dots + X_n$  de variables aléatoires non indépendantes est donnée par l'expression

$$\mathbb{P}_{X_1+\dots+X_n}(A) = \mathbb{P}(X_1 + \dots + X_n \in A) = \int \int \mathbf{1}_A(x_1 + \dots + x_n) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n).$$

**Remarque 6.2.1 (Produits tensoriel et de convolution)** Étant données deux mesures  $\mu$  et  $\nu$  sur un espace  $(E, \mathcal{A})$ , il ne faut pas confondre la mesure produit  $\mu \otimes \nu$  et la mesure produit de convolution  $\mu * \nu$ .

— La mesure produit  $\mu \otimes \nu$  est la mesure sur l'espace produit  $E^2 = E \times E$  définie par

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), \quad \forall A \times B \in \mathcal{A} \otimes \mathcal{A}.$$

Dans le contexte probabiliste, si  $\mu$  et  $\nu$  sont les lois de probabilité de variables aléatoires  $X, Y$  alors  $\mu \otimes \nu$  est la loi du vecteur aléatoire  $(X, Y)$  lorsque  $X, Y$  sont indépendantes.

— La mesure produit de convolution  $\mu * \nu$  est la mesure sur l'espace  $E$  définie par (6.1). Dans le contexte probabiliste, si  $\mu$  et  $\nu$  sont les lois de probabilités de variables aléatoires  $X, Y$  indépendantes alors  $\mu * \nu$  est la loi de la variable aléatoire  $X + Y$ .

### Remarque 6.2.2 (Interprétation probabiliste des propriétés de la convolution)

L'interprétation probabiliste permet de retrouver facilement plusieurs propriétés de la convolution :

- (élément neutre)  $\delta_0 * \mathbb{P}_X = \mathbb{P}_X (0 + X = X)$ .
- (commutativité)  $\mathbb{P}_X * \mathbb{P}_Y = \mathbb{P}_Y * \mathbb{P}_X (X + Y = Y + X)$ .
- (associativité)  $\mathbb{P}_X * (\mathbb{P}_Y * \mathbb{P}_Z) = (\mathbb{P}_X * \mathbb{P}_Y) * \mathbb{P}_Z (X + (Y + Z) = (X + Y) + Z)$ .
- (distributivité)  $\mathbb{P}_X * (\lambda \mathbb{P}_Y + (1 - \lambda) \mathbb{P}_Z) = \lambda(\mathbb{P}_X * \mathbb{P}_Y) + (1 - \lambda)(\mathbb{P}_X * \mathbb{P}_Z)$ , ie.

$$X + (\mathbf{1}_{\{\varepsilon=1\}}Y + \mathbf{1}_{\{\varepsilon=0\}}Z) = \mathbf{1}_{\{\varepsilon=1\}}(X + Y) + \mathbf{1}_{\{\varepsilon=0\}}(X + Z)$$

où  $\varepsilon \sim b(\lambda)$  est indépendante de  $X, Y, Z$ .

On considère dans la suite les deux cas particuliers des variables aléatoires discrètes (et même entières) et à densité.

## 6.3 Variables aléatoires à densité indépendantes

Dans le cas de variables aléatoires à densité, la convolée des lois est la loi de densité la convolée des densités :

**Proposition 6.3.1** Soient  $X, Y$  des variables aléatoires réelles de densités respectives  $f$  et  $g$ . Alors la variable aléatoire  $X + Y$  admet pour densité  $f * g$ , (ie.  $\mathbb{P}_X(dx) = f(x)dx$  et  $\mathbb{P}_Y(dy) = g(y)dy$  alors  $\mathbb{P}_X * \mathbb{P}_Y(dx) = (f * g)(x)dx$  ou

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}_{X+Y}(A) = \mathbb{P}(X + Y \in A) = \int_A (f * g)(x) dx.$$

**Démonstration :** On a

$$\begin{aligned} \mathbb{P}_X * \mathbb{P}_Y(A) &= \int \int \mathbf{1}_A(x + y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int \int_{x+y \in A} f(x)g(y) dx dy \\ &= \int \int_{\mathbb{R} \times A} f(u)g(v-u) du dv = \int_A \left( \int_{\mathbb{R}} f(u)g(v-u) du \right) dv = \int_A (f * g)(u) dv \end{aligned}$$

avec le changement de variable  $(x, y) \mapsto (u, v) = (x, x + y)$ . Noter que comme  $(x, y)$  varie dans  $\mathbb{R}^2$  de façon que  $x + y \in A$ ,  $u$  décrit tout  $\mathbb{R}$  et  $v$  décrit  $A$  et que le jacobien du changement de variable est

$$Jac = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

□

**Remarque 6.3.1** On connaît bien la loi de la somme  $X + Y$  si  $X$  et  $Y$  sont indépendantes, sinon, il faut connaître la loi du couple  $(X, Y)$  et par exemple sa densité  $h(x, y)$  si elle existe pour avoir la loi de  $X + Y$  par

$$\mathbb{P}(X + Y \in A) = \int_{(x,y):x+y \in A} h(x, y) dx dy = \int_A \int_{-\infty}^{+\infty} h(x, y - x) dx dy.$$

### Exemples de loi de somme de variables aléatoires indépendantes

**Proposition 6.3.2** Soient  $X_1$  de loi  $\mathcal{N}(m_1, \sigma_1^2)$  et  $X_2$  de loi  $\mathcal{N}(m_2, \sigma_2^2)$  indépendantes alors  $X_1 + X_2$  est de loi normale  $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

**Démonstration :** On a  $X_1 = m_1 + X'_1$  et  $X_2 = m_2 + X'_2$  avec  $X'_1 \sim \mathcal{N}(0, \sigma_1^2) \perp\!\!\!\perp X'_2 \sim \mathcal{N}(0, \sigma_2^2)$  où l'indépendance vient de  $X'_1 \perp\!\!\!\perp X'_2$ . On a alors

$$X_1 + X_2 = m_1 + m_2 + X'_1 + X'_2$$

et le résultat est obtenu si on montre que  $X'_1 + X'_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$ . Dans la suite, il suffit donc de considérer le cas où  $m_1 = m_2 = 0$ . Le point clef dans ce calcul est donné par la normalisation de la loi normale  $\mathcal{N}(0, \sigma^2)$  :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$

Notons alors  $f_1$  et  $f_2$  les densités de  $X_1$  et de  $X_2$ . Celle de  $X_1 + X_2$  est donnée d'après la Proposition 6.3.1 par

$$\begin{aligned} f_1 * f_2(x) &= \int_{-\infty}^{+\infty} f_1(t)f_2(x-t)dt \\ &= \int_{-\infty}^{+\infty} \exp\left(-t^2/(2\sigma_1^2)\right) \exp\left(-(x-t)^2/(2\sigma_2^2)\right) \frac{dt}{\sqrt{2\pi\sigma_1^2}\sqrt{2\pi\sigma_2^2}} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{(\sigma_1^2 + \sigma_2^2)t^2 - 2\sigma_1^2xt + \sigma_1^2x^2}{2\sigma_1^2\sigma_2^2}\right) \frac{dt}{2\pi\sigma_1\sigma_2} \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x\right)^2 - \frac{\sigma_1^4}{(\sigma_1^2 + \sigma_2^2)}x^2 + \sigma_1^2x^2}{2\sigma_1^2\sigma_2^2}\right) dt \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x\right)^2 + \frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}x^2}{2\sigma_1^2\sigma_2^2}\right) dt \\ &= \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{\left((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x\right)^2}{2\sigma_1^2\sigma_2^2}\right) dt \end{aligned}$$

$$= \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2\sigma_1^2\sigma_2^2}\right) \frac{du}{(\sigma_1^2 + \sigma_2^2)^{1/2}}$$

avec le changement de variable  $u = (\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x$ . Puis d'après la normalisation de la loi normale  $\mathcal{N}(0, \sigma_1^2\sigma_2^2)$ , la dernière intégrale vaut  $\frac{\sqrt{2\pi\sigma_1^2\sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}}$ , on a finalement :

$$f_1 * f_2(x) = \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \frac{\sqrt{2\pi\sigma_1^2\sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}} = \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}.$$

On a obtenu la densité de la loi  $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$ . On a donc

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

□

On peut aussi prouver les Propositions 6.3.2 en utilisant les fonctions caractéristiques, par exemple pour les lois normales, lorsque  $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$  et  $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$  on a :

$$\begin{aligned} \varphi_{X_1+X_2}(t) &= \varphi_{X_1}(t)\varphi_{X_2}(t) = \exp(im_1t + \sigma_1^2 t)\exp(im_2t + \sigma_2^2 t) \\ &= \exp(i(m_1 + m_2)t + (\sigma_1^2 + \sigma_2^2)t) \end{aligned}$$

qui est la fonction caractéristique de  $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

## 6.4 Cas de variables aléatoires discrètes indépendantes

Si  $X$  et  $Y$  sont des variables aléatoires discrètes indépendantes, alors  $X + Y$  est encore une variable aléatoire discrète de domaine

$$\mathcal{S}(X)(X + Y) = \mathcal{S}(X) + \mathcal{S}(Y) = \{x_i + y_j : x_i \in \mathcal{S}(X), y_j \in \mathcal{S}(Y)\}$$

et on peut calculer ses probabilités ponctuelles par une « convolution discrète ».

Supposons pour simplifier que  $\mathcal{S}(X) = \mathcal{S}(Y) = \mathbb{N}$ , et notons  $(p_k)_{k \in \mathbb{N}}, (q_k)_{k \in \mathbb{N}}$  les probabilités ponctuelles de  $X$  et de  $Y$  alors  $\mathcal{S}(X+Y) = \mathbb{N}$  et comme on a pour chaque  $k \in \mathbb{N}$  la partition suivante  $\{X + Y = n\} = \bigcup_{k=0}^n \{X = k, Y = n - k\}$ , il vient

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \mathbb{P}\left(\bigcup_{k=0}^n \{X = k, Y = n - k\}\right) = \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) \\ &= \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k) = \sum_{k=0}^n p_k q_{n-k}. \end{aligned}$$

Les probabilités ponctuelles de  $X + Y$  sont donc données par la suite  $\left(\sum_{k=0}^n p_k q_{n-k}\right)_{n \in \mathbb{N}}$ .

# Chapitre 7

## Convergences de variables aléatoires

La convergence de variables aléatoires est une notion essentielle en probabilités et mène à certains résultats fondamentaux (LGN, TCL). À titre d'exemple introductif, notons que lors d'un jeu de pile ou face où on lance une pièce dont la fréquence d'apparition de pile est  $p$ , la fréquence observée du nombre de pile obtenu après  $n$  lancers est « proche » de  $p$ , pourvu que  $n$  soit « assez grand ». Donc, si  $p$  est inconnue, cette observation offre un moyen d'approximer  $p$  en comptant les fréquences de pile pour un grand nombre de lancers. En probabilité, plusieurs modes de convergence sont possibles, ils sont introduits dans ce chapitre, dans lequel les suites de variables aléatoires sont supposées construites sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Pour simplifier on ne considère que des variables aléatoires réelles, mais les énoncés et les résultats restent vrais pour des vecteurs aléatoires à valeurs dans  $\mathbb{R}^d$ . (On peut même les adapter pour des variables aléatoires à valeurs dans un espace métrique  $(S, d)$ .)

### 7.1 Convergence presque sûre

La convergence de variables aléatoires correspond à la convergence de fonctions. Pour les fonctions, la convergence la plus faible est la convergence simple qui s'énonce pour des variables aléatoires de la façon suivante : pour tout  $\omega \in \Omega$ ,  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ . En probabilité (et plus généralement en théorie de la mesure), il est trop restrictif de demander la convergence pour **tous** les  $\omega \in \Omega$ . À la place, on considère la convergence presque sûre :

**Définition 7.1.1 (Convergence presque sûre)** *Une suite  $(X_n)_{n \geq 0}$  de variables aléatoires converge presque sûrement vers  $X$  si la convergence est vraie avec une probabilité 1*

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

On note la convergence presque sûre :  $X_n \xrightarrow{ps} X$ .

**Remarque 7.1.1** — Notons que l'ensemble  $A = \{ \lim_{n \rightarrow +\infty} X_n = X \}$  est bien un évènement car la convergence  $X_n \xrightarrow{ps} X$  s'écrit pour presque chaque  $\omega \in \Omega$  :

$$\forall \varepsilon > 0, \exists m \in \mathbb{N}, \forall n \geq m : |X_n(\omega) - X(\omega)| < \varepsilon. \quad (7.1)$$

On peut donc écrire car

$$A = \left\{ \lim_{n \rightarrow +\infty} X_n = X \right\} = \bigcap_{k \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{|X_n - X| < 1/k\} \in \mathcal{F}.$$

— D'un point de vue analytique, c'est une convergence plus faible que la convergence simple (déjà très faible). Mais d'un point de vue probabiliste, ce sera une des convergences les plus fortes que nous considérerons.

Dans la suite, on étudiera notamment les liens entre cette convergence et les autres.

Par ailleurs comme pour une suite d'évènements  $(A_n)_{n \geq 1}$ ,  $\mathbb{P}(\bigcap_{n \geq 1} A_n) = 1$  est équivalent à  $\mathbb{P}(A_n) = 1$  pour tout  $n \geq 1$ , on a que  $X_n$  converge ps vers  $X$  si et seulement si

$$\forall k \geq 1, \quad \mathbb{P}\left(\bigcup_{m \geq 1} \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}\right) = 1$$

c'est à dire

$$\begin{aligned} & \forall \varepsilon > 0, \quad \mathbb{P}\left(\bigcup_{m \geq 1} \bigcap_{n \geq m} \{|X_n - X| \leq \varepsilon\}\right) = 1 \\ \iff & \forall \varepsilon > 0, \quad \mathbb{P}\left(\bigcap_{m \geq 1} \bigcup_{n \geq m} \{|X_n - X| > \varepsilon\}\right) = 0 \\ \iff & \forall \varepsilon > 0, \quad \mathbb{P}(|X_n - X| > \varepsilon \text{ i.s.}) = 0. \end{aligned} \quad (7.2)$$

Par convergence monotone des probabilités, c'est encore équivalent à

$$\begin{aligned} & \forall \varepsilon > 0, \quad \lim_{m \rightarrow +\infty} \mathbb{P}\left(\bigcup_{n \geq m} \{|X_n - X| > \varepsilon\}\right) = 0 \\ \iff & \forall \varepsilon > 0, \quad \lim_{m \rightarrow +\infty} \mathbb{P}\left(\sup_{n \geq m} |X_n - X| > \varepsilon\right) = 0. \end{aligned} \quad (7.3)$$

De la même façon, en partant du critère de Cauchy (réel)

$$\forall \omega \in \Omega, \forall \varepsilon > 0, \exists p \in \mathbb{N}, \forall n, m \geq p : |X_n(\omega) - X_m(\omega)| < \varepsilon$$

à la place de (7.1), on a le critère suivant de convergence presque sûre. Ce critère a l'avantage de s'exprimer sans la limite  $X$ . On peut donc déterminer la convergence presque sûre sans connaître la limite.

**Proposition 7.1.1 (Critère de Cauchy)** Une suite de variables aléatoires  $(X_n)_{n \geq 1}$  converge ps si et seulement si

$$\forall \varepsilon > 0, \quad \mathbb{P} \left( \bigcup_{m \geq 1} \bigcap_{n \geq 1} |X_n - X_m| \leq \varepsilon \right) = 1.$$

**Proposition 7.1.2 (Borel-Cantelli pour la convergence ps)** Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires réelles.

1. Si pour tout  $\varepsilon > 0$ ,  $\sum_{n=0}^{+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty$  alors  $X_n \xrightarrow{\text{ps}} X$ .
2. Si les variables aléatoires  $(X_n)_{n \geq 0}$  sont mutuellement indépendantes alors  $X_n \xrightarrow{\text{ps}} 0$  si et seulement si pour tout  $\varepsilon > 0$ , on a  $\sum_{n=0}^{+\infty} \mathbb{P}(|X_n| \geq \varepsilon) < +\infty$ .

**Démonstration :** Pour 1), on considère  $\varepsilon > 0$  et les évènements

$$A_n = \{|X_n - X| \geq \varepsilon\}, \quad n \in \mathbb{N}.$$

En appliquant le lemme de Borel-Cantelli aux  $(A_n)_{n \geq 0}$ , on déduit que  $\mathbb{P}(A_n \text{ i.s.}) = 0$ , ce qui donne le critère (7.2) pour la convergence ps. Pour 2), on utilise la deuxième partie du lemme de Borel-Cantelli pour les évènements indépendants. Noter qu'il est nécessaire de supposer  $X = 0$  pour conserver dans ce cas des évènements  $A_n$  indépendants, sinon, chaque  $A_n$  dépendrait de  $X$  et les  $A_n$ ,  $n \geq 0$ , seraient dépendants.  $\square$

**Proposition 7.1.3** Soit  $X_n \xrightarrow{\text{ps}} X$  et  $f$  une fonction continue alors  $f(X_n) \xrightarrow{\text{ps}} f(X)$ .

**Démonstration :** Le résultat est immédiat puisque pour presque chaque  $\omega \in \Omega$  on a  $X_n(\omega) \rightarrow X(\omega)$  (convergence presque sûre). Comme  $f$  est continue, il vient, pour ces  $\omega : f(X_n(\omega)) \rightarrow f(X(\omega))$ , on a donc  $f(X_n) \rightarrow f(X)$  presque sûrement.  $\square$

## 7.2 Convergence en probabilité

La convergence en probabilité de  $X_n$  vers  $X$  garantit que la probabilité que les variables aléatoires  $X_n$  et  $X$  restent « éloignées » tend vers 0.

**Définition 7.2.1 (Convergence en probabilité)** Une suite  $(X_n)_{n \geq 0}$  de variables aléatoires converge en probabilité vers la variable aléatoire  $X$  si

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X - X_n| \geq \varepsilon) = 0. \quad (7.4)$$

On note la convergence en probabilité :  $X_n \xrightarrow{\mathbb{P}} X$ .

**Proposition 7.2.1 (Unicité ps de la limite en probabilité)** Si  $X_n \xrightarrow{\mathbb{P}} X$  et  $X_n \xrightarrow{\mathbb{P}} X'$  alors  $X' = X$  ps.

**Démonstration :** Pour tout  $\varepsilon > 0$ , on a

$$\left\{ |X - X'| > \varepsilon \right\} \subset \left\{ |X_n - X| > \varepsilon/2 \right\} \cup \left\{ |X_n - X'| > \varepsilon/2 \right\}$$

et

$$\mathbb{P}(|X - X'| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - X'| > \varepsilon/2)$$

et on déduit de  $X_n \xrightarrow{\mathbb{P}} X$  et  $X_n \xrightarrow{\mathbb{P}} X'$  que pour tout  $\varepsilon > 0$

$$\mathbb{P}(|X - X'| > \varepsilon) = 0.$$

On a alors

$$\mathbb{P}(X \neq X') = \mathbb{P}\left(\bigcup_{k \geq 1} \{|X - X'| > 1/k\}\right) = 0.$$

□

Noter que la convergence presque sûre exige un sup supplémentaire dans la définition de la convergence en probabilité, cf. (7.3). La convergence en probabilité s'exprime

$$\forall \varepsilon > 0, \forall \delta > 0, \exists n_1(\varepsilon, \delta) \text{ tel que pour } n \geq n_1, \quad \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \delta. \quad (7.5)$$

ce qui est équivalent à

$$\forall \varepsilon > 0, \exists n_2(\varepsilon) \text{ tel que pour } n \geq n_2, \quad \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \varepsilon. \quad (7.6)$$

En effet, (7.5) implique facilement (7.6) avec le choix  $\delta = \varepsilon$ .

Puis si (7.6) est vraie :

- lorsque  $\varepsilon < \delta$ , il suffit de prendre  $n_1(\varepsilon, \delta) = n_2(\varepsilon)$ .
- lorsque  $\varepsilon \geq \delta$ , on prend  $n_1(\varepsilon, \delta) = n_2(\delta)$  et on a

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq \delta) \leq \delta.$$

En fait  $n_1(\varepsilon, \delta) = n_2(\delta \wedge \varepsilon)$ .

On ainsi

**Proposition 7.2.2 (Métrique de la convergence en probabilité)** Soit

$$d_{\mathbb{P}}(X, Y) = \inf(c > 0 : \mathbb{P}(|X - Y| \geq c) \leq c).$$

Alors  $d_{\mathbb{P}}$  définit une distance sur l'ensemble des variables aléatoires et cette distance métrise la convergence en probabilité.

**Démonstration :** D'abord, on montre que  $d_{\mathbb{P}}$  est une distance :

109

- Il est clair que  $d_{\mathbb{P}}(X, X) = 0$  car  $\mathbb{P}(|X - X| \geq c) = 0 \leq c$  et donc  $\{c > 0 : \mathbb{P}(|X - Y| \geq c) \leq c\} = \mathbb{R}_+$  et l'inf est 0.
- On a évidemment  $d_{\mathbb{P}}(X, Y) = d_{\mathbb{P}}(Y, X)$ .
- Puis comme

$$\{|X - Y| \geq c + d\} \subset \{|X - Z| \geq c\} \cup \{|Z - Y| \geq d\}$$

puisque lorsque  $|X - Z| < c$  et  $|Z - Y| < d$  alors  $|X - Y| \leq |X - Z| + |Z - Y| < c + d$  prouve l'inclusion convenable des complémentaires. Pour  $c$  et  $d$  tels que  $\mathbb{P}(|X - Z| \geq c) \leq c$  et  $\mathbb{P}(|Z - Y| \geq d) \leq d$ , on a

$$\mathbb{P}(|X - Y| \geq c + d) \leq \mathbb{P}(|X - Z| \geq c) + \mathbb{P}(|Z - Y| \geq d) \leq c + d$$

et par définition  $d_{\mathbb{P}}(X, Y) \leq c + d$ . En faisant  $c \searrow d(X, Z)$  et  $d \searrow d_{\mathbb{P}}(Z, Y)$ , on a l'inégalité triangulaire

$$d_{\mathbb{P}}(X, Y) \leq d_{\mathbb{P}}(Z, Y) + d_{\mathbb{P}}(Z, Y).$$

Ensuite, la formulation (7.6) de la convergence en probabilité assure que  $d_{\mathbb{P}}(X_n, X) \rightarrow 0$  est équivalent à  $X_n \xrightarrow{\mathbb{P}} X$ .  $\square$

**Proposition 7.2.3 (Convergences ps et proba)** *La convergence presque sûre implique la convergence en probabilité : si la suite  $(X_n)_{n \geq 0}$  converge presque sûrement vers  $X$ , elle converge en probabilité.*

**Démonstration :** Soit  $\varepsilon > 0$  fixé. On pose  $Y_n = \mathbf{1}_{\{|X_n - X| > \varepsilon\}}$ . Les fonctions  $Y_n$  sont mesurables, positives et majorées par 1. Par ailleurs, d'après la convergence presque sûre, pour presque tout  $\omega \in \Omega$ , il existe  $n_0$  tel que pour tout  $n \geq n_0$ ,

$$|X_n(\omega) - X(\omega)| \leq \varepsilon$$

c'est à dire  $Y_n(\omega) = 0$  pour tout  $n \geq n_0$ . On a donc écrit que la suite  $Y_n$  converge presque sûrement vers 0. D'après le théorème de convergence dominée

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}[Y_n] \longrightarrow \mathbb{E}\left[\lim_{n \rightarrow +\infty} Y_n\right] = \mathbb{E}[0] = 0.$$

 $\square$ 

La réciproque est fausse (cf. exemple ci-dessous). Cependant, ce résultat admet une réciproque partielle pour les sous-suites qu'on obtiendra par le lemme de Borel-Cantelli, cf. Proposition 7.2.6.

**Exemple.** La convergence en probabilité n'implique pas la convergence presque sûre. Soit l'espace de probabilité  $([0, 1], \mathcal{B}([0, 1]), \lambda)$  et on considère le tableau triangulaire de variables aléatoires

$$Y_{2^n, j}(\omega) = \mathbf{1}_{[\frac{j}{2^n}, \frac{j+1}{2^n}]}(\omega), \quad 0 \leq j \leq 2^n - 1, \quad n \geq 1$$

On considère alors la suite  $(X_n)_{n \geq 1}$  définie par

$$X_p = Y_{2^n, j} \text{ lorsque } p = 2^n + j \text{ avec } 0 \leq j \leq 2^n - 1, \quad n \geq 1$$

(prendre  $n = [\ln p / \ln 2]$  et  $j = p - 2^n$ ). Alors pour tout  $\omega \in [0, 1]$ , on a  $X_p(\omega) = 1$  si et seulement si  $\omega \in ]\frac{j}{2^n}, \frac{j+1}{2^n}]$  et  $X_p(\omega) = 0$  sinon. Ainsi,  $X_p(\omega)$  vaut (alternativement) 0 (souvent) et 1 (parfois) et  $\liminf_{n \rightarrow +\infty} X_n(\omega) = 0$  et  $\limsup_{n \rightarrow +\infty} X_n(\omega) = 1$  si bien que  $X_n(\omega)$  ne converge pas ps. Mais pour tout  $\varepsilon \in ]0, 1[$  et  $p = 2^n + j$ ,  $0 \leq j \leq 2^n - 1$ , puisque

$$\mathbb{P}(|X_p| \geq \varepsilon) = \lambda\left(\left[\frac{j}{2^n}, \frac{j+1}{2^n}\right]\right) = 2^{-n},$$

$X_p$  converge en probabilité car  $n = [\ln p / \ln 2]$  et  $p$  tendent vers  $+\infty$  en même temps.

**Exemple 7.2.1 (Cas de lois de Bernoulli)** Soient  $X_n \sim b(p_n)$  des variables aléatoires alors

- $X_n \xrightarrow{\mathbb{P}} 0$  si et seulement si  $p_n \rightarrow 0$  car  $\mathbb{P}(|X_n| \geq \varepsilon) = p_n$  pour  $\varepsilon \in (0, 1)$  ;
- Lorsque la suite  $(X_n)_n$  est indépendante :  $X_n \xrightarrow{ps} 0$  si et seulement si  $\sum_{n \geq 1} p_n < +\infty$  d'après le lemme de Borel-Cantelli pour la convergence ps (Proposition 7.1.2).

**Définition 7.2.2 (Autre métrique pour la convergence en proba)** Soient  $X, Y \in L^0(\Omega, \mathcal{F}, \mathbb{P})$ .

On définit

$$\widehat{d}_{\mathbb{P}}(X, Y) = \mathbb{E}[\min(|X - Y|, 1)].$$

Il est facile de voir que  $\widehat{d}_{\mathbb{P}}(X, Y) \geq 0$ ,  $\widehat{d}_{\mathbb{P}}(X, Y) = 0$  si et seulement si  $X = Y$  ps et  $\widehat{d}_{\mathbb{P}}(X, Y) \leq \widehat{d}_{\mathbb{P}}(X, Z) + \widehat{d}_{\mathbb{P}}(Z, Y)$  si bien que  $\widehat{d}$  est une métrique. Pour l'inégalité triangulaire, on note que si  $\min(|X - Z|, 1) = 1$  alors

$$\min(|X - Y|, 1) \leq 1 = \min(|X - Z|, 1) \leq \min(|X - Z|, 1) + \min(|Z - Y|, 1);$$

puis si  $\min(|X - Z|, 1) = |X - Z|$  et  $\min(|Z - Y|, 1) = |Z - Y|$  alors

$$\min(|X - Y|, 1) \leq |X - Y| \leq |X - Z| + |Z - Y| = \min(|X - Z|, 1) + \min(|Y - Z|, 1)$$

L'inégalité triangulaire pour  $\widehat{d}_{\mathbb{P}}$  vient en prenant l'espérance dans l'inégalité ci-dessus. Le résultat suivant indique qu'elle métrise la convergence en probabilité.

**Proposition 7.2.4** On a  $X_n \xrightarrow{\mathbb{P}} X$  si et seulement si  $\widehat{d}_{\mathbb{P}}(X_n, X) \rightarrow 0$ ,  $n \rightarrow +\infty$ .

**Démonstration :** Par l'inégalité de Markov, on a pour tout  $\varepsilon \in ]0, 1]$  :

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(\min(|X_n - X|, 1) \geq \varepsilon) \leq \frac{\widehat{d}_{\mathbb{P}}(X_n, X)}{\varepsilon}. \quad (7.7)$$

Si  $\varepsilon > 1$ , on utilise

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \mathbb{P}(|X_n - X| \geq 1).$$

Pour la réciproque, on a

$$\begin{aligned} \widehat{d}_{\mathbb{P}}(X_n, X) &= \mathbb{E}[\min(|X_n - X|, 1)] \\ &= \mathbb{E}[\min(|X_n - X|, 1)\mathbf{1}_{\{|X_n - X| \leq \varepsilon\}}] + \mathbb{E}[\min(|X_n - X|, 1)\mathbf{1}_{\{|X_n - X| > \varepsilon\}}] \\ &= \varepsilon + \mathbb{P}(|X_n - X| > \varepsilon) \leq 2\varepsilon \end{aligned} \quad (7.8)$$

si  $n$  est assez grand puisque  $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$  quand  $X_n \xrightarrow{\mathbb{P}} X$ . On a donc bien  $\lim_{n \rightarrow +\infty} \widehat{d}_{\mathbb{P}}(X_n, X) = 0$ .  $\square$

**Proposition 7.2.5 (Lien entre les distances  $d_{\mathbb{P}}$  et  $\widehat{d}_{\mathbb{P}}$ )** *On a*

$$\frac{1}{2}\widehat{d}_{\mathbb{P}}(X, Y) \leq d_{\mathbb{P}}(X, Y) \leq \sqrt{\widehat{d}_{\mathbb{P}}(X, Y)}. \quad (7.9)$$

**Démonstration :** On observe que  $d_{\mathbb{P}}(X, Y) \leq 1$  et  $\widehat{d}_{\mathbb{P}}(X, Y) \leq 1$ . En reprenant la borne (7.8) on a  $\widehat{d}_{\mathbb{P}}(X, Y) \leq 2c$  pour tout  $c > d_{\mathbb{P}}(X, Y)$  d'où il vient  $\widehat{d}_{\mathbb{P}}(X, Y) \leq 2d_{\mathbb{P}}(X, Y)$ , ce qui donne la première partie de (7.9). Ensuite en combinant (7.7) avec le choix  $\varepsilon = \sqrt{\widehat{d}_{\mathbb{P}}(X_n, X)}$ , il vient

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\widehat{d}_{\mathbb{P}}(X_n, X)}{\varepsilon} = \sqrt{\widehat{d}_{\mathbb{P}}(X_n, X)} = \varepsilon$$

soit  $d_{\mathbb{P}}(X, Y) \leq \varepsilon = \sqrt{\widehat{d}_{\mathbb{P}}(X_n, X)}$ , ce qui conclut avec la droite de (7.9).  $\square$

**Proposition 7.2.6 (Critère de sous-suites pour la convergence en probabilité)** *Soient  $X$  et  $(X_n)_{n \geq 0}$  des variables aléatoires réelles. Alors  $X_n \xrightarrow{\mathbb{P}} X$  si et seulement si de toute suite croissante d'entiers  $(n')$  on peut extraire  $(n'') \subset (n')$  telle que  $X_{n''} \rightarrow X$  ps.*

**Démonstration :** Supposons que  $X_n \xrightarrow{\mathbb{P}} X$ . Pour tout  $k \in \mathbb{N}^*$ , soit  $n_k$  le plus petit entier tel que  $\mathbb{P}(|X_{n_k} - X| \geq 1/k) < 1/2^k$ . Alors  $\sum_{k \in \mathbb{N}^*} \mathbb{P}(|X_{n_k} - X| \geq 1/k) < +\infty$ . Par le lemme de Borel-Cantelli, pour presque chaque  $\omega \in \Omega$ , il existe  $k_0(\omega)$  tel que pour  $k \geq k_0(\omega)$   $|X_{n_k}(\omega) - X(\omega)| < 1/k$ . Cela assure  $X_{n_k} \xrightarrow{ps} X$ .

On en déduit le critère de sous-suite en appliquant pour toute sous-suite  $(n') = (n_p)_{p \geq 1} \subset (n)$  le résultat précédent à  $X_{n'}$ .

Réiproquement, si  $(X_n)_{n \geq 0}$  ne converge pas en probabilité, alors il existe une suite croissante  $(n')$ ,  $\delta > 0$  et  $\varepsilon > 0$  tels que  $\mathbb{P}(|X_{n'} - X| \geq \varepsilon) \geq \delta$ . Mais par hypothèse, on peut extraire une sous-suite  $(n'') \subset (n')$  telle que  $X_{n''}$  converge ps vers  $X$  et donc en probabilité (d'après la Proposition 7.2.3). Mais, c'est absurde car  $\mathbb{P}(|X_{n'} - X| \geq \varepsilon) \geq \delta$  et  $(n'') \subset (n')$ .  $\square$

De la première partie de la preuve précédente, on déduit en particulier :

**Corollaire 7.2.1** *Si  $X_n \xrightarrow{\mathbb{P}} X$  alors il existe  $(n') \subset (n)$  tel que  $X_{n'} \xrightarrow{ps} X$ .*

**Remarque 7.2.1** De cette Proposition 7.2.6, on voit que la convergence presque sûre n'est pas métrique : supposons qu'elle est métrisée par  $d_{ps}$  et considérons  $X_n \xrightarrow{\mathbb{P}} X$ . On aurait alors  $X_n \rightarrow X$  ps. En effet, si ce n'est pas le cas  $X_n \not\xrightarrow{ps} X$ , ie.  $d_{ps}(X_n, X) \not\rightarrow 0$ . Il existe donc  $\varepsilon > 0$  et  $(n_k)_{k \geq 1}$  sous-suite croissante telle que  $d_{ps}(X_{n_k}, X) > \varepsilon$ . D'après la proposition précédente, comme  $X_n \xrightarrow{\mathbb{P}} X$ , on peut extraire  $(n_{k_p})_{p \geq 1}$  de  $(n_k)_{k \geq 1}$  telle que  $X_{n_{k_p}} \rightarrow X$  ps ou encore  $d_{ps}(X_{n_{k_p}}, X) \rightarrow 0$ , ce qui nie  $d_{ps}(X_{n_{k_p}}, X) \geq \varepsilon$ . On doit donc avoir  $X_n \rightarrow X$  ps. Autrement formulé : on aurait l'équivalence entre les convergences presque sûre et en probabilité, ce qui est faux (cf. contre-exemple ci-dessus).

**Conclusion :** La convergence presque sûre n'est pas métrisable.

**Proposition 7.2.7** *Soient  $f : \mathbb{R} \rightarrow \mathbb{R}$  continue et  $(X_n)_{n \geq 0}$  une suite de variables aléatoires réelles qui converge en probabilité vers une variable aléatoire  $X$ . Alors  $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ . En particulier si en plus  $Y_n \xrightarrow{\mathbb{P}} Y$ , pour tout  $\alpha, \beta \in \mathbb{R}$ ,  $\alpha X_n + \beta Y_n \xrightarrow{\mathbb{P}} \alpha X + \beta Y$  et  $X_n Y_n \xrightarrow{\mathbb{P}} XY$ .*

**Démonstration :** La preuve se fait facilement avec la Proposition 7.2.6 en utilisant la convergence ps de sous-suite extraite. Par exemple, si  $X_n \xrightarrow{\mathbb{P}} X$ , soit  $(n') \subset (n)$  une sous-suite, on peut extraire  $(n'') \subset (n')$ , telle que  $X_{n''}$  converge ps vers  $X$ . Comme  $f$  est continue, on a aussi  $f(X_{n''}) \rightarrow f(X)$  ps quand  $n'' \rightarrow +\infty$ . D'après le critère de sous-suites précédent,  $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ . On fait de même pour les autres assertions.  $\square$

On dit que  $(X_n)_{n \geq 0}$  est une suite de variables aléatoires réelles vérifie le **critère de Cauchy en probabilité** si

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 : \mathbb{P}(|X_n - X_m| \geq \varepsilon) < \varepsilon. \quad (7.10)$$

soit  $\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 : d_{\mathbb{P}}(X_n, X_m) < \varepsilon$ . En utilisant (7.8), le critère (7.10) est encore équivalent au critère de Cauchy pour  $\widehat{d}_{\mathbb{P}}$  :

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 : \widehat{d}_{\mathbb{P}}(X_n, X_m) < \varepsilon.$$

**Proposition 7.2.8 (Critère de Cauchy pour la CV en proba)** *Une suite de variables aléatoires  $(X_n)_{n \geq 0}$  converge en probabilité si et seulement si  $X_n$  satisfait le critère de Cauchy en probabilité (7.10). Autrement dit, l'espace  $L^0(\Omega, \mathcal{F}, \mathbb{P})$  est complet pour les distances  $d_{\mathbb{P}}$  et  $\hat{d}_{\mathbb{P}}$  métrisant la convergence en probabilité.*

**Démonstration :** On suppose d'abord que  $(X_n)_{n \geq 0}$  converge en probabilité vers  $X$  : pour tout  $\varepsilon > 0$  il existe  $n_0$  tel que dès que  $n \geq n_0$  on ait  $\mathbb{P}(|X_n - X| > \varepsilon) < \varepsilon$ . Par l'inégalité triangulaire on peut écrire pour tout  $n \geq m \geq n_0$  :

$$\{|X_n - X_m| > \varepsilon\} \subset \{|X_n - X| > \varepsilon/2\} \cup \{|X_m - X| > \varepsilon/2\}$$

et on en déduit le critère de Cauchy en probabilité de la convergence en probabilité. Réciproquement, soit  $\varepsilon = 2^{-k}$  dans la condition de Cauchy en probabilité. On pose  $n_1 = 1$  et par récurrence soit  $n_k$  le plus petit entier tel que pour  $n, m \geq n_k$  :  $\mathbb{P}(|X_n - X_m| > 2^{-k}) \leq 2^{-k}$ , ie.

$$n_k = \min \left( N > n_{k-1} : \mathbb{P}(|X_n - X_m| > 2^{-k}) < 2^{-k}, \quad \forall n, m \geq N \right).$$

Ainsi  $n_k > n_{k-1}$  et

$$\mathbb{P}(|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}) < 2^{-k}.$$

Par le lemme de Borel-Cantelli, pour presque tout  $\omega \in \Omega$ , il existe un entier  $k_0(\omega) < +\infty$  tel que si  $m \geq k_0(\omega)$ , on a  $|X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leq 2^{-m}$ . Alors la suite de réels  $(X_{n_m}(\omega))_{m \geq 1}$  est de Cauchy. En effet, soit  $\varepsilon > 0$  et  $p \geq l \geq \max(k_0, l_0)$  avec  $l_0 = \ln(2/\varepsilon)/\ln(2)$ , on a

$$|X_{n_p}(\omega) - X_{n_l}(\omega)| \leq \sum_{m=l}^{p-1} |X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leq \sum_{m=l}^{p-1} 2^{-m} \leq \sum_{m=l}^{+\infty} 2^{-m} \leq 2^{-l+1} \leq \varepsilon.$$

La limite de  $(X_{n_m}(\omega))_{m \geq 1}$ , suite de Cauchy dans  $\mathbb{R}$ , existe. On la note  $X(\omega) = \lim_{k \rightarrow +\infty} X_{n_k}(\omega)$ . On construit ainsi  $X(\omega)$  pour presque chaque  $\omega \in \Omega$  et on prend  $X(\omega) = 0$  pour les  $\omega$ , négligeables, tels que  $(X_{n_m}(\omega))_{m \geq 1}$  n'est pas de Cauchy. On définit ainsi la variable aléatoire  $X$ . De cette façon,  $X_{n_k}$  converge presque sûrement vers  $X$  et en particulier cette sous-suite converge en probabilité vers  $X$ . Pour conclure, on écrit :

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(|X_n - X_{n_k}| > \varepsilon/2) + \mathbb{P}(|X_{n_k} - X| > \varepsilon/2).$$

La condition de Cauchy garantit que le premier terme tend vers 0 pour  $n$  et  $k$  assez grands. La convergence en probabilité de la sous-suite  $X_{n_k}$  vers  $X$  garantit la convergence vers 0 du second terme quand  $k \rightarrow +\infty$ . Finalement

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

ie.  $X_n \xrightarrow{\mathbb{P}} X$ .

□

## 7.3 Convergence en norme $p$

### 7.3.1 Définition et propriétés

On rappelle que pour  $p > 1$ , on définit les espaces  $L^p$  :

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{R} : \|X\|_p = (\mathbb{E}[|X|^p])^{1/p} < +\infty\} / \sim_{ps}$$

où  $\sim_{ps}$  désigne la relation d'équivalence « être égale presque sûrement », cf. Section 3.4. Ce sont des espaces de Banach pour la norme  $\|\cdot\|_p$ . Pour  $p = 2$ ,  $L^2(\Omega, \mathcal{F}, \mathbb{P})$  est un espace de Hilbert car la norme  $\|\cdot\|_2$  dérive du produit scalaire  $\langle X, Y \rangle = \mathbb{E}[XY]$ .

**Définition 7.3.1 (Convergence en norme  $p$ )** *On dit qu'une suite de variables aléatoires  $(X_n)_{n \geq 0}$  converge en norme  $p$  vers  $X$  si  $\|X_n - X\|_p \rightarrow 0$  quand  $n \rightarrow +\infty$ , c'est à dire*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

On note la convergence en norme  $p$  :  $X_n \xrightarrow{L^p} X$ .

**Proposition 7.3.1 (Comparaison des convergences  $L^p$  et  $L^q$ )** *Si  $p \geq q$ , la convergence en norme  $p$  entraîne la convergence en norme  $q$ .*

Il s'agit de la même preuve que pour l'inclusion topologique  $L^p(\Omega, \mathcal{F}, \mathbb{P}) \subset L^q(\Omega, \mathcal{F}, \mathbb{P})$  fondée sur une inégalité entre norme  $L^p$  et  $L^q$ , cf. Proposition 3.4.1.

**Démonstration :** Soit  $p > q$ . Avec  $\alpha = \frac{p}{q} > 1$  et  $\beta$  l'indice conjugué de  $\alpha$  ( $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ ), d'après l'inégalité de Hölder, pour toute variable aléatoire  $X$ , on a :

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

On en déduit  $\|X\|_q \leq \|X\|_p$ . Appliquée à la variable aléatoire  $X_n - X$ , on a  $\|X_n - X\|_q \leq \|X_n - X\|_p$ , ce qui conclut.  $\square$

En particulier, la convergence en norme  $p$  entraîne toujours la convergence en norme  $\|\cdot\|_1$ . La convergence en norme  $\|\cdot\|_\infty$  entraîne toute convergence en norme  $p$ .

**Proposition 7.3.2 (Convergences  $L^p$  et proba)** *Si  $X_n$  et  $X$  sont dans  $L^p(\Omega, \mathcal{F}, \mathbb{P})$  alors la convergence en norme  $\|\cdot\|_p$   $X_n \xrightarrow{L^p} X$  entraîne la convergence en probabilité  $X_n \xrightarrow{\mathbb{P}} X$ .*

**Démonstration :** Il suffit d'appliquer l'inégalité de Markov : soit  $\varepsilon > 0$ ,

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} = \frac{\|X_n - X\|_p^p}{\varepsilon^p} \rightarrow 0, n \rightarrow +\infty.$$

La réciproque est fausse : la convergence en probabilité seule n'implique pas la convergence (par exemple) dans  $L^1$  :

**Exemple :** Soit  $X_n$  de loi donnée par  $\mathbb{P}(X_n = 0) = 1 - 1/n$  et  $\mathbb{P}(X_n = a_n) = 1/n$ . On a  $X_n \xrightarrow{\mathbb{P}} X = 0$  puisque

$$\mathbb{P}(|X_n| \geq \varepsilon) = \mathbb{P}(X_n = a_n) = 1/n \rightarrow 0.$$

Puis  $\mathbb{E}[|X_n - X|] = \mathbb{E}[|X_n|] = |a_n|/n$  qui ne tend pas vers 0 en général : par exemple si  $a_n = 2n$  alors  $\mathbb{E}[X_n] = 2 \rightarrow 2$ , si  $a_n = n^2$  alors  $\mathbb{E}[X_n] \rightarrow +\infty$ .

Dans la section suivante, on introduit la notion qui permet de donner une réciproque à la Proposition 7.3.2, cf. le Théorème de Vitali (Théorème 7.3.1).

### 7.3.2 Uniforme intégrabilité

**Définition 7.3.2 (Uniforme intégrabilité)** Une suite de variables aléatoires intégrables  $(X_n)_{n \geq 0}$  est dite uniformément intégrable (u.i.) si

$$\lim_{c \rightarrow +\infty} \sup_{n \geq 0} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}] = 0.$$

- Remarque 7.3.1**
- Une famille de variables aléatoires avec un seul élément (intégrable) est uniformément intégrable (par convergence dominée!).
  - Une suite de variables aléatoires dominées par une variable aléatoire  $Z$  intégrable est uniformément intégrable. En effet, par croissance  $|X_n| \mathbf{1}_{\{|X_n| > c\}} \leq Z \mathbf{1}_{\{Z > c\}}$ , d'où :

$$\lim_{c \rightarrow +\infty} \sup_{n \geq 0} \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > c\}}] \leq \lim_{c \rightarrow +\infty} \mathbb{E}[Z \mathbf{1}_{\{Z > c\}}] = 0.$$

- Une suite finie de variables intégrables est uniformément intégrable. En effet, une telle suite  $(X_k)_{k=1,\dots,n}$  est dominée par  $\sum_{k=1}^n |X_k|$  intégrable.
- Si  $(X_n)_{n \geq 0}$  et  $(Y_n)_{n \geq 0}$  sont deux suites de variables aléatoires avec, pour tout  $n$ ,  $|X_n| \leq |Y_n|$  et  $(Y_n)_{n \geq 0}$  uniformément intégrable alors  $(X_n)_{n \geq 0}$  l'est aussi.

**Proposition 7.3.3** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires dans  $L^p$  telle que pour  $\delta > 0$  la suite est bornée dans  $L^{p+\delta}$ . Alors la suite  $(X_n^p)_{n \geq 0}$  est uniformément intégrable.

**Démonstration :** En effet,

$$\begin{aligned} \sup_{n \geq 0} \mathbb{E}[|X_n|^p \mathbf{1}_{\{|X_n|^p > c\}}] &= \sup_{n \geq 0} \mathbb{E}[|X_n|^p \times 1 \times \mathbf{1}_{\{|X_n|/c^{1/p} > 1\}}] \\ &\leq \sup_{n \geq 0} \mathbb{E}\left[|X_n|^p \times \frac{|X_n|^\delta}{c^{\delta/p}} \times \mathbf{1}_{\{|X_n|/c^{1/p} > 1\}}\right] \\ &\leq \frac{1}{c^{\delta/p}} \sup_{n \geq 0} \mathbb{E}[|X_n|^{p+\delta}] = O(c^{-\delta/p}) \rightarrow 0, \quad c \rightarrow +\infty. \end{aligned}$$

□

Rappelons que si une variable aléatoire  $X$  est intégrable alors pour tout  $\varepsilon > 0$ , il existe  $\eta > 0$  tel que si  $A \in \mathcal{F}$  avec  $\mathbb{P}(A) < \eta$  alors  $\mathbb{E}[|X|\mathbf{1}_A] < \varepsilon$ . En effet, c'est une conséquence du théorème de convergence dominée : pour  $c$  assez grand,  $\mathbb{E}[|X|\mathbf{1}_{\{|X|>c\}}] \leq \varepsilon/2$ . Puis pour  $\eta < \varepsilon/(2c)$ , on a

$$\mathbb{E}[|X|\mathbf{1}_A] = \mathbb{E}[|X|\mathbf{1}_{A \cap \{|X|>c\}}] + \mathbb{E}[|X|\mathbf{1}_{A \cap \{|X|\leq c\}}] \leq \frac{\varepsilon}{2} + c\mathbb{P}(A) \leq \frac{\varepsilon}{2} + c\eta < \varepsilon.$$

Pour des variables aléatoires uniformément intégrables, on a la généralisation suivante de ce rappel :

**Proposition 7.3.4 (Critère d'uniforme intégrabilité)** *Une suite de variables aléatoires réelles  $(X_n)_{n \geq 0}$  est uniformément intégrable si et seulement si*

1.  $\forall \varepsilon > 0$ ,  $\exists \eta > 0$  tel que pour  $A \in \mathcal{F}$  avec  $\mathbb{P}(A) < \eta$  on a  $\mathbb{E}[|X_n|\mathbf{1}_A] < \varepsilon$  pour tout  $n \in \mathbb{N}$ .
2.  $\sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$ .

**Démonstration :** On suppose d'abord que  $(X_n)_{n \geq 0}$  est uniformément intégrable : pour tout  $\varepsilon > 0$ ,  $\exists c > 0$  tel que  $\sup_{n \geq 0} \mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] < \varepsilon/2$ . Alors pour  $A \in \mathcal{F}$  et  $n \in \mathbb{N}$ , on a

$$\mathbb{E}[|X_n|\mathbf{1}_A] = \mathbb{E}[|X_n|\mathbf{1}_{A \cap \{|X_n|>c\}}] + \mathbb{E}[|X_n|\mathbf{1}_{A \cap \{|X_n|\leq c\}}] \leq \frac{\varepsilon}{2} + c\mathbb{P}(A).$$

On obtient alors 1) avec  $\eta = \varepsilon/(2c)$  et 2) avec  $A = \Omega$ .

Réciproquement, on fixe  $\varepsilon > 0$  et on considère  $\eta > 0$  donné par 1) et  $M = \sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$  par 2). D'après l'inégalité de Markov, pour tout  $c \geq M/\eta$ , on a

$$\mathbb{P}(|X_n| > c) \leq \frac{\mathbb{E}[|X_n|]}{c} \leq \frac{M}{c} \leq \eta.$$

En appliquant le 1) pour chaque  $n$  avec  $A = \{|X_n| > c\}$ , on a  $\mathbb{E}[|X_n|\mathbf{1}_{\{|X_n|>c\}}] \leq \varepsilon$ .  $\square$

**Proposition 7.3.5** *Soit  $X$  une variable aléatoire intégrable. Alors la famille de variables aléatoires  $\{\mathbb{E}[X|\mathcal{G}] : \mathcal{G}$  sous-tribu de  $\mathcal{F}\}$  est uniformément intégrable.*

**Démonstration :** Notons  $X_{\mathcal{G}} = \mathbb{E}[X|\mathcal{G}]$ . Comme  $\{X_{\mathcal{G}} > c\}$  est  $\mathcal{G}$ -mesurable, par définition de l'espérance conditionnelle  $X_{\mathcal{G}} = \mathbb{E}[X|\mathcal{G}]$ , on a :

$$\mathbb{E}[|X_{\mathcal{G}}|\mathbf{1}_{\{|X_{\mathcal{G}}|>c\}}] = \mathbb{E}[|X|\mathbf{1}_{\{|X_{\mathcal{G}}|>c\}}]. \quad (7.11)$$

Mais par l'inégalité de Markov

$$\mathbb{P}(|X_{\mathcal{G}}| > c) \leq \frac{\mathbb{E}[|X_{\mathcal{G}}|]}{c} = \frac{\mathbb{E}[\mathbb{E}[|X|\mathcal{G}]]}{c} \leq \frac{\mathbb{E}[\mathbb{E}[|X| |\mathcal{G}]]}{c} = \frac{\mathbb{E}[|X|]}{c}.$$

Puis comme  $X$  est intégrable, pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$  tel que si  $\mathbb{P}(A) < \delta$  alors  $\mathbb{E}[|X|\mathbf{1}_A] < \varepsilon$ . Avec  $c > \mathbb{E}[|X|]/\delta$ , on a  $\mathbb{P}(|X_{\mathcal{G}}| > c) \leq \delta$  et donc  $\mathbb{E}[|X|\mathbf{1}_{\{|X_{\mathcal{G}}|>c\}}] < \varepsilon$ . Finalement, avec l'égalité (7.11), on a  $\mathbb{E}[|X_{\mathcal{G}}|\mathbf{1}_{\{|X_{\mathcal{G}}|>c\}}] < \varepsilon$ , c'est à dire

$$\lim_{c \rightarrow +\infty} \sup_{\mathcal{G} \text{ sous-tribu de } \mathcal{F}} \mathbb{E}[|X_{\mathcal{G}}|\mathbf{1}_{\{|X_{\mathcal{G}}|>c\}}] = 0,$$

ce qui prouve la Proposition 7.3.5.  $\square$

**Théorème 7.3.1 (Vitali)** Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires intégrables. Il y a équivalence entre

1.  $(X_n)_{n \geq 0}$  converge dans  $L^1$  ;
2.  $(X_n)_{n \geq 0}$  est une suite de Cauchy dans  $L^1$ , ie.  $\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n, m \geq n_0, \mathbb{E}[|X_n - X_m|] < \varepsilon$  ;
3.  $(X_n)_{n \geq 0}$  est uniformément intégrable et  $(X_n)_{n \geq 0}$  converge en probabilité.

**Remarque 7.3.2** — L'implication  $2) \Rightarrow 1)$  montre que  $L^1$  est un espace complet pour  $\|\cdot\|_1$ . Il s'agit du **théorème de Riesz-Fisher** (Th. 3.4.3) pour  $L^1$ .

— Le même résultat est vrai avec  $L^p$ ,  $p \geq 1$ , à la place de  $L^1$  si  $X_n^p$  est uniformément intégrable, cf. le Corollaire 7.3.1 pour un énoncé précis. En particulier,  $(L^p, \|\cdot\|_p)$  est complet (Th. 3.4.3).

**Démonstration :** 1)  $\Rightarrow$  2) : la convergence  $L^1$  implique la condition de Cauchy dans  $L^1$  (inégalité triangulaire).

2)  $\Rightarrow$  3) : Par la condition de Cauchy, pour tout  $\varepsilon > 0$ , il existe  $n_0 = n_0(\varepsilon)$  tel que si  $n, m \geq n_0$  alors  $\mathbb{E}[|X_n - X_m|] < \varepsilon/2$ . Ainsi pour tout  $A \in \mathcal{F}$  et  $n \geq n_0$ ,

$$\begin{aligned} \mathbb{E}[|X_n|\mathbf{1}_A] &\leq \mathbb{E}[|X_n - X_{n_0}|\mathbf{1}_A] + \mathbb{E}[|X_{n_0}|\mathbf{1}_A] \leq \mathbb{E}[|X_n - X_{n_0}|] + \mathbb{E}[|X_{n_0}|\mathbf{1}_A] \\ &\leq \varepsilon/2 + \mathbb{E}[|X_{n_0}|\mathbf{1}_A]. \end{aligned}$$

Il suit  $\sup_{n \geq n_0} \mathbb{E}[|X_n|\mathbf{1}_A] \leq \varepsilon/2 + \mathbb{E}[|X_{n_0}|\mathbf{1}_A]$  et

$$\sup_{n \geq 0} \mathbb{E}[|X_n|\mathbf{1}_A] \leq \varepsilon/2 + \sup_{k \leq n_0} \mathbb{E}[|X_k|\mathbf{1}_A].$$

Avec  $A = \Omega$ , on déduit  $\sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$ . Puis, comme la suite finie  $(X_k)_{k \leq n_0}$  est uniformément intégrable, il existe  $\eta > 0$  tel que  $\mathbb{P}(A) < \eta$  implique  $\sup_{k \leq n_0} \mathbb{E}[|X_k|\mathbf{1}_A] \leq \varepsilon/2$ . Finalement,  $\sup_{n \geq 0} \mathbb{E}[|X_n|\mathbf{1}_A] \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$  pour  $\mathbb{P}(A) < \eta$ . La Proposition 7.3.4 s'applique et donne l'uniforme intégrabilité de  $(X_n)_{n \geq 0}$ .

Ensuite, par l'inégalité de Markov,

$$\mathbb{P}(|X_n - X_m| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n - X_m|]}{\varepsilon} \rightarrow 0, \quad n, m \rightarrow +\infty.$$

La suite  $(X_n)_{n \geq 0}$  est donc de Cauchy en probabilité, ce qui donne la convergence en probabilité de toute la suite d'après la Proposition 7.2.8.

3)  $\Rightarrow$  1) : Comme  $X_n$  converge en probabilité vers  $X$ , de toute sous-suite  $(n') \subset (n)$ , on peut extraire une sous-suite  $(n'') \subset (n')$  telle que  $X_{n''}$  converge ps vers  $X$ . Le lemme de Fatou, avec l'uniforme intégrabilité, garantit  $X \in L^1$  :

$$\mathbb{E}[|X|] = \mathbb{E}\left[\liminf_{n'' \rightarrow +\infty} |X_{n''}|\right] \leq \liminf_{n'' \rightarrow +\infty} \mathbb{E}[|X_{n''}|] \leq \sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$$

d'après l'uniforme intégrabilité. Puis pour tout  $\varepsilon > 0$ , on a

$$\begin{aligned} \mathbb{E}[|X_n - X|] &\leq \mathbb{E}[|X_n - X| \mathbf{1}_{\{|X_n - X| < \varepsilon\}}] + \mathbb{E}[|X_n - X| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] \\ &\leq \mathbb{E}[|X_n - X| \mathbf{1}_{\{|X_n - X| < \varepsilon\}}] + \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] + \mathbb{E}[|X| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] \\ &\leq \varepsilon + \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] + \mathbb{E}[|X| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}]. \end{aligned}$$

Comme  $\{X, X_1, \dots, X_n, \dots\}$  est uniformément intégrable, il existe  $\eta$  tel que pour  $\mathbb{P}(A) \leq \eta$  :

$$\mathbb{E}[|X_n| \mathbf{1}_A] < \varepsilon, \quad \mathbb{E}[|X| \mathbf{1}_A] < \varepsilon.$$

Puis d'après la convergence en probabilité  $X_n \xrightarrow{\mathbb{P}} X$  pour  $n$  assez grand  $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \eta$  si bien que

$$\mathbb{E}[|X_n| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] < \varepsilon, \quad \mathbb{E}[|X| \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] < \varepsilon.$$

Finalement, pour  $n$  assez grand, on a  $\mathbb{E}[|X_n - X|] \leq 3\varepsilon$ , ce qui prouve le 1). □

**Corollaire 7.3.1 (Vitali version  $L^p$ )** Soient  $p \geq 1$  et  $(X_n)_{n \geq 0}$  une suite de variables aléatoires de  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ . Il y a équivalence entre

1.  $(X_n)_{n \geq 0}$  converge dans  $L^p$ .
2.  $(X_n)_{n \geq 0}$  est une suite de Cauchy dans  $L^p$ , ie.  $\mathbb{E}[|X_n - X_m|^p] \rightarrow 0$ ,  $m, n \rightarrow +\infty$ .
3.  $(X_n^p)_{n \geq 0}$  est uniformément intégrable et  $(X_n)_{n \geq 0}$  converge en probabilité.

**Démonstration :** La preuve est similaire à celle du Th. 7.3.1 en utilisant les propriétés de la norme  $\|\cdot\|_p$ . □

## 7.4 Convergence en loi

**Définition 7.4.1 (Convergence en loi)** Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires réelles. On dit que  $(X_n)_{n \geq 0}$  converge en loi vers la variable aléatoire  $X$  si la fonction de répartition  $F_n(t)$  de  $X_n$  converge vers  $F(t)$ , celle de  $X$  en tout point  $t$  où  $F$  est continue (c'est à dire en les points  $t$  tels que  $\mathbb{P}(X = t) = 0$ ). On note la convergence en loi :  $X_n \xrightarrow{\text{en loi}} X$ .

On peut noter que, pour cette convergence, les variables aléatoires ne sont pas nécessairement définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Cela est par contre nécessaire pour les convergences presque sûre, en probabilité, en norme  $\|\cdot\|_p$ .

**Remarque 7.4.1** Si  $X$  est une variable aléatoire à densité, sa fonction de répartition est continue et il n'y a plus d'hypothèse restrictive à faire sur les points de continuité  $t$  dans la définition de la convergence en loi puisque pour tout  $t$  on a  $\mathbb{P}(X = t) = 0$ .

**Exemple :** Soit  $X_n$  de loi de Bernoulli donnée par  $\mathbb{P}(X_n = 1 + 1/n) = \mathbb{P}(X_n = 0) = \frac{1}{2}$ . Alors la suite  $(X_n)_{n \geq 0}$  converge en loi vers  $X$  de loi  $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$ .

**Théorème 7.4.1** La convergence en loi  $X_n \Rightarrow X$  est équivalente à avoir pour toute fonction  $g : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée :

$$\lim_{n \rightarrow +\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]. \quad (7.12)$$

**Remarque 7.4.2** En fait, la définition de la convergence en loi par les fonctions de répartition ne permet de considérer la convergence en loi que pour les variables ou vecteurs aléatoires à valeurs dans un espace fini-dimensionnel (ie. ce qu'on a appelé variables aléatoires réelles ou vecteurs aléatoires). Par contre, la condition (7.12) permet de définir la convergence en loi pour les variables aléatoires à valeurs dans un espace métrique  $S$  quelconque avec  $g : S \rightarrow \mathbb{R}$  continue bornée. En général, dans les espaces métriques, on prend donc (7.12) pour définition de la convergence en loi.

**Démonstration :** • **Sens direct.** On suppose d'abord  $g$  de classe  $C^1$  à support compact. En particulier  $g'$  est bornée et à support compact et la mesure  $|g'(y)|dy$  a sa masse totale finie :

$$\int_{\mathbb{R}} |g'(y)|dy \leq \sup_{y \in \text{Supp}(g')} (|g'(y)|) \lambda(\text{Supp}(g')) < +\infty. \quad (7.13)$$

Par le théorème de Fubini, on a

$$\begin{aligned} \mathbb{E}[g(X_n)] &= \int_{\mathbb{R}} g(x)\mathbb{P}_{X_n}(dx) = \int_{\mathbb{R}} \left( \int_{-\infty}^x g'(y)dy \right) \mathbb{P}_{X_n}(dx) \\ &= \int_{\mathbb{R}} g'(y) \int_y^{+\infty} \mathbb{P}_{X_n}(dx) dy = \int_{\mathbb{R}} (1 - F_n(y))g'(y)dy. \end{aligned}$$

(Noter que le théorème de Fubini s'applique puisque, d'après (7.13),  $g'(y)$  est intégrable par rapport à  $dy\mathbb{P}_{X_n}(dx)$ .) La fonction  $F$  a au plus un nombre dénombrable de discontinuité et, en ses points de continuité,  $1 - F_n$  converge vers  $1 - F$  (par convergence en loi) en restant bornée par 1 (car une fonction répartition est à valeurs dans  $[0, 1]$ ). Par convergence dominée (dans le membre de droite), on a :

$$\lim_{n \rightarrow +\infty} \mathbb{E}[g(X_n)] = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}} (1 - F_n(y))g'(y)dy = \int_{\mathbb{R}} (1 - F(y))g'(y)dy = \mathbb{E}[g(X)]$$

ce qui prouve (7.12) lorsque  $g \in C_c^1$ .

On suppose maintenant que  $g$  est continue à support compact mais pas nécessairement dérivable. Pour tout  $\varepsilon > 0$ , il existe une fonction  $h$  de classe  $C^1$  à support compact telle que  $|g(x) - h(x)| \leq \frac{\varepsilon}{4}$  pour tout  $x \in \mathbb{R}$  (par exemple, prendre  $h = h_k = g * e_k$  la convolée de  $g$  avec une approximation de l'unité  $e_k$  qui est  $C^1$  à support compact assez petit). D'après le cas précédent, il existe  $n_0$  tel que pour  $n \geq n_0$ , on a  $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \frac{\varepsilon}{2}$ . Pour  $n \geq n_0$ , on a alors :

$$\begin{aligned} & |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \\ & \leq |\mathbb{E}[g(X_n)] - \mathbb{E}[h(X_n)]| + |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| + |\mathbb{E}[h(X)] - \mathbb{E}[g(X)]| \\ & \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon \end{aligned}$$

ce qui prouve (7.12) lorsque  $g \in C_c^0$ .

On suppose enfin que  $g$  est continue et bornée (et, pour simplifier, disons par 1). Étant donné  $\varepsilon > 0$ , il existe  $h$  continue à support compact telle que  $0 \leq h \leq 1$  et avec  $\mathbb{E}[h(X)] > 1 - \varepsilon/5$  :  $\lim_{n \rightarrow +\infty} \mathbb{P}(X \in [-n, n]) = 1$  donc il existe  $n_0$  tel que  $\mathbb{P}(X \in [-n_0, n_0]) \geq 1 - \varepsilon/5$ , on prend alors

$h(x) = 1$  si  $x \in [-n_0, n_0]$ ,  $h(x) = 0$  si  $x \notin [-n_0 - 1, n_0 + 1]$  et  $h$  linéaire sur  $[-n_0 - 1, -n_0]$  et  $[n_0, n_0 + 1]$ ; on a alors  $\mathbb{E}[h(X)] \geq \mathbb{E}[\mathbf{1}_{[-n_0, n_0]}(X)] = \mathbb{P}(X \in [-n_0, n_0]) \geq 1 - \varepsilon/5$ .

On déduit, par la convergence en loi avec  $h$  continue à support compact, qu'il existe  $n_1$  tel que pour  $n \geq n_1$ , on a  $\mathbb{E}[h(X_n)] > 1 - \varepsilon/4$ . La fonction  $gh$  est continue à support compact. Il existe donc un entier  $n_2$  tel que pour  $n \geq n_2$  on a

$$|\mathbb{E}[(gh)(X_n)] - \mathbb{E}[(gh)(X)]| \leq \frac{\varepsilon}{2}.$$

On a alors pour tout  $n \geq \max(n_1, n_2)$  (en écrivant  $g = gh + g(1 - h)$ ) :

$$\begin{aligned} & |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \\ & \leq |\mathbb{E}[(gh)(X_n)] - \mathbb{E}[(gh)(X)]| + |\mathbb{E}[g(1 - h)(X_n)]| + |\mathbb{E}[g(1 - h)(X)]| \\ & \leq \frac{\varepsilon}{2} + \mathbb{E}|(1 - h)(X_n)| + \mathbb{E}|(1 - h)(X)| \leq \varepsilon \end{aligned}$$

ce qui termine la preuve du sens direct.

• **Sens réciproque.** Soit  $x \in \mathbb{R}$  un point de continuité de  $F$ , la fonction de répartition de  $X$ . Choisissons  $y > x$  (à préciser dans un instant) et soit  $f$  la fonction égale à 1 pour  $u \leq x$  et égale à 0 pour  $u \geq y$ , et affine entre  $x$  et  $y$ . Cette fonction est continue et bornée par 1. Puis comme  $\mathbf{1}_{(-\infty, x]} \leq f \leq \mathbf{1}_{(-\infty, y]}$ , on a

$$\mathbb{P}(X_n \leq x) = \mathbb{E}[\mathbf{1}_{(-\infty, x]}(X_n)] \leq \mathbb{E}[f(X_n)] \leq \mathbb{E}[\mathbf{1}_{(-\infty, y]}(X_n)] = \mathbb{P}(X_n \leq y) \quad (7.14)$$

c'est à dire  $F_n(x) \leq \mathbb{E}[f(X_n)] \leq F_n(y)$ . De la même manière avec  $X$  à la place de  $X_n$ , on a  $F(x) \leq \mathbb{E}[f(X)] \leq F(y)$ .

Fixons  $\varepsilon > 0$ , par continuité (à droite) de  $F$  en  $x$ , il existe  $y_0 \geq x$  tel que  $F(x) \leq F(y_0) \leq F(x) + \varepsilon/2$ . On choisit la fonction  $f_0$  associée comme précédemment à ce  $y_0$ . Par hypothèse pour cette fonction  $f_0$  continue bornée, il existe  $n_0$  tel que

$$\forall n \geq n_0, \quad \mathbb{E}[f_0(X_n)] \leq \mathbb{E}[f_0(X)] + \varepsilon/2.$$

Il en résulte que pour tout  $n \geq n_0$ ,

$$F_n(x) \leq \mathbb{E}[f_0(X_n)] \leq \mathbb{E}[f_0(X)] + \varepsilon/2 \leq F(y_0) + \varepsilon/2 \leq F(x) + \varepsilon. \quad (7.15)$$

En choisissant de façon analogue  $y < x$ , on construit des fonctions  $\tilde{f}$  pour lesquelles on obtient des inégalités inversées :

$$F_n(y) \leq \mathbb{E}[\tilde{f}(X_n)] \leq F_n(x), \quad F(y) \leq \mathbb{E}[\tilde{f}(X)] \leq F(x). \quad (7.16)$$

Par continuité (à gauche) de  $F$  en  $x$ , on peut alors choisir  $y_1 \leq x$  tel que  $F(x) - \varepsilon/2 \leq F(y_1) \leq F(x)$ ; puis pour la fonction  $\tilde{f}_1$  associée à ce  $y_1$  comme ci-dessus, l'hypothèse donne l'existence de  $n_1$  tel que

$$\forall n \geq n_1, \quad \mathbb{E}[\tilde{f}_1(X_n)] \geq \mathbb{E}[\tilde{f}_1(X)] - \varepsilon/2$$

puis pour tout  $n \geq n_1$ ,

$$F_n(x) \geq \mathbb{E}[\tilde{f}_1(X_n)] \geq \mathbb{E}[\tilde{f}_1(X)] - \varepsilon/2 \geq F(y_1) - \varepsilon/2 \geq F(x) - \varepsilon. \quad (7.17)$$

Finalement avec (7.15), (7.17), on a obtenu la convergence de  $F_n(x)$  vers  $F(x)$  : pour tout  $n \geq \max(n_1, n_2)$ , on a

$$F(x) - \varepsilon \leq F_n(x) \leq F(x) + \varepsilon.$$

□

Un des intérêts de la convergence en loi est de permettre d'approcher les lois inconnues (ou connues mais, parfois, difficilement calculables) des  $X_n$  par celle (souvent plus simple) de la limite  $X$ , comme par exemple dans le théorème central limite (cf. Section 8.2). D'où l'importance de critères simples garantissant cette convergence sans repasser par la définition. On en donne ci-dessous quelques uns : le Théorème 7.4.2 (plutôt utile dans les contextes théoriques) et le Théorème 7.4.3 (plus pratique à l'aide des fonctions caractéristiques).

**Théorème 7.4.2 (Portmanteau)** *Les assertions suivantes sont équivalentes quand  $n \rightarrow +\infty$  pour des variables aléatoires  $X_n$  à valeurs dans un espace métrique  $S$ .*

1.  $X_n \Rightarrow X$ , ie. pour toute fonction  $f : S \rightarrow \mathbb{R}$  continue et bornée

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

2. Pour toute fonction  $f : S \rightarrow \mathbb{R}$  uniformément continue et bornée

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

3. Pour tout fermé  $F$ ,  $\limsup_{n \rightarrow +\infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ .
4. Pour tout ouvert  $G$ ,  $\liminf_{n \rightarrow +\infty} \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ .
5. Pour tout  $A$  tel que  $\mathbb{P}(X \in \bar{A} \setminus A^\circ) = 0$ , on a  $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in A) = \mathbb{P}(X \in A)$ .

**Théorème 7.4.3** La convergence en loi  $X_n \Rightarrow X$  est équivalente à la convergence simple des fonctions caractéristiques :

$$\lim_{n \rightarrow +\infty} \varphi_{X_n}(t) = \varphi_X(t), \quad \forall t \in \mathbb{R}.$$

**Démonstration :** Le sens direct est immédiat avec le Théorème 7.4.1 puisqu'on a  $\varphi_X(t) = \mathbb{E}[g_t(X)]$  avec  $g_t(x) = e^{itx}$  continue bornée.

Pour la réciproque, on suppose d'abord que  $g : \mathbb{R} \rightarrow \mathbb{R}$  est  $C^2$  à support compact. Sa transformée de Fourier  $\hat{g}(t) = \int_{\mathbb{R}} e^{itx} g(x) dx$  est alors dans  $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ . En effet,  $\hat{g}$  est bornée ainsi que celle de  $g''$  qui est égale à  $t^2 \hat{g}(t)$ . On a donc  $\hat{g}(t) = O(1/t^2)$  en  $+\infty$  et  $\hat{g}$  est bornée si bien que  $\hat{g} \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ .

Comme  $\hat{g} \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ , on peut appliquer le théorème d'inversion de Fourier pour écrire

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \hat{g}(t) dt.$$

Comme  $\hat{g}$  est intégrable, le théorème de Fubini s'applique et donne :

$$\begin{aligned} \mathbb{E}[g(X_n)] &= \frac{1}{2\pi} \mathbb{E} \left[ \int_{\mathbb{R}} e^{-itX_n} \hat{g}(t) dt \right] = \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}[e^{-itX_n}] \hat{g}(t) dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_{X_n}(-t) \hat{g}(t) dt. \end{aligned}$$

Mais comme  $\varphi_{X_n}(-t)$  converge vers  $\varphi_X(-t)$  pour chaque  $t$  et que  $|\varphi_{X_n}(-t)| \leq 1$  pour tout  $t \in \mathbb{R}$  avec en plus  $\hat{g}$  intégrable, par le théorème de convergence dominée, on a

$$\lim_{n \rightarrow +\infty} \mathbb{E}[g(X_n)] = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(-t) \hat{g}(t) dt = \mathbb{E}[g(X)].$$

Lorsque l'on suppose seulement  $g$  continue à support compact puis ensuite continue bornée, on procède comme précédemment dans la preuve du sens direct du Théorème 7.4.1.  $\square$

Une version plus générale du Théorème 7.4.3 donne la convergence en loi quand il y a convergence simple des fonctions caractéristiques avec une hypothèse simple sur la limite pour qu'elle soit une fonction caractéristique.

**Théorème 7.4.4 (Paul Lévy)** On considère une suite de fonctions caractéristiques  $(\varphi_n)_{n \geq 0}$  de variables aléatoires  $X_n$  réelles. On suppose que, pour chaque  $t \in \mathbb{R}$ ,  $\lim_{n \rightarrow +\infty} \varphi_n(t) = \varphi(t)$  et que  $\varphi$  est une fonction continue en  $t = 0$ . Alors  $\varphi$  est la fonction caractéristique d'une variable aléatoire  $X$  et on a la convergence en loi  $X_n \Rightarrow X$ ,  $n \rightarrow +\infty$ .

**Démonstration :** On note  $F_n(x) = \mathbb{P}(X_n \leq x)$  la fonction de répartition de chaque variable aléatoire  $X_n$ .

**Étape 1.** Soit  $\{r_1, r_2, \dots, r_n, \dots\}$  une énumération de  $\mathbb{Q}$ . Pour chaque  $j \geq 1$ , on considère la suite  $(F_n(r_j))_n$ . Comme c'est une suite de  $[0, 1]$ , elle contient une sous-suite convergente (Théorème de Bolzano-Weierstrass). Par une procédure d'extraction diagonale, on peut trouver une sous-suite  $G_k = F_{n_k}$  de  $F_n$  telle que  $\lim_{k \rightarrow +\infty} G_k(r) = b_r$  existe pour tout  $r \in \mathbb{Q}$ . Comme  $F_{n_k}$  est croissante, on a facilement  $b_r \leq b_{r'}$  si  $r < r'$ .

**Étape 2.** On reconstruit une fonction à partir du squelette  $b_r$ ,  $r \in \mathbb{Q}$ . Pour cela, on note  $G(t) = \inf_{r > t} b_r$ . Il est clair que  $G$  est croissante car si  $t < t'$  alors  $G(t) \leq G(t')$ . Puis si  $t_n \searrow t$  et si  $r > t$  alors, pour  $n$  assez grand,  $r > t_n$ , on en déduit que  $G(t) = \inf_{n \geq 1} G(t_n)$  pour toute suite  $t_n \searrow t$ . Cela justifie la continuité à droite de  $G$ .

**Étape 3.** On montre qu'en tout point de continuité  $t$  de  $G$  alors  $\lim_{n \rightarrow +\infty} G_n(t) = G(t)$ . Soit  $r \in \mathbb{Q}$ ,  $r > t$ . Alors  $G_n(t) \leq G_n(r)$  et  $\lim_{n \rightarrow +\infty} G_n(r) = b_r$ . On en déduit que  $\limsup_{n \rightarrow +\infty} G_n(t) \leq b_r$ . Comme c'est vrai pour tout  $r \in \mathbb{Q}$ ,  $r > t$ , on a  $\limsup_{n \rightarrow +\infty} G_n(t) \leq G(t) = \inf_{r > t} b_r$ . Soit maintenant  $s < t$  et  $r \in \mathbb{Q}$  tel que  $s < r < t$ . Alors  $\liminf_{n \rightarrow +\infty} G_n(t) \geq \liminf_{n \rightarrow +\infty} G_n(r) = b_r \geq G(s)$ . Comme cela est vrai pour chaque  $s < t$ , on en déduit  $\liminf_{n \rightarrow +\infty} G_n(t) \geq \sup_{s < t} G_n(s) = G(t^-) = G(t)$  puisque  $t$  est un point de continuité de  $G$ . Finalement, pour un tel  $t$ , on a  $G_n(t) \rightarrow G(t)$ .

Attention, cependant, à ce stade  $G$  n'est pas nécessairement une fonction de répartition, par exemple  $F_n(t) = \delta_n((-\infty, t])$  dont la limite  $G(t)$  existe et est nulle pour tout  $t$  mais  $G \equiv 0$  n'est pas une fonction de répartition.

**Étape 4.** On utilise le résultat suivant :

**Lemme 7.4.1** Soit  $X$  une variable aléatoire de fonction caractéristique  $\psi$  et de fonction de répartition  $G$ . Alors pour tout  $T > 0$ ,

$$1 - G(2/T) + G(-2/T) \leq 2 \left( 1 - \frac{1}{2T} \int_{-T}^T \psi(t) dt \right).$$

D'après ce lemme, pour tout  $k \geq 1$ , on a

$$1 - G_k(2/T) + G_k(-2/T) = 1 - F_{n_k}(2/T) + F_{n_k}(-2/T) \leq 2 \left( 1 - \frac{1}{2T} \int_{-T}^T \varphi_{n_k}(t) dt \right).$$

Il est possible de choisir  $T$  tel que  $\pm \frac{2}{T}$  soient des points de continuité de  $G$ . Par continuité de  $G$  (à gauche) et convergence dominée (à droite), avec  $k \rightarrow +\infty$ , on a

$$1 - G(2/T) + G(-2/T) \leq 2 \left( 1 - \frac{1}{2T} \int_{-T}^T \varphi(t) dt \right).$$

Comme  $\varphi(0) = 1$ , et  $\varphi$  est continue en 0, on a

$$\lim_{T \rightarrow 0} \frac{1}{2T} \int_{-T}^T \varphi(t) dt = 1.$$

On laisse alors  $T \rightarrow 0$  de façon que  $\pm \frac{2}{T}$  restent des points de continuité de  $G$ . On en déduit que  $1 - G(+\infty) + G(-\infty) = 0$ , ce qui exige  $G(-\infty) = 0$  et  $G(+\infty) = 1$ .

Finalement,  $G$  est croissante, continue à droite avec des limites à gauche de limite 0 et 1 en  $-\infty$  et  $+\infty$ . D'après le Théorème 2.3.1,  $G$  est une fonction de répartition.

On note  $Q$  la loi de fonction de répartition  $G$ . Cette loi est unique par caractérisation des lois par leur fonction de répartition.

**Étape 5.** Dans l'étape 1, on a montré qu'en tout point de continuité de  $G$ , fonction de répartition, on a

$$\lim_{k \rightarrow +\infty} G_k = \lim_{k \rightarrow +\infty} F_{n_k} = G.$$

On a donc  $\mathbb{P}_{X_{n_k}} \Rightarrow Q$ ,  $k \rightarrow +\infty$ , et donc  $\varphi_{n_k} \rightarrow \varphi_Q$  d'après le Théorème 7.4.1. Comme par hypothèse  $\varphi_n \rightarrow \varphi$ , cela exige  $\varphi_Q = \varphi$ .

Toute sous-suite  $(n')$  telle que  $\mathbb{P}_{X_{n'}}$  converge en loi doit converger vers  $Q$  puisque on doit avoir  $\varphi_{n'} \rightarrow \varphi$  et que  $\varphi$  est nécessairement la fonction caractéristique de la loi limite, cette loi limite doit donc être  $Q$ . Mais si  $X_n \not\Rightarrow Q$  alors il existe  $g$  continue bornée telle que  $\mathbb{E}[g(X_n)] \not\rightarrow \mathbb{E}[g(X)]$  où  $X \sim Q$ , c'est à dire qu'il existe  $\varepsilon > 0$  et une sous-suite  $(n')$  tels que

$$|\mathbb{E}[g(X_{n'})] - \mathbb{E}[g(X)]| \geq \varepsilon. \quad (7.18)$$

Mais en appliquant les étapes 1, 2 et 3 de cette preuve à  $X_{n'}$ , on montre que  $X_{n'}$  converge en loi. D'après ce qui précède, cette loi doit nécessairement être  $Q$ , ce qui contredit (7.18) pour  $n'$  assez grand.  $\square$

**Démonstration :** [Lemme 7.4.1] Soit  $X$  une variable aléatoire de fonction caractéristique  $\psi$  et soit  $T > 0$ . D'après le théorème de Fubini,

$$\begin{aligned} \frac{1}{2T} \int_{-T}^T \psi(t) dt &= \mathbb{E} \left[ \frac{1}{2T} \int_{-T}^T e^{itX} dt \right] = \mathbb{E} \left[ \frac{\sin(TX)}{TX} \right] \leq \mathbb{E} \left[ \left| \frac{\sin(TX)}{TX} \right| \right] \\ &\leq \mathbb{E} \left[ \left| \frac{\sin(TX)}{TX} \mathbf{1}_{\{|X|<\alpha\}} \right| \right] + \mathbb{E} \left[ \left| \frac{\sin(TX)}{TX} \mathbf{1}_{\{|X|\geq\alpha\}} \right| \right] \\ &\leq \mathbb{E}[\mathbf{1}_{\{|X|<\alpha\}}] + \frac{1}{T\alpha} \mathbb{E}[\mathbf{1}_{\{|X|\geq\alpha\}}] \end{aligned}$$

où dans la dernière égalité, on a utilisé  $|\sin x| \leq |x|$  et  $|\sin x| \leq 1$ . On en déduit

$$\begin{aligned} 1 - \frac{1}{2T} \int_{-T}^T \psi(t) dt &\geq 1 - \mathbb{P}(|X| < \alpha) - \frac{1}{T\alpha} \mathbb{P}(|X| \geq \alpha) \\ &= \left(1 - \frac{1}{T\alpha}\right) \mathbb{P}(|X| \geq \alpha) \geq \left(1 - \frac{1}{T\alpha}\right) (1 - G(\alpha) + G(-\alpha)) \end{aligned}$$

car  $\mathbb{P}(X < \alpha) \leq \mathbb{P}(X \leq \alpha) = G(\alpha)$ . On conclut en prenant  $\alpha = \frac{2}{T}$ .  $\square$

**Théorème 7.4.5 (Continuous mapping theorem)** Soit  $f : S \rightarrow S'$  une application entre espaces métriques et  $(X_n)_{n \geq 0}$  une suite de variables aléatoires à valeurs dans  $S$ . Alors si  $X_n \Rightarrow X$  et  $f$  est continue  $\mathbb{P}_X$ -presque partout, on a aussi  $f(X_n) \Rightarrow f(X)$ ,  $n \rightarrow +\infty$ .

**Démonstration :** Notons  $Y_n = f(X_n)$ . Si  $f$  est continue alors le résultat vient facilement du Théorème 7.4.1 : on considère  $g$  continue bornée,  $g \circ f$  l'est encore et le Théorème 7.4.1 (sens direct) donne pour la convergence  $X_n \Rightarrow X$

$$\mathbb{E}[g(Y_n)] = \mathbb{E}[(g \circ f)(X_n)] \rightarrow \mathbb{E}[(g \circ f)(X)] = \mathbb{E}[g(Y_n)].$$

Le sens indirect du Théorème 7.4.1 assure alors  $X_n \Rightarrow X$ .

Dans le cas général où  $f$  est seulement  $\mathbb{P}_X$ -ps continue, il faut utiliser le Théorème 7.4.2 (Portmanteau).

□

La convergence en loi ne vérifie pas de bonnes propriétés arithmétiques : si par exemple  $X_n$  tend en loi vers  $X$ , il n'est pas vrai que  $X_n - X$  tend vers 0. En effet, prendre par exemple,  $X_n$  et  $X$  des variables aléatoires de loi normale centrée réduite  $\mathcal{N}(0, 1)$  alors la loi des  $X_n$  coïncide avec celle de  $X$  mais aussi celle de  $-X$  (car  $X$  et  $-X$  ont même loi, par parité de la densité de  $\mathcal{N}(0, 1)$ !). Donc les  $X_n$  convergent en loi vers  $-X$ , pourtant si  $X_n \perp\!\!\!\perp X$ ,  $X_n - X$  suit une loi normale  $\mathcal{N}(0, 2)$  qui n'est pas la loi de  $0 = X - X$ . Cependant, dans certains cas, le lemme de Slutsky donne des résultats positifs.

**Théorème 7.4.6 (Slutsky)** Soit  $X_n \xrightarrow{*} X$  et  $Y_n \Rightarrow c$  où  $c$  est une constante. Alors

$$(X_n, Y_n) \xrightarrow{*} (X, c) \tag{7.19}$$

où la convergence  $\xrightarrow{*}$  est en loi ou en probabilité selon le même mode que  $X_n$ . De plus

1.  $X_n + Y_n \xrightarrow{\mathbb{P}} X + c$
2.  $X_n Y_n \xrightarrow{\mathbb{P}} cX$
3. si  $c \neq 0$ ,  $X_n / Y_n \xrightarrow{\mathbb{P}} X/c$ .

On commence par une version plus faible de Slutsky qui contient l'essentiel de la difficulté et à laquelle on va se ramener :

**Proposition 7.4.1 (Slutsky)** Si la suite  $(X_n)_{n \geq 0}$  converge en loi vers  $X$  et si  $(Y_n)_{n \geq 0}$  converge en loi vers 0 alors  $X_n + Y_n$  converge en loi vers  $X$ .

**Remarque 7.4.3** On peut expliquer intuitivement ce résultat : dire que  $(X_n)_{n \geq 0}$  tend en loi vers  $X$ , c'est dire que les répartitions statistiques de  $X_n$  et  $X$  ont tendance à être voisines. Dire que  $Y_n$  tend vers 0, c'est dire que statistiquement,  $Y_n$  est souvent voisin de 0. Il en résulte qu'ajouter  $Y_n$  à  $X_n$  modifie souvent peu  $X_n$  et donc également la répartition statistique des valeurs prises par  $X_n$ .

**Démonstration :** Soit  $t \in \mathbb{R}$  tel que la fonction de répartition  $F_X$  de  $X$  soit continue en  $t$ . Soit  $\alpha > 0$ , pour avoir  $X_n + Y_n \leq t$ , on a soit  $X_n \leq t + \alpha$ , soit  $Y_n \leq -\alpha$ , c'est à dire  $\{X_n + Y_n \leq t\} \subset \{X_n \leq t + \alpha\} \cup \{Y_n \leq -\alpha\}$  d'où

$$\mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(X_n \leq t + \alpha) + \mathbb{P}(Y_n \leq -\alpha).$$

D'une façon analogue

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X_n \leq t - \alpha) - \mathbb{P}(Y_n \geq \alpha).$$

Fixons  $\varepsilon > 0$ , par continuité de  $F_X$  en  $t$ , il existe  $\alpha$  tel que

$$F_X(t) = \mathbb{P}(X \leq t) \leq F_X(t + \alpha) = \mathbb{P}(X \leq t + \alpha) \leq \mathbb{P}(X \leq t) + \frac{\varepsilon}{3}.$$

On peut s'arranger pour choisir  $\alpha$  tel que  $F_X$  soit continue en  $t + \alpha$ . En effet, comme  $F_X$  est croissante, (c'est une fonction réglée et donc) elle a au plus une infinité dénombrable de points de discontinuité, en particulier ses points de continuité sont partout denses. On peut donc, si besoin, choisir  $\alpha'$  tel que  $t < t + \alpha' < t + \alpha$  et tel que  $F_X$  soit continue en  $t + \alpha'$ . On a alors par croissance de  $F_X$

$$\mathbb{P}(X \leq t) \leq \mathbb{P}(X \leq t + \alpha') \leq \mathbb{P}(X \leq t + \alpha) \leq \mathbb{P}(X \leq t) + \frac{\varepsilon}{3}.$$

De plus, par la convergence en loi et la continuité de  $F_X$  en  $t + \alpha'$  :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \leq t + \alpha') = \mathbb{P}(X \leq t + \alpha').$$

Il existe donc  $n_0$  tel que pour tout  $n \geq n_0$ ,

$$\mathbb{P}(X_n \leq t + \alpha') \leq \mathbb{P}(X \leq t + \alpha') + \frac{\varepsilon}{3}.$$

D'autre part, la convergence en loi de  $Y_n$  vers 0 entraîne que  $\mathbb{P}(Y_n \leq -\alpha')$  tend vers  $\mathbb{P}(0 \leq -\alpha') = 0$  (la fonction de répartition de la variable aléatoire 0 est continue en  $-\alpha'$ ). Il existe donc  $n_1$  tel que pour  $n \geq n_1$ ,

$$\mathbb{P}(Y_n \leq -\alpha') \leq \frac{\varepsilon}{3}.$$

Finalement, pour tout  $n \geq N_0 = \max(n_0, n_1)$ , on a

$$\begin{aligned} \mathbb{P}(X_n + Y_n \leq t) &\leq \mathbb{P}(X_n \leq t + \alpha') + \mathbb{P}(Y_n \leq -\alpha') \\ &\leq \mathbb{P}(X \leq t + \alpha') + \varepsilon/3 + \varepsilon/3 \\ &\leq \mathbb{P}(X \leq t) + \varepsilon. \end{aligned}$$

On montre de même (en raisonnant de l'autre côté) l'existence de  $N_1$  tel que pour tout  $n \geq N_1$ ,

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X \leq t) - \varepsilon.$$

La convergence en loi en résulte : pour tout  $n \geq \max(N_0, N_1)$ , on a en tout point  $t$  de continuité de  $F_X$  :

$$\mathbb{P}(X \leq t) - \varepsilon \leq \mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(X \leq t) + \varepsilon$$

soit

$$F_X(t) - \varepsilon \leq F_{X_n+Y_n}(t) \leq F_X(t) + \varepsilon.$$

□

**Démonstration :** [Théorème de Slutsky] Les affirmations 1), 2), 3) sont des conséquences faciles de (7.19) combiné avec le « continuous mapping theorem » (Théorème 7.4.5) appliqué avec les fonctions continues  $g_1(x, y) = x + y$ ,  $g_2(x, y) = xy$  et  $g_3(x, y) = x/y$ . Il reste à voir (7.19).

Montrons le d'abord dans le cas de l'hypothèse  $X_n \Rightarrow X$ . Pour cela, on considère  $f(x, y)$  continue bornée et on montre que  $\mathbb{E}[f(X_n, Y_n)] \rightarrow \mathbb{E}[f(X, c)]$ . Pour  $g(x) = f(x, c)$ , on a  $g$  continue bornée et donc  $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$ , ie.  $\mathbb{E}[f(X_n, c)] \rightarrow \mathbb{E}[f(X, c)]$ , ou encore  $(X_n, c) \Rightarrow (X, c)$ . Puis  $|(X_n, Y_n) - (X_n, c)|_1 = |Y_n - c| \xrightarrow{\mathbb{P}} 0$  ( $Y_n$  converge en loi vers une constante donc aussi en probabilité, cf. Proposition 7.4.4) ce qui garantit que

$$(X_n, Y_n) = (X_n, c) + (X_n, Y_n) - (X_n, c) \Rightarrow (X, c) + 0 = (X, c)$$

puisque d'après la Proposition 7.4.1, si  $Z_n \Rightarrow Z$  et  $\alpha_n \xrightarrow{\mathbb{P}} 0$  alors  $Z_n + \alpha_n \Rightarrow Z$ .

De la même façon, si  $X_n \xrightarrow{\mathbb{P}} X$  alors  $(X_n, c) \xrightarrow{\mathbb{P}} (X, c)$  car

$$\mathbb{P}(|(X_n, c) - (X, c)| \geq \epsilon) = \mathbb{P}(|X_n - X| \geq \epsilon) \rightarrow 0, n \rightarrow +\infty,$$

et on a toujours  $|(X_n, Y_n) - (X_n, c)| = |Y_n - c| \xrightarrow{\mathbb{P}} 0$ . Finalement, on a :

$$(X_n, Y_n) = (X_n, c) + (X_n, Y_n) - (X_n, c) \xrightarrow{\mathbb{P}} (X, c) + 0 = (X, c).$$

□

### 7.4.1 Convergence en loi et autres convergences

**Proposition 7.4.2 (Convergences ps et en loi)** Si la suite  $(X_n)_{n \geq 0}$  converge presque sûrement vers  $X$ , elle converge en loi vers  $X$ .

**Démonstration :** On pourrait le justifier en utilisant que la convergence en probabilité (impliquée par la convergence presque sûre cf. Proposition 7.2.3) implique la convergence en loi (cf. ci-dessous la Proposition 7.4.3).

On en donne une preuve indépendante à l'aide du critère de convergence en loi énoncé par le

**Théorème 7.4.3.** Soit  $f$  une fonction continue bornée. Si  $X_n \rightarrow X$  ps, alors  $f(X_n) \rightarrow f(X)$  presque sûrement et  $f(X_n)$  est bornée par  $\|f\|_\infty$ , qui comme toutes les constantes est  $\mathbb{P}$ -intégrable. Mais alors par convergence dominée,  $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ ,  $n \rightarrow +\infty$ , ce qui justifie la convergence en loi par le Théorème 7.4.2.  $\square$

**Proposition 7.4.3 (Convergences en proba et en loi)** *La convergence en probabilité implique la convergence en loi.*

**Démonstration :** Il s'agit de montrer que si  $(X_n)_{n \geq 0}$  converge en probabilité vers  $X$ , la suite  $F_n(x) = \mathbb{P}(X_n \leq x)$  converge vers  $F_X(x) = \mathbb{P}(X \leq x)$  en tout point  $x$  où  $F_X$  est continue.

Soit  $\varepsilon > 0$ ,  $X_n = X + (X_n - X)$  est inférieure à  $x$  si soit  $X \leq x + \varepsilon$  soit  $X_n - X \leq -\varepsilon$ . On a donc  $\{X_n \leq x\} \subset \{X \leq x + \varepsilon\} \cup \{X_n - X \leq -\varepsilon\}$  puis

$$\begin{aligned}\mathbb{P}(X_n \leq x) &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(X_n - X \leq -\varepsilon) \\ &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| \geq \varepsilon).\end{aligned}$$

Soit  $\eta > 0$ , par continuité de  $F_X$  en  $x$ , il existe  $\varepsilon_0$  tel que  $F_X(x) \leq F_X(x + \varepsilon_0) \leq F_X(x) + \eta/2$ . Pour ce choix de  $\varepsilon_0$ , d'après la convergence en probabilité, il existe  $n_0$  tel que pour tout  $n \geq n_0$ ,  $\mathbb{P}(|X - X_n| \geq \varepsilon_0) \leq \eta/2$ . Finalement, pour tout  $n \geq n_0$ ,

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) \leq F_X(x + \varepsilon_0) + \mathbb{P}(|X_n - X| \geq \varepsilon_0) \leq F_X(x) + \eta/2 + \eta/2 \leq F_X(x) + \eta.$$

Un raisonnement analogue partant de l'inégalité

$$\mathbb{P}(X \leq x - \varepsilon) \leq \mathbb{P}(X_n \leq x) + \mathbb{P}(X - X_n \leq -\varepsilon)$$

réécrite sous la forme

$$\begin{aligned}\mathbb{P}(X_n \leq x) &\geq \mathbb{P}(X \leq x - \varepsilon) - \mathbb{P}(X - X_n \leq -\varepsilon) \\ &\geq \mathbb{P}(X \leq x - \varepsilon) - \mathbb{P}(|X - X_n| \geq \varepsilon)\end{aligned}$$

puis pour  $n$  assez grand :

$$F_{X_n}(x) \geq F_X(x - \varepsilon) - \mathbb{P}(|X - X_n| \geq \varepsilon) \geq F_X(x) - \frac{\eta}{2} - \frac{\eta}{2}$$

( $\eta$  étant fixé, on choisit  $\varepsilon$  assez petit tel que  $F_X(x - \varepsilon) > F_X(x) - \eta/2$  par continuité de  $F_X$  en  $x$ ; puis pour cet  $\varepsilon > 0$ , la convergence en probabilité rend  $\mathbb{P}(|X - X_n| \geq \varepsilon) \leq \eta/2$  pour  $n$  assez grand) permet de montrer la seconde inégalité, ie.  $F_{X_n}(x) \geq F_X(x) - \eta$  pour  $n$  assez grand.

Finalement, on a établit la convergence de  $F_{X_n}(x)$  vers  $F_X(x)$ , en  $x$  point de continuité de  $F_X$ .  $\square$

**Remarque 7.4.4** La réciproque est fausse en général : la convergence en loi n'entraîne pas la convergence en probabilité. Pour le voir, considérons par exemple  $(X_n)_{n \in \mathbb{N} \cup \{+\infty\}}$  une suite de variable iid de loi  $b(p)$ . Comme la suite est de même loi, on a immédiatement  $X_n \Rightarrow X$ . Puis

$$X_n - X = \begin{cases} 1 - 1 = 0 & \text{avec probabilité } 1/4 \\ 1 - 0 = 1 & \text{avec probabilité } 1/4 \\ 0 - 1 = -1 & \text{avec probabilité } 1/4 \\ 0 - 0 = 0 & \text{avec probabilité } 1/4. \end{cases}$$

Pour  $\varepsilon \in ]0, 1[$ , on a alors

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(X_n - X = \pm 1) = 1/2 \not\rightarrow 0.$$

Cependant, dans un cas bien particulier, on a une réciproque partielle :

**Proposition 7.4.4 (Convergence en proba et en loi 2)** *Si  $X_n$  converge en loi vers la constante  $c$  alors  $X_n$  converge en probabilité vers  $c$ .*

**Démonstration :** En effet, soit pour  $\varepsilon > 0$ ,  $h = \mathbf{1}_{]-\infty, c-\varepsilon] \cup [c+\varepsilon, +\infty[}$  et  $g$  la fonction qui vaut 1 si  $x \leq c - \varepsilon$  ou  $x \geq c + \varepsilon$ , 0 en  $c$  et affine entre  $c - \varepsilon$  et  $c$  puis entre  $c$  et  $c + \varepsilon$ . La fonction  $g$  est continue, bornée et  $h \leq g$ . On a

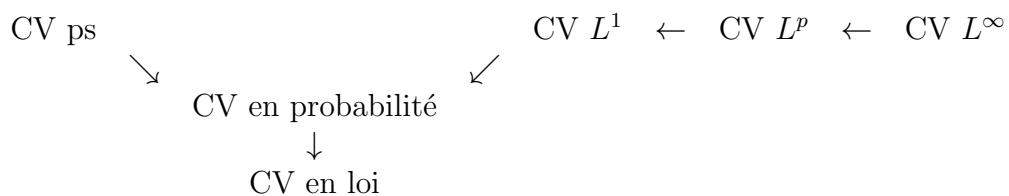
$$\mathbb{P}(|X_n - c| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}}] = \int h(X_n) d\mathbb{P} \leq \int g(X_n) d\mathbb{P} \rightarrow \int g(c) d\mathbb{P} = g(c) = 0$$

en appliquant le Théorème 7.4.1. On a donc  $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - c| \geq \varepsilon) = 0$ .  $\square$

La convergence  $L^1$  (ou dans  $L^p$ ) implique la convergence en loi (puisque elle implique la convergence en probabilité qui implique celle en loi). La réciproque est fausse comme on le voit avec le même (contre) exemple qu'après la Proposition 7.3.2.

## 7.4.2 Liens entre les différents modes de convergence

Les liens entre les différentes convergences sont donnés par le diagramme suivant :



avec en plus les réciproques partielles :

- CV proba  $\Rightarrow$  CV ps pour une sous-suite ;
- CV proba  $\Rightarrow$  CV  $L^1$  sous UI ;
- CV en loi  $\Rightarrow$  CV proba si  $X_\infty = c$ .

# Chapitre 8

## Théorèmes limite

Dans ce chapitre, on considère des variables aléatoires réelles et on s'intéresse au comportement asymptotique de  $S_n = \sum_{i=1}^n X_i$  où  $(X_i)_{i \geq 1}$  est une suite de variables aléatoires *iid*. Les principaux résultats sont la loi des grands nombres (LGN) qui donne le comportement de  $S_n/n$  et le théorème central limite (TCL) qui précise ce résultat en montrant que la vitesse de convergence de  $S_n/n$  vers  $\mathbb{E}[X_1]$  est en  $\sqrt{n}$ .

### 8.1 Lois des grands nombres (LGN)

La loi des grands nombres est la formulation rigoureuse des faits intuitifs suivants : si on lance un grand nombre de fois une pièce en l'air, il y aura en moyenne 50% de piles. De même, si on lance un grand nombre de fois un dé à 6 faces en l'air, il y aura en moyenne 1/6-ème des faces qui seront, par exemple, des 4 (si la pièce et le dé sont équilibrés).

#### 8.1.1 Version faible de la LGN

Il existe une première version de la LGN (dite faible) pour la convergence en probabilité. Sous des hypothèses  $L^2$ , c'est une conséquence simple de l'inégalité de Tchebychev (Théorème 8.1.1) ; sous des hypothèses  $L^1$ , c'est une conséquence du Théorème de Paul Lévy et de la Proposition 7.4.4 en calculant les fonctions caractéristiques (Théorème 8.1.2) combiné avec la Proposition 7.4.4.

#### LGN faible $L^2$

**Théorème 8.1.1 (LGN faible  $L^2$ )** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes et de même loi avec un moment d'ordre 2. Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_1], \quad n \rightarrow +\infty.$$

On a donc la convergence de la moyenne arithmétique (dite aussi moyenne empirique) vers la moyenne probabiliste (l'espérance probabiliste).

**Démonstration :** La variable aléatoire limite est la constante  $\mathbb{E}[X_1](=\mathbb{E}[X_i]$  pour tout  $i$  car les variables aléatoires  $X_i$  ont même loi, donc même espérance). Il s'agit de vérifier

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1] \right| \geq \varepsilon \right) = 0.$$

Posons  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ , par linéarité, on a  $\mathbb{E}[M_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_1]$ . D'autre part, par indépendance des  $X_i$ , on a

$$\text{Var}(M_n) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \text{Var} \left( \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \times (n \text{Var}(X_1)) = \frac{\text{Var}(X_1)}{n}.$$

L'inégalité de Tchebychev appliquée à  $M_n$  donne alors pour tout  $\varepsilon > 0$  et  $n \in \mathbb{N}^*$  :

$$\mathbb{P} \left( \left| \frac{1}{n} \left( \sum_{i=1}^n X_i \right) - \mathbb{E}[X_1] \right| \geq \varepsilon \right) = \mathbb{P}(|M_n - \mathbb{E}[M_n]| \geq \varepsilon) \leq \frac{\text{Var}(M_n)}{\varepsilon^2} \leq \frac{\text{Var}(X_1)}{n \varepsilon^2} \quad (8.1)$$

On conclut en faisant tendre  $n$  vers  $+\infty$ . □

**Remarque 8.1.1** Plus que la convergence, nous avons obtenu la *vitesse* de convergence : d'après (8.1) elle est en  $O(1/n)$ . Si on connaît  $\text{Var}(X_1)$ , on peut donc pour une proportion donnée, trouver un rang  $n_0$  tel que pour  $n \geq n_0$  et pour cette proportion de  $\omega \in \Omega$ , on a la moyenne arithmétique  $\varepsilon$ -proche de l'espérance. Ceci est utile pour déterminer des intervalles de confiance.

Souvent, on se trouve dans le cas particulier où les variables aléatoires considérées sont de loi de Bernoulli, la LGN faible se réécrit alors :

**Corollaire 8.1.1** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes de Bernoulli de même paramètre  $p$ . Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p, \quad n \rightarrow +\infty.$$

**Démonstration :** On a  $\mathbb{E}[X_i] = p$  car  $X_i \sim b(p)$  et comme  $\text{Var}(X_i) = p(1-p) < +\infty$ , la LGN faible version  $L^2$  s'applique. □

C'est ce résultat qui formalise le résultat intuitif sur le lancer des dés ou des pièces : avec  $X_i = \mathbf{1}_{\{\text{obtenir un } 4 \text{ au i-ème lancer}\}}$ , on a  $X_i \sim b(1/6)$  et  $p = 1/6$  et  $\frac{1}{n} \sum_{i=1}^n X_i$  désigne la fréquence d'apparition du 4 sur les  $n$  premiers lancers.

## Application de la LGN $L^2$ : estimation d'une proportion inconnue

On se propose d'estimer le paramètre  $p$  inconnu d'une loi de Bernoulli en observant un grand nombre de fois un phénomène aléatoire de loi de Bernoulli  $b(p)$ , c'est à dire en observant les réalisations d'une suite de variables aléatoires indépendantes  $X_i(\omega)$ ,  $1 \leq i \leq n$ , de loi de Bernoulli  $b(p)$ .

Modélisons le problème de la façon suivante. On considère une urne comportant des boules rouges en proportion inconnue  $p$  et des boules vertes (en proportion  $1 - p$ ). On effectue  $n$  tirages d'une boule avec remise. Notons

$$X_i = \begin{cases} 0 & \text{si la boule tirée est rouge} \\ 1 & \text{si la boule tirée est verte} \end{cases} = \mathbf{1}_{\{\text{la boule est rouge au } i\text{-ème tirage}\}}.$$

Désignons toujours par  $M_n$  la moyenne arithmétique des  $n$  premières variables aléatoires  $X_i$ , ici cela correspond à la fréquence d'apparition des boules rouges lors des  $n$  premiers tirages. D'après la loi faible des grands nombres (ou plutôt son corollaire 8.1.1 pour les fréquences),  $M_n$  converge en probabilité vers  $p$  :

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p.$$

Il est donc légitime d'estimer  $p$  par  $M_n$  pour  $n$  assez grand. En pratique, on observe *une* valeur particulière  $M_n(\omega)$  calculée à partir des  $n$  tirages réellement effectués correspondant à la réalisation  $\omega$  du hasard. La question pratique qui se pose est de donner une fourchette pour l'approximation de  $p$  par la valeur observée  $M_n(\omega)$  et contrôler le risque que cette approximation ne soit pas valable. L'inégalité de Tchebychev pour  $M_n$  s'écrit ici :

$$\mathbb{P}(|M_n - p| \geq t) \leq \frac{\text{Var}(X_1)}{nt^2} = \frac{p(1-p)}{nt^2} \leq \frac{1}{4nt^2}$$

car on majore facilement  $p(1-p)$  par  $1/4$  (la fonction  $x \mapsto x(1-x)$  sur  $[0, 1]$  a son maximum en  $1/2$  et il vaut  $1/4$ ). D'où

$$\mathbb{P}(M_n - t < p < M_n + t) = 1 - \mathbb{P}(|M_n - p| \geq t) \geq 1 - \frac{1}{4nt^2}. \quad (8.2)$$

En pratique, on observe  $M_n(\omega)$  et on dit que  $I_n(\omega) = ]M_n(\omega) - t, M_n(\omega) + t[$  est un intervalle de confiance (ou fourchette). L'équation (8.2) permet de voir que la probabilité de se tromper (ie. en fait  $p \notin I_n(\omega)$ ) est majorée par  $1/(4nt^2)$ .

Si on se fixe un seuil d'erreur  $\alpha$ , on trouve  $t_\alpha$  tel que  $1/(4nt_\alpha^2) = \alpha$  et l'intervalle  $I_{n,\alpha}(\omega) = ]M_n(\omega) - t_\alpha, M_n(\omega) + t_\alpha[$  est alors l'intervalle de confiance au niveau  $\alpha$ .

**Exemple (Sondage) :** Avant un référendum, un institut de sondage interroge au hasard 1000 personnes dans la rue. On note  $p$  la proportion d'électeurs décidés à voter OUI dans la population totale. Dans l'échantillon sondé, cette proportion est égale à 0,54. Proposer un intervalle de confiance pour  $p$  au niveau 0,95.

Le sondage peut être assimilé à un tirage avec remise (la réponse d'un électeur de l'échantillon correspondant au tirage d'une boule d'une certaine couleur selon son choix de vote), on est alors ramené à la situation de l'exemple précédent. Ici, la fréquence observée du choix du OUI sur les 1000 électeurs est  $M_{1000}(\omega) = 0,54$  et l'intervalle de confiance est

$$I = ]0,54 - t; 0,54 + t[ \quad \text{avec un niveau } \alpha \geq 1 - 1/(4 \times 1000 \times t^2).$$

Pour un seuil de confiance  $\alpha \geq 0,95$ , il faut alors

$$1 - \frac{1}{4000 \times t^2} \geq 0,95 \iff \frac{1}{4000 \times t^2} \leq 0,05 \iff t \geq \frac{1}{10\sqrt{2}} \simeq 0,0707.$$

Avec  $t_\alpha = 0,071$ , on trouve  $I_\alpha = ]0,469; 0,611[$ . On constate en particulier qu'une zone de l'intervalle de confiance correspond à  $p < 1/2$ , pour lequel le OUI ne serait pas gagnant alors que la proportion observée semblait lui garantir la victoire. On ne peut pas garantir la victoire du OUI avec une probabilité d'erreur inférieure à 5%.

Combien de personne faut-il alors interroger pour donner une fourchette à  $\pm 1\%$  avec un seuil de 95% ?

Repartons de (8.2), avec une fourchette  $t = 0,01$ . On veut  $\alpha \geq 0,95$  soit  $1 - \alpha \leq 0,05$  :

$$\frac{1}{4n \times 0,01^2} \leq 0,05.$$

On trouve  $n = 50000$ , ce qui donne au sondage un coût prohibitif. En gros, pour améliorer la précision d'un facteur 10, on observe qu'il faut interroger 100 fois plus de personnes et donc multiplier les coûts par 100.

## LGN faible $L^1$

On obtient toujours une LGN faible pour des variables aléatoires *iid*  $L^1$  en utilisant le théorème de Paul Lévy (Théorème 7.4.3) et la Proposition 7.4.4 en calculant les fonctions caractéristiques (Théorème 8.1.1). L'hypothèse  $L^1$  est optimale puisqu'il faut que  $X_1$  soit intégrable pour que son espérance ait bien un sens. Pour cela, on utilise les résultats préliminaires suivants :

**Lemme 8.1.1** *On a*

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \min \left( \frac{|x|^{p+1}}{(p+1)!}, \frac{2|x|^p}{p!} \right).$$

Par conséquent,

$$\left| \varphi_X(t) - \sum_{k=0}^p \frac{i^k t^k}{k!} \mathbb{E}[X^k] \right| \leq \mathbb{E} \left[ \min \left( \frac{t^{p+1}|X|^{p+1}}{(p+1)!}, \frac{2t^p|X|^p}{p!} \right) \right].$$

En effet, on a

134

$$\begin{aligned}
 \left| \varphi_X(t) - \sum_{k=0}^p \frac{i^k t^k}{k!} \mathbb{E}[X^k] \right| &= \left| \mathbb{E}[\exp(itX)] - \sum_{k=0}^p \frac{i^k t^k}{k!} \mathbb{E}[X^k] \right| \\
 &\leq \mathbb{E} \left[ \left| \exp(itX) - \sum_{k=0}^p \frac{(itX)^k}{k!} \right| \right] \\
 &\leq \mathbb{E} \left[ \min \left( \frac{t^{p+1}|X|^{p+1}}{(p+1)!}, \frac{2t^p|X|^p}{p!} \right) \right].
 \end{aligned}$$

**Démonstration :** [lemme] On utilise la formule de Taylor avec reste intégral :

$$e^{ix} = \sum_{k=0}^p \frac{(ix)^k}{k!} + \frac{i^{p+1}}{p!} \int_0^x (x-s)^p e^{is} ds$$

dont on déduit

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \frac{1}{p!} \left| \int_0^x (x-s)^p ds \right| \leq \frac{|x|^{p+1}}{(p+1)!}.$$

Puis, par un calcul simple à partir d'intégrations par parties

$$e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} = \frac{i^p}{(p-1)!} \int_0^x (x-s)^{p-1} e^{is} ds - \frac{(ix)^p}{p!}$$

d'où

$$\left| e^{ix} - \sum_{k=0}^p \frac{(ix)^k}{k!} \right| \leq \frac{|x|^p}{p!} + \frac{|x|^p}{p!} = 2 \frac{|x|^p}{p!}$$

ce qui donne la conclusion du lemme. □

**Lemme 8.1.2** Soient  $a_i, b_i \in \mathbb{C}$ ,  $1 \leq i \leq n$ , avec  $|a_i|, |b_i| \leq 1$ , alors

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i - b_i|.$$

**Démonstration :** On procède par récurrence sur  $n$ . Le cas  $n = 1$  est évident. On suppose le résultat vrai pour  $n$ . On écrit alors

$$\begin{aligned}
 \left| \prod_{i=1}^{n+1} a_i - \prod_{i=1}^{n+1} b_i \right| &\leq \left| \prod_{i=1}^{n+1} a_i - a_{n+1} \prod_{i=1}^n b_i \right| + \left| a_{n+1} \prod_{i=1}^n b_i - \prod_{i=1}^{n+1} b_i \right| \\
 &\leq |a_{n+1}| \left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| + \left| \prod_{i=1}^n b_i \right| |a_{n+1} - b_{n+1}|
 \end{aligned}$$

$$\begin{aligned} &\leq |a_{n+1}| \sum_{i=1}^n |a_i - b_i| + \left| \prod_{i=1}^n b_i \right| |a_{n+1} - b_{n+1}| \\ &\leq \sum_{i=1}^{n+1} |a_i - b_i|, \end{aligned}$$

ce qui donne le résultat pour  $n + 1$  et achève la récurrence.  $\square$

**Théorème 8.1.2 (LGN faible  $L^1$ )** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires iid avec un moment d'ordre 1 ( $\mathbb{E}[|X_1|] < +\infty$ ) et soit  $S_n = X_1 + \dots + X_n$ . Alors

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1], \quad n \rightarrow +\infty.$$

**Remarque 8.1.2** Si les  $X_i$  ne sont pas  $L^1$ , alors il n'y a pas convergence en proba de  $S_n/n$ . En effet, par exemple si les  $X_i$  ont iid de loi de Cauchy  $\mathcal{C}(1)$  (on a bien  $X_i \notin L^1$  alors puisque  $\varphi_X(t) = e^{-|t|}$ , on a

$$\varphi_{S_n/n}(t) = \varphi_{S_n}(t/n) = \varphi_X(t/n)^n = e^{-|t|}$$

c'est à dire  $S_n/n \sim \mathcal{C}(1)$  et la suite  $S_n/n$  est stationnaire en loi et ne converge donc pas en proba.

**Démonstration :** Quitte à écrire  $X_i = m + X'_i$ , puis  $S_n/n = m + S'_n/n$  avec  $S'_n = \sum_{k=1}^n X'_i$ , on se ramène sans perte de généralité à  $m = 0$  et à montrer  $\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0$ . Pour cela, on se contente de montrer une convergence en loi (Proposition 7.4.4) et on l'établit à l'aide des fonctions caractéristiques (théorème de Paul Lévy, Th. 7.4.4).

Comme  $X \in L^1$ , la fonction caractéristique  $\varphi_X$  de  $X$  est dérivable et on a  $\varphi_X(0) = 1$ ,  $\varphi'_X(0) = i\mathbb{E}[X]$ . On a donc le développement limité suivant pour  $t$  proche de 0

$$\varphi_X(t) = 1 + it\mathbb{E}[X] + o(t).$$

Comme les variables aléatoires  $X_n$ ,  $n \geq 1$ , sont iid, la fonction caractéristique de  $S_n$  est

$$\varphi_{S_n}(t) = \varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \dots \varphi_{X_n}(t) = \varphi_X(t)^n$$

et donc

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \varphi_X\left(\frac{t}{n}\right)^n = \left(1 + i\frac{t}{n}\mathbb{E}[X] + o(t/n)\right)^n.$$

Il faut contrôler correctement le reste  $o(t/n)$ . Pour cela, on applique le Lemme 8.1.2 avec  $a_i = \varphi_X(t)$  et  $b_i = 1$ ,  $1 \leq i \leq n$  (on a bien  $|a_i| \leq 1$ ). On a alors

$$\left| \varphi_X\left(\frac{t}{n}\right)^n - 1^n \right| \leq n \left| \varphi_X\left(\frac{t}{n}\right) - 1 \right|.$$

Pour montrer que ce majorant tend vers 0, on estime  $\left| \varphi_X\left(\frac{t}{n}\right) - 1 \right|$  avec le Lemme 8.1.1 avec  $p = 1$  (et  $\mathbb{E}[X_1] = 0$ ) :

$$\begin{aligned} n \left| \varphi_X\left(\frac{t}{n}\right) - 1 \right| &= n \left| \mathbb{E}[e^{it/n} - 1] \right| = n \left| \mathbb{E}[e^{it/n} - 1 - i\frac{t}{n}X] \right| \\ &\leq n \mathbb{E} \left[ \left| e^{it/n} - 1 - i\frac{t}{n}X \right| \right] \leq \mathbb{E} \left[ n \min \left( \frac{t^2 X^2}{2n^2}, \frac{2t|X|}{n} \right) \right]. \end{aligned}$$

On observe alors que

$$n \min \left( \frac{t^2 X^2}{2n^2}, \frac{2t|X|}{n} \right) \leq 2t|X| \in L^1 \quad \text{et} \quad n \min \left( \frac{t^2 X^2}{2n^2}, \frac{2t|X|}{n} \right) \leq \frac{t^2 X^2}{n} \rightarrow 0, \quad n \rightarrow +\infty.$$

Par le théorème de convergence dominée, il vient

$$\lim_{n \rightarrow +\infty} n \left| \varphi_X\left(\frac{t}{n}\right) - 1 \right| = 0 \tag{8.3}$$

c'est à dire

$$\lim_{n \rightarrow +\infty} n \varphi_{S_n/n}(t) \rightarrow 1 = \varphi_{\mathbb{E}[X]}(t).$$

On a établi  $S_n/n \Rightarrow \mathbb{E}[X] = 0$ . Comme la limite est constante, la convergence en loi se renforce en une convergence en probabilité comme souhaité, cf. Proposition 7.4.4.  $\square$

### 8.1.2 Version forte (presque sûre) de la LGN

La loi des grands nombres s'énonce aussi pour la convergence presque sûre (ou en norme  $L^p$ ). On parle alors de LGN forte. Comme la convergence ps (ou en norme  $L^p$ ) implique la convergence en probabilité, la LGN forte implique la LGN faible déjà prouvée. On propose dans cette section plusieurs preuves de la LGN forte pour des variables *iid* avec des conditions d'intégrabilité  $L^4$ ,  $L^2$ ,  $L^1$ . Bien sûr, plus les hypothèses d'intégrabilité sont faibles, meilleur est le résultat mais souvent plus difficiles sont les preuves.

**Théorème 8.1.3 (LGN forte ps)** *Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires iid. Il y a équivalence entre*

1.  $X_1 \in L^1$  ie.  $\mathbb{E}[|X_1|] < +\infty$ .
2. (LGNF ps)

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{ps}} c \in \mathbb{R}, \quad n \rightarrow +\infty. \tag{8.4}$$

Dans ce cas,  $c = \mathbb{E}[X_1]$ .

**Remarque 8.1.3** la LGN ps reste vraie sous l'hypothèse d'indépendance par paires, cf. la deuxième preuve ci-dessous.

La LGN s'énonce encore de la façon suivante : si  $\mathbb{E}[|X_1|] < +\infty$  alors  $\frac{S_n}{n} = \mathbb{E}[X_1] + o(1)$  ps lorsque  $n \rightarrow +\infty$ . Le TCL donne en un certain sens un terme de plus dans le développement asymptotique en précisant le comportement en loi du terme  $o(1)$  sous une hypothèse supplémentaire sur la loi des  $X_i$ . On a ainsi une estimation très précise de l'erreur commise en approchant l'espérance mathématique par la moyenne empirique. Le résultat permet en plus d'approximer la loi de  $\frac{S_n}{n}$  lorsque  $n$  est grand.

**Démonstration du sens indirect 2)  $\Rightarrow 1)$ .**

Si la suite  $\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$  converge presque sûrement vers  $c$ , alors

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{S_{n-1}}{n-1} \times \frac{n-1}{n} \rightarrow c - c \times 1 = 0$$

donc  $X_n/n$  converge presque sûrement vers 0. Comme les variables aléatoires  $(X_n)_{n \geq 1}$  sont iid, par le lemme de Borel-Cantelli, pour tout (ou pour un)  $\varepsilon > 0$  on a nécessairement :

$$\sum_{n \geq 1} \mathbb{P}\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) < +\infty$$

sinon on aurait ps  $|X_n/n| > \varepsilon$  infiniment souvent. Mais

$$\sum_{n \geq 1} \mathbb{P}\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) = \sum_{n \geq 1} \mathbb{P}(|X_n| \geq \varepsilon n) = \sum_{n \geq 1} \mathbb{P}(|X_1| \geq \varepsilon n).$$

La Proposition 3.2.3 garantit alors que  $\mathbb{E}[|X_1|] < +\infty$  puisque

$$\sum_{n \geq 0} \mathbb{P}(|X| > n) \leq \mathbb{E}[|X|] \leq \sum_{n \geq 1} \mathbb{P}(|X| > n).$$

□

### Preuve de la LGN $L^4$

On prouve la LGN, c'est à dire le sens  $1) \Rightarrow 2)$  du Théorème 8.1.3 en commençant par le cas  $L^4$  (existence d'un moment d'ordre 4 :  $\mathbb{E}[X_1^4] < +\infty$ ). Dans ce cas, il suffit d'appliquer le lemme de Borel-Cantelli en vérifiant l'hypothèse de la convergence d'une série de probabilité en les majorant par l'inégalité de Tchebychev. L'existence de moment d'ordre 4 permet d'avoir de bonnes majorations de ces probabilités, cf. (8.5).

**Démonstration de 1)  $\Rightarrow$  2) dans le cas  $L^4$ .** Il suffit de prouver le théorème quand  $\mathbb{E}[X_1] = 0$ , le cas général s'obtenant par soustraction de l'espérance  $\mathbb{E}[X_1]$ . Posons

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n = \sum_{i=1}^n X_i.$$

Soit  $\varepsilon > 0$  et  $D_\varepsilon = \limsup_{n \rightarrow +\infty} \{|M_n| \geq \varepsilon\}$ . On utilise le lemme de Borel-Cantelli pour montrer que  $\mathbb{P}(D_\varepsilon) = 0$ . Il suivra que l'ensemble  $D = \bigcup_{n \geq 1} D_{1/n}$  est un événement de probabilité nulle, ce qui prouvera le résultat car  $D = \{M_n \not\rightarrow 0\}$ , cf. le détail des explications ci-dessous.

Afin d'utiliser le lemme de Borel-Cantelli, on montre la convergence de la série de terme général  $\mathbb{P}(|M_n| \geq \varepsilon)$ . Or

$$\mathbb{P}(|M_n| \geq \varepsilon) = \mathbb{P}(|S_n| \geq n\varepsilon) = \mathbb{P}(S_n^4 \geq n^4 \varepsilon^4) \leq \frac{\mathbb{E}[S_n^4]}{n^4 \varepsilon^4}.$$

en utilisant l'inégalité de Markov. Il reste à estimer  $\mathbb{E}[S_n^4]$ .

$$\begin{aligned} S_n^4 &= (X_1 + X_2 + \cdots + X_n)^4 \\ &= \sum_{k_1, k_2, k_3, k_4 \in \{1, \dots, n\}^4} X_{k_1} X_{k_2} X_{k_3} X_{k_4} \\ &= \left\{ \begin{array}{ll} M(4) \sum_{i=1}^n X_i^4 & + M(1, 3) \sum_{1 \leq i < j \leq n} X_i^3 X_j \\ + M(2, 2) \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 & + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} X_i^2 X_j X_k \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} X_i X_j X_k X_l \end{array} \right. \end{aligned}$$

où  $M(4) = \binom{4}{4} = 1$ ,  $M(1, 3) = \binom{4}{3} = 4$ ,  $M(2, 2) = \binom{4}{2} = 6$ ,  $M(2, 1, 1) = \binom{4}{2} \binom{2}{1} = 12$ ,  $M(1, 1, 1, 1) = \binom{4}{1} \binom{3}{1} \binom{2}{1} = 24$ . La linéarité et l'indépendance des  $X_i$  donnent alors  $\mathbb{E}[S_n^4] =$

$$\left\{ \begin{array}{ll} M(4) \sum_{i=1}^n \mathbb{E}[X_i^4] & + M(1, 3) \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^3] \mathbb{E}[X_j] \\ + M(2, 2) \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] & + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j] \mathbb{E}[X_k] \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j] \mathbb{E}[X_k] \mathbb{E}[X_l] \end{array} \right.$$

Comme  $\mathbb{E}[X_i] = 0$ , les deuxième, quatrième et cinquième termes sont nuls. Comme on montre que  $M(4) = 1$ ,  $M(2, 2) = 6$ , on obtient

$$\mathbb{E}[S_n^4] = \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2]$$

$$\begin{aligned}
&= n\mathbb{E}[X_1^4] + 6\binom{n}{2}(\mathbb{E}[X_1^2])^2 \\
&= n\mathbb{E}[X_1^4] + 3n(n-1)(\mathbb{E}[X_1^2])^2 \\
&\leq n\mathbb{E}[X^4] + 3n(n-1)\mathbb{E}[X^4] \\
&\leq 3n^2\mathbb{E}[X^4] < +\infty
\end{aligned}$$

car  $\mathbb{E}[X^2]^2 \leq \mathbb{E}[X^4]$  par l'inégalité de Cauchy-Schwarz. On a alors

$$\mathbb{P}(|M_n| \geq \varepsilon) \leq \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4} \leq \frac{3\mathbb{E}[X^4]}{n^2\varepsilon^4}. \quad (8.5)$$

Comme  $3\mathbb{E}[X^4]/n^2\varepsilon^4$  est le terme général d'une série convergente,  $\mathbb{P}(|M_n| \geq \varepsilon)$  aussi. Le lemme de Borel-Cantelli s'applique et donne  $\mathbb{P}(D_\varepsilon) = 0$ . Posons alors

$$D = \bigcup_{p=1}^{+\infty} D_{1/p}.$$

On a  $\mathbb{P}(D) = 0$  car  $D$  est réunion dénombrable d'ensembles  $D_{1/p}$  de probabilités nulles. Prenons alors

$$\Omega_0 := D^c = \bigcap_{p \geq 1} D_{1/p}^c = \bigcap_{p \geq 1} \bigcup_{k \geq 1} \bigcap_{n \geq k} \{|M_n| \leq 1/p\}.$$

On a  $\mathbb{P}(\Omega_0) = 1$  et pour tout  $\omega \in \Omega_0$ , par traduction dans le langage logique des symboles ensemblistes, pour tout  $p \in \mathbb{N}^*$ , il existe un entier  $k$  tel que pour tout  $n \geq k$  :  $|M_n| \leq 1/p$ . On a donc  $M_n$  qui converge presque sûrement vers 0 ; ce qui achève la preuve de la LGN forte.  $\square$

### Preuve de la LGN $L^2$

On prouve maintenant la LGN  $L^2$  qui suppose l'existence d'un moment d'ordre 2 :  $\mathbb{E}[X_1^2] < +\infty$ . Elle utilise sur le lemme de Kronecker (Lemme 8.1.4) qui repose lui-même sur celui de Césaro (Lemme 8.1.3) qu'on commence par rappeler.

**Lemme 8.1.3 (Césaro)** *Soit  $(u_n)_{n \geq 1}$  une suite qui converge vers  $\ell$  et  $(\alpha_n)_{n \geq 1}$  une suite positive, de sommes partielles  $a_n = \sum_{k=1}^n \alpha_k$  qui tendent vers  $+\infty$ . Alors on a*

$$\lim_{n \rightarrow +\infty} \frac{1}{a_n} \sum_{k=1}^n \alpha_k u_k = \ell.$$

**Démonstration :** Soit  $\varepsilon > 0$  et  $n_0 \geq 1$  tel que, pour  $n \geq n_0$ , on a  $|u_n - \ell| < \varepsilon$ . Comme  $\sum_{k=1}^n \alpha_k = a_n$ , on a

$$\left| \frac{1}{a_n} \left( \sum_{k=1}^n \alpha_k u_k \right) - \ell \right| = \left| \frac{1}{a_n} \sum_{k=1}^n \alpha_k (u_k - \ell) \right|$$

$$\begin{aligned} &\leq \frac{1}{a_n} \sum_{k=1}^{n_0} \alpha_n |u_k - \ell| + \frac{1}{a_n} \sum_{k=n_0+1}^n \alpha_n |u_k - \ell| \\ &\leq \frac{1}{a_n} \sum_{k=1}^{n_0} \alpha_n |u_k - \ell| + \frac{a_n - a_{n_0}}{a_n} \varepsilon. \end{aligned}$$

Comme  $\frac{a_n - a_{n_0}}{a_n} \leq 1$  et  $\lim_{n \rightarrow +\infty} \frac{1}{a_n} \sum_{k=1}^{n_0} \alpha_n |u_k - \ell| = 0$  ( $a_n \rightarrow +\infty$ ), on a

$$\limsup_{n \rightarrow +\infty} \left| \frac{1}{a_n} \sum_{k=1}^n \alpha_n u_k - \ell \right| \leq \varepsilon,$$

ce qui permet de conclure puisque  $\varepsilon > 0$  est arbitraire.

□

**Lemme 8.1.4 (Kronecker)** Soit  $\sum_{n \geq 1} x_n$  une série convergente dans un evn  $E$  ( $x_n \in E$ ). Soient  $a_n > 0$  une suite de réels  $a_n \rightarrow +\infty$  en croissant. Alors

$$\lim_{n \rightarrow +\infty} \frac{1}{a_n} \sum_{k=1}^n a_k x_k = 0.$$

**Démonstration :** Il s'agit de faire une transformation d'Abel en écrivant  $x_k = S_k - S_{k-1}$  :

$$\sum_{k=1}^n a_k x_k = \sum_{k=1}^n a_k (S_k - S_{k-1}) = \sum_{k=1}^{n-1} (a_k - a_{k-1}) S_k + a_n S_n$$

avec  $S_0 = 0$  par convention et donc

$$\frac{1}{a_n} \sum_{k=1}^n a_k x_k = S_n - \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} S_k. \quad (8.6)$$

Comme  $S_k$  converge vers  $S$ , le Lemme 8.1.3 (Césaro) avec  $\alpha_k = a_k - a_{k-1} \geq 0$  de somme partielle  $a_n \rightarrow +\infty$  assure

$$\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} S_k \rightarrow S, \quad n \rightarrow +\infty,$$

ce qui conclut le Lemme 8.1.4 (Kronecker) quand on passe à la limite dans (8.6). □

**Théorème 8.1.4** Soient  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes centrées  $L^2$  et  $a_n > 0$  des réels qui convergent en croissant vers  $+\infty$ . Alors si  $\sum_{n=1}^{+\infty} \frac{\mathbb{E}[X_n^2]}{a_n^2} < +\infty$ , on a  $\frac{S_n}{a_n} \xrightarrow{ps,L^2} 0$ .

**Corollaire 8.1.2** Si les variables aléatoires  $X_n$ ,  $n \geq 1$ , indépendantes sont non centrées, alors  $\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n^2)}{a_n^2} < +\infty$  implique

$$\frac{S_n - \mathbb{E}[S_n]}{a_n} \xrightarrow{ps,L^2} 0.$$

**Démonstration :** [Corollaire 8.1.2] Appliquer le Théorème 8.1.4 aux variables aléatoires  $Y_n = X_n - \mathbb{E}[X_n]$ ,  $n \geq 1$ , qui sont centrées et indépendantes.  $\square$

**Corollaire 8.1.3 (LGN L<sup>2</sup>)** Soit  $X_n$ ,  $n \geq 1$ , iid avec un moment d'ordre 2 fini et de moyenne  $m$ . Alors

$$\frac{S_n}{n} \xrightarrow{ps,L^2} m.$$

**Démonstration :** [Corollaire 8.1.3] Le Corollaire 8.1.2 s'applique avec  $a_n = n$  puisque

$$\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n^2)}{a_n^2} = \text{Var}(X_1) \sum_{n \geq 1} \frac{1}{n^2} < +\infty$$

et donne la conclusion cherchée (LGN L<sup>2</sup>).  $\square$

**Démonstration :** [Th. 8.1.4] On montre d'abord la convergence L<sup>2</sup> : Comme les variables aléatoires  $X_n$ ,  $n \geq 1$ , sont indépendantes et centrées, on a

$$\begin{aligned} \frac{\mathbb{E}[S_n^2]}{a_n^2} &= \frac{1}{a_n^2} \mathbb{E}\left[ \sum_{k=1}^n X_k^2 + \sum_{k \neq l} X_k X_l \right] \\ &= \frac{1}{a_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2] + \sum_{k \neq l} \mathbb{E}[X_k] \mathbb{E}[X_l] \\ &= \frac{1}{a_n^2} \sum_{k=1}^n \frac{a_k^2 \mathbb{E}[X_k^2]}{a_k^2}. \end{aligned}$$

Comme  $\sum_{k=1}^n \frac{\mathbb{E}[X_k^2]}{a_k^2}$  converge, le Lemme 8.1.4 (Kronecker) s'applique avec  $a_n^2 \rightarrow +\infty$  et assure  $\mathbb{E}[S_n^2]/a_n^2 \rightarrow 0$ , d'où la convergence L<sup>2</sup> de  $S_n/a_n$  vers 0.

Ensuite, pour la convergence presque sûre : Comme les variables aléatoires  $X_n$ ,  $n \geq 1$ , sont indépendantes et centrées, on a

$$\mathbb{E}\left[\left(\sum_{k=1}^{+\infty} \frac{X_k}{a_k}\right)^2\right] = \mathbb{E}\left[\sum_{k=1}^{+\infty} \frac{X_k^2}{a_k^2} + \sum_{k \neq l} \frac{X_k X_l}{a_k a_l}\right]$$

$$\begin{aligned}
&= \sum_{k=1}^{+\infty} \frac{\mathbb{E}[X_k^2]}{a_k^2} + \sum_{k \neq l} \frac{\mathbb{E}[X_k]\mathbb{E}[X_l]}{a_k a_l} \\
&= \sum_{k=1}^{+\infty} \frac{\mathbb{E}[X_k^2]}{a_k^2} < +\infty,
\end{aligned}$$

et donc  $\sum_{k \geq 1} \frac{X_k}{a_k} < +\infty$  presque sûrement. Le Lemme 8.1.4 (Kronecker) s'applique encore et donne  $\lim_{n \rightarrow +\infty} (\sum_{k \leq n} X_k)/a_n = 0$  ps et cela achève la preuve du Théorème 8.1.4.  $\square$

Du Théorème 8.1.4, on déduit aisément le corollaire suivant qui couvre de nombreuses situations dont la LGN  $L^2$  en 3) :

**Corollaire 8.1.4** *Étant donnée  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes centrées  $L^2$ , de variances bornées  $\mathbb{E}[X_n^2] \leq c$  (par exemple avec les  $X_n$  toutes de même loi), on a*

1.  $\sum_{n \geq 1} \frac{1}{a_n^2} < +\infty \implies \frac{S_n}{a_n} \rightarrow 0$   $L^2$  et ps.
2. Dés que  $\alpha > 1/2$ , on a  $S_n/n^\alpha \rightarrow 0$   $L^2$  et ps.
3.  $S_n/n \rightarrow 0$   $L^2$  et ps.
4. Pour une suite  $(X_n)_{n \geq 1}$  de variable aléatoires iid  $L^2$ , on a  $S_n/n \rightarrow \mathbb{E}[X_1]$   $L^2$  et ps.

**Démonstration :** Les différents cas sont des situations particulières du Théorème 8.1.4 qui s'applique et donne les résultats annoncés.  $\square$

### 8.1.3 Applications de la LGN

#### Méthode de Monte Carlo

Soient  $B$  un domaine mesurable borné,  $f : B \rightarrow \mathbb{R}$  une application mesurable et  $(X_n)_{n \geq 1}$  une suite de variables aléatoires iid uniformes sur  $B$ . On suppose que  $f(X_1) \in L^1$  (par exemple, il suffit d'avoir  $f$  bornée sur  $B$ ). On a alors

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{ps} \frac{1}{\lambda(B)} \int_B f d\lambda, \quad n \rightarrow +\infty.$$

En effet,  $\mathbb{E}[f(X_1)] = \int f(x) \mathbb{P}_{X_1}(dx) = \int f(x) \frac{\mathbf{1}_B(x) dx}{\lambda(B)}$ . Le résultat s'obtient donc en appliquant la LGN forte aux variables  $f(X_i)$  iid  $L^1$ .

La méthode de Monte-Carlo permet donc de faire un calcul approché d'intégrale en utilisant la LGN. L'avantage par rapport aux méthodes classiques d'analyse numérique (trapèze, Simpson, interpolation) est qu'aucune régularité n'est à supposer pour  $f$ . L'inconvénient est que la convergence n'est que presque sûre (pas vraiment gênant en pratique) mais surtout qu'il n'y a pas facilement de contrôle de l'erreur commise dans l'approximation.

**Définition 8.1.1 (Moyenne et variance empiriques)** Étant donnée une suite de variables aléatoires  $(X_n)_{n \geq 1}$  iid, on définit :

- la moyenne empirique  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ .
- la variance empirique  $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2$ .

**Proposition 8.1.1 (Moyenne et variance empiriques)** Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoires iid de même loi  $Q$ . On suppose que  $Q$  admet un moment d'ordre 2 fini. Alors les moyenne et variance empirique  $\bar{X}_n$ ,  $S_n^2$  de l'échantillon  $(X_1, \dots, X_n)$  estiment l'espérance et la variance de  $Q$ , ie. quand  $n \rightarrow +\infty$ , on a

$$\begin{aligned}\bar{X}_n &\xrightarrow{ps} \mathbb{E}[X_1] = \int_{\mathbb{R}} x Q(dx), \\ S_n^2 &\xrightarrow{ps} \text{Var}(X_1) = \int_{\mathbb{R}} x^2 Q(dx) - \left( \int_{\mathbb{R}} x Q(dx) \right)^2.\end{aligned}$$

De plus,  $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1]$  et  $\mathbb{E}[S_n^2] = \frac{n-1}{n} \text{Var}(X_1)$ .

On préfère parfois considérer la variance empirique corrigée  $S_{n,c}^2 = \frac{n}{n-1} S_n^2$  qui vérifie  $\mathbb{E}[S_{n,c}^2] = \text{Var}(X_1)$ . Dans ce cas, on obtient un estimateur sans biais. La moyenne empirique, elle, est toujours sans biais. Dans le cas gaussien, on peut préciser le comportement de  $M_n$  et de  $S_n^2$ , cf. Théorème 9.3.2.

**Démonstration :** Il s'agit d'une application immédiate de la LGN. Pour la moyenne empirique  $\bar{X}_n$  : c'est immédiat. Pour la variance empirique, on a

$$\begin{aligned}S_n^2 &= \frac{1}{n} \left( \sum_{k=1}^n X_k^2 - 2 \sum_{k=1}^n \bar{X}_n X_k + n \bar{X}_n^2 \right) = \frac{1}{n} \left( \sum_{k=1}^n X_k^2 - 2n \bar{X}_n^2 + n \bar{X}_n^2 \right) \\ &= \frac{1}{n} \left( \sum_{k=1}^n X_k^2 - \bar{X}_n^2 \right) = \left( \frac{1}{n} \sum_{k=1}^n X_k^2 \right) - \bar{X}_n^2 \\ &\rightarrow \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{Var}(X) \quad (\text{par la LGN}).\end{aligned}$$

Puis  $\mathbb{E}[S_n^2] = \frac{1}{n} \left( n \mathbb{E}[X^2] - n \mathbb{E}[\bar{X}_n^2] \right) = \mathbb{E}[X^2] - \mathbb{E}[\bar{X}_n^2]$ . Or

$$\mathbb{E}[\bar{X}_n^2] = \frac{1}{n^2} \mathbb{E} \left[ \left( \sum_{k=1}^n X_k \right)^2 \right] = \frac{1}{n^2} (n \mathbb{E}[X^2] + n(n-1) \mathbb{E}[X]^2)$$

car les variables aléatoires sont iid. On a donc

$$\mathbb{E}[S_n^2] = \frac{1}{n} (n \mathbb{E}[X^2] - (\mathbb{E}[X^2] - (n-1) \mathbb{E}[X]^2)) = \frac{n-1}{n} \text{Var}(X)$$

et  $\mathbb{E}[S_{n,c}^2] = \text{Var}(X_1)$ . □

On renvoie à [Dress] pour de nombreux exemples simples d'estimation.

## 8.2 Théorème central limite (TCL)

Le TCL est un résultat fondamental qui explique que les fluctuations autour de leur moyenne des effets cumulés de phénomènes répétés indépendamment, après normalisation, se comporte comme la loi  $\mathcal{N}(0, 1)$ . Autrement dit (et grossièrement dit), la somme d'un grand nombre de variables aléatoires indépendantes (et de variance finie) suit, à peu près, une loi normale. On propose le TCL classique (pour des variables aléatoires *iid*) dans la Section 8.2.1 avant de le généraliser à des variables indépendantes non identiquement distribuées dans la Section ??.

### 8.2.1 TCL classique et applications

La première version du TCL (théorème de Moivre-Laplace) concerne les lois de Bernoulli et est due à De Moivre (1733) qui l'a intuitionnée et Laplace (1809) qui l'a prouvée. La version standard du TCL date du 20ème siècle et est prouvée par Pólya (1920)<sup>1</sup>.

**Théorème 8.2.1 (TCL)** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles *iid*  $L^2$ , d'espérance  $m \in \mathbb{R}$  et de variance finie  $\sigma^2 > 0$ . Soit  $S_n = X_1 + \dots + X_n$  la somme partielle. Alors quand  $n \rightarrow +\infty$

$$\frac{S_n - nm}{\sqrt{n\sigma^2}} \xrightarrow{} \mathcal{N}(0, 1).$$

**Remarque 8.2.1** — D'un certain point de vue ce résultat complète la loi des grands nombres : la LGN dit que la moyenne arithmétique

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$$

converge quand  $n \rightarrow +\infty$  vers l'espérance  $m = \mathbb{E}[X_1]$  (la moyenne probabiliste) en probabilité ou presque sûrement selon la version de la LGN. On sait donc que  $S_n = nm + o(n)$ . Le TCL précise le comportement du  $o(n)$  et indique que la vitesse de la convergence  $S_n/n \rightarrow 0$  est en  $1/\sqrt{n}$  (en un sens faible car la convergence du TCL est une convergence en loi soit une convergence assez faible).

- En plus dans le TCL, apparaît à la limite la loi  $\mathcal{N}(0, 1)$  alors que les variables aléatoires  $X_i$  sont de lois arbitraires (avec une variance finie) : ce résultat justifie le rôle universel de la loi normale. Elle modélise les petites variations de n'importe quelle loi (avec un moment d'ordre 2) par rapport à sa moyenne.

**Démonstration :**[TCL] En remplaçant  $X_j$  par  $\frac{X_j - \mathbb{E}[X_j]}{\sqrt{\text{Var}(X_j)}} = \frac{X_j - m}{\sigma}$ , on peut se ramener à  $\mathbb{E}[X_j] = 0$  et  $\text{Var}(X_j) = 1$ . Comme  $X \in L^2$ , sa fonction caractéristique est  $C^2$  et la formule

---

1. über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem [Sur le théorème central du calcul probabiliste, parmi ceux ayant rapport à la notion de limite, et le problème des moments]

de Taylor à l'ordre 2 s'écrit

$$\varphi_X(t) = 1 + t\varphi'_X(0) + \frac{t^2}{2}\varphi''_X(0) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2) \quad \text{lorsque } t \rightarrow 0.$$

Puis comme les  $X_k$  sont *iid*, on a  $\varphi_{S_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t) = \varphi_X(t)^n$ , et donc

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_{S_n}(t/\sqrt{n}) = \varphi_X(t/\sqrt{n})^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \quad \text{lorsque } n \rightarrow +\infty.$$

On applique le Lemme 8.1.2 avec  $a_i = \varphi_X(t/\sqrt{n})$ ,  $b_i = (1 - \frac{t^2}{2n})$ ,  $1 \leq i \leq n$  :

$$\left| \varphi_X\left(\frac{t}{\sqrt{n}}\right)^n - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq n \left| \varphi_X\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right) \right|.$$

Puis le Lemme 8.1.1 assure

$$n \left| \varphi_X\left(\frac{t}{\sqrt{n}}\right) - \left(1 - \frac{t^2}{2n}\right) \right| \leq n \mathbb{E} \left[ \min \left( \frac{t^3 |X|^3}{6n^{3/2}}, \frac{t^2 |X|^2}{n} \right) \right] = \mathbb{E} \left[ \min \left( \frac{t^3 |X|^3}{6n^{1/2}}, t^2 |X|^2 \right) \right]$$

avec de plus

$$\min \left( \frac{t^3 |X|^3}{6n^{1/2}}, t^2 |X|^2 \right) \leq t^2 |X|^2 \in L^1, \quad \text{et} \quad \min \left( \frac{t^3 |X|^3}{6n^{1/2}}, t^2 |X|^2 \right) \leq \frac{t^3 |X|^3}{6\sqrt{n}} \rightarrow 0.$$

Le théorème de convergence dominée assure donc

$$\lim_{n \rightarrow +\infty} \varphi_{S_n/\sqrt{n}}(t) = \lim_{n \rightarrow +\infty} \left(1 - \frac{t^2}{2n}\right)^n = \exp(-t^2/2),$$

ie.  $S_n/\sqrt{n} \Rightarrow \mathcal{N}(0, 1)$  par le théorème de Paul Lévy (Théorème 7.4.3).  $\square$

**Remarque 8.2.2 (Utilisation pratique du TCL)** En général, lorsque  $n$  est grand, on approxime la loi d'une somme de variables aléatoires *iid* de  $L^2(\Omega)$  par une loi normale grâce au TCL de la façon suivante. Soit  $S_n = X_1 + \dots + X_n$  la somme de variables aléatoires *iid*  $X_i$  avec  $\sigma^2 < +\infty$ , on a d'après le TCL

$$\frac{X_1 + \dots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \implies \mathcal{N}(0, 1).$$

Quand  $n$  est grand on approxime alors la loi de  $\frac{X_1 + \dots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}}$  par celle de  $Y = \mathcal{N}(0, 1)$ . Si bien que la loi de la somme  $S_n = X_1 + \dots + X_n$  est approximée par celle de

$$n\mathbb{E}[X_1] + \sigma\sqrt{n}Y \simeq \mathcal{N}(n\mathbb{E}[X_1], \sigma^2 n).$$

**Règle statistique :** La somme  $S_n$  d'une suite de variables aléatoires iid  $L^2$  de moyenne  $m$  et de variance  $\sigma^2$  s'approxime par

$$S_n \approx \mathcal{N}(nm, \sigma^2 n).$$

**Théorème de Moivre-Laplace.** En particulier comme une loi binomiale  $\mathcal{B}(n, p)$  peut se voir comme la loi d'une somme  $S_n$  de  $n$  variables aléatoires  $X_i$ ,  $1 \leq i \leq n$ , de loi de Bernoulli  $b(p)$  indépendantes, on a l'approximation d'une loi binomiale par une loi normale

$$\mathcal{B}(n, p) \approx \mathcal{N}(np, np(1-p)).$$

**Exemple.** Un joueur lance une pièce équilibrée : lorsqu'il obtient pile, il gagne 100 Euros, lorsqu'il obtient face, il perd 100 Euros. Estimer le nombre maximal de lancers à effectuer pour que ce joueur ait plus de 95 chances sur 100 de perdre au plus 2000 Euros.

Notons  $n$  le nombre de lancers effectués, la variable aléatoire  $S_n$  égale au nombre de piles obtenus sur les  $n$  premiers lancers suit une loi  $\mathcal{B}(n, 1/2)$  et le gain (algébrique) vaut :

$$G_n = 100 \times S_n - 100 \times (n - S_n) = 200S_n - 100n.$$

On cherche alors  $n$  tel que  $\mathbb{P}(G_n \geq -2000) \geq 0,95$ . Or  $\{G_n \geq -2000\} = \{S_n - n/2 \geq -10\}$ . Comme  $S_n$  de loi binomiale, peut être vue comme une somme  $S_n = \epsilon_1 + \dots + \epsilon_n$  de variables aléatoires de loi  $b(1/2)$ , d'après le TCL

$$S_n^* = \frac{S_n - n/2}{\sqrt{n/4}} \Rightarrow \mathcal{N}(0, 1).$$

Si bien qu'on approxime la loi de  $S_n^*$  par  $\mathcal{N}(0, 1)$  et donc celle de  $S_n$  par la loi normale  $\mathcal{N}\left(\frac{n}{2}, \frac{n}{4}\right)$ .

Chercher  $n$  tel que  $\mathbb{P}(G_n \geq -2000) = \mathbb{P}(S_n - n/2 \geq -10) \geq 0,95$  revient à estimer  $n$  tel que

$$\mathbb{P}(N \geq -20/\sqrt{n}) \geq 0,95 \quad \text{ou} \quad \mathbb{P}(N \leq 20/\sqrt{n}) \geq 0,95$$

où  $N \sim \mathcal{N}(0, 1)$  et par symétrie de la loi  $\mathcal{N}(0, 1)$ . La table de la loi  $\mathcal{N}(0, 1)$  donne alors

$$\frac{20}{\sqrt{n}} = 1,65 \quad \text{c'est à dire} \quad n = \left(\frac{20}{1,65}\right)^2 = 146.$$

**Remarque 8.2.3 (Autres applications statistiques)** Le TCL est fondamental en statistique pour établir des intervalles de confiance (asymptotiques) ou pour faire des tests d'hypothèses à partir d'échantillons aléatoires (test du  $\chi^2$  (chi-deux), test de Student ou autres).

# Chapitre 9

## Vecteurs gaussiens

Le TCL justifie l'importance de la loi normale. En effet, on approxime (en loi) de nombreuses variables aléatoires par des variables aléatoires normales ou des vecteurs gaussiens en dimension  $d \geq 2$ . Il importe donc de savoir manipuler correctement ce type de vecteurs. Ces vecteurs sont caractérisés par leur (vecteur) moyenne et leur (matrice de) covariance et la plupart des calculs liés utilisent abondamment la structure hilbertienne de  $L^2$ .

### 9.1 Variables aléatoires gaussiennes

**Définition 9.1.1 (Variable aléatoire gaussienne (normale))** Une variable aléatoire réelle  $X$  suit la loi normale standard  $\mathcal{N}(0, 1)$  si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

De façon générale, si  $m \in \mathbb{R}$  et  $\sigma^2 > 0$ , une variable aléatoire réelle  $X$  suit la loi normale  $\mathcal{N}(m, \sigma^2)$  si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Si  $\sigma^2 = 0$ , la loi est dégénérée et la variable aléatoire  $X$  est constante égale à  $m$ . Sa loi est un Dirac en  $m$  :  $\mathbb{P}_X = \delta_m$ .

On rappelle que  $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$  justifie la normalisation de la loi  $\mathcal{N}(0, 1)$ . Par ailleurs, rappelons que :

**Proposition 9.1.1** Une variable aléatoire  $X \sim \mathcal{N}(m, \sigma^2)$  peut se voir comme la translatée et dilatée d'une variable aléatoire  $X_0$  de loi normale standard  $\mathcal{N}(0, 1)$  par

$$X = m + \sigma X_0.$$

Autrement dit si  $X \sim \mathcal{N}(m, \sigma^2)$ ,  $\sigma^2 > 0$ , on définit la variable centrée réduite  $\tilde{X} = \frac{X - m}{\sigma}$ . Elle suit la loi  $\mathcal{N}(0, 1)$ . Rappelons encore que :

**Proposition 9.1.2** *Une variable aléatoire  $X$  de loi  $\mathcal{N}(m, \sigma^2)$  a pour*

- *Espérance :  $\mathbb{E}[X] = m$ .*
- *Variance :  $\text{Var}(X) = \sigma^2$ .*
- *Fonction caractéristique :  $\varphi_X(t) = \exp(imt - \sigma^2 t^2/2)$ .*

**Proposition 9.1.3** *Soient  $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$  et  $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$  indépendantes. Alors  $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .*

**Démonstration :** Cf. Proposition 6.3.2. Une fois prouvé que la loi est normale  $\mathcal{N}(m, \sigma^2)$ , on retrouve facilement les paramètres de la loi de  $X_1 + X_2$  puisque

$$\begin{aligned} m &= \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = m_1 + m_2 \\ \sigma^2 &= \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = \sigma_1^2 + \sigma_2^2 \end{aligned}$$

où on utilise l'indépendance pour le calcul de la variance.  $\square$

## 9.2 Vecteurs gaussiens

Dans la suite, pour simplifier la présentation, on note sous la forme de transposée de vecteurs lignes les vecteurs colonnes :  $X = (X_1, \dots, X_d)^t$ . On considère le produit scalaire euclidien : pour  $x = (x_1, \dots, x_d)^t, y = (y_1, \dots, y_d)^t \in \mathbb{R}^d$ , on a  $\langle x, y \rangle = x^t y = \sum_{i=1}^d x_i y_i$ . On étudie maintenant la version multidimensionnelle des variables normales.

**Définition 9.2.1 (Vecteur gaussien)** *Un vecteur aléatoire  $X = (X_1, \dots, X_d)^t$  est gaussien si et seulement si toutes les combinaisons linéaires de ses coordonnées  $\langle a, X \rangle = a_1 X_1 + \dots + a_d X_d$  suit une loi gaussienne dans  $\mathbb{R}$  (pour tout  $a = (a_1, \dots, a_d)^t \in \mathbb{R}^d$ ).*

**Proposition 9.2.1** *L'image linéaire d'un vecteur gaussien est encore un vecteur gaussien : soit  $X = (X_1, \dots, X_d)^t$  un vecteur gaussien de dimension  $d$  et  $A \in \mathcal{M}_{p,d}(\mathbb{R})$ . Alors  $AX$  est un vecteur gaussien en dimension  $p$ .*

**Démonstration :** En effet si  $a \in \mathbb{R}^p$  alors  $\langle a, AX \rangle = \langle A^t a, X \rangle$  suit une loi normale car combinaison linéaire des marginales du vecteur gaussien  $X$ .  $\square$

De façon générale, un vecteur  $X = (X_1, \dots, X_d)^t$  est dit  $L^1$ , resp.  $L^2$ , si  $\mathbb{E}[|X_i|] < +\infty$ , resp.  $\mathbb{E}[X_i^2] < +\infty$ , pour tout  $1 \leq i \leq d$ . Un vecteur gaussien est nécessairement  $L^2$  puisque, par définition, chacune de ses marginales  $X_i$  est gaussienne donc  $L^2$ .

**Définition 9.2.2** La matrice de covariance d'un vecteur aléatoire  $X = (X_1, \dots, X_d)^t \in L^2$  est la matrice carrée symétrique, positive

$$K = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}.$$

L'espérance d'un vecteur aléatoire  $X = (X_1, \dots, X_d)^t \in L^1$  est le vecteur des espérances de ses marginales

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^t.$$

Si  $\mathbb{E}[X] = 0$ , le vecteur  $X$  est dit centré.

### Fonction caractéristique gaussienne en dimension $d$

Si  $X = (X_1, \dots, X_d)^t$  est un vecteur gaussien alors  $\langle a, X \rangle = \sum_{i=1}^d a_i X_i$  suit une loi normale de paramètres

$$\begin{aligned} \mathbb{E}[\langle a, X \rangle] &= \mathbb{E}[a_1 X_1 + \dots + a_d X_d] = a_1 \mathbb{E}[X_1] + \dots + a_d \mathbb{E}[X_d] = \langle a, \mathbb{E}[X] \rangle \\ \text{Var}(\langle a, X \rangle) &= \text{Var}(a_1 X_1 + \dots + a_d X_d) = \sum_{i,j=1}^d a_i a_j \text{Cov}(X_i, X_j) = a^t \text{Cov}(X) a \end{aligned}$$

en utilisant des notations matricielles. Ces égalités caractérisent la moyenne  $m = \mathbb{E}[X]$  et la covariance  $K = \text{Cov}(X)$  de  $X$ . La variable aléatoire gaussienne  $\langle a, X \rangle$  suit donc la loi  $\mathcal{N}(\langle a, \mathbb{E}[X] \rangle, a^t \text{Cov}(X) a)$ , sa fonction caractéristique est alors donnée par

$$\varphi_{\langle a, X \rangle}(x) = \exp \left( ix \langle a, \mathbb{E}[X] \rangle - \frac{1}{2} (a^t \text{Cov}(X) a) x^2 \right). \quad (9.1)$$

D'après les définitions des fonctions caractéristiques d'une variable aléatoire et d'un vecteur aléatoire

$$\varphi_X(x) = \mathbb{E}[e^{i\langle x, X \rangle}] = \varphi_{\langle x, X \rangle}(1).$$

On déduit alors de l'expression (9.1) :

**Proposition 9.2.2** La fonction caractéristique d'un vecteur gaussien  $X = (X_1, \dots, X_d)^t$  est donnée par

$$\varphi_X(x) = \exp \left( i \langle x, \mathbb{E}[X] \rangle - \frac{1}{2} \langle x, \text{Cov}(X) x \rangle \right). \quad (9.2)$$

La loi d'un vecteur gaussien est donc connue dès qu'on a le vecteur moyenne  $\mathbb{E}[X]$  et la matrice de covariance  $\text{Cov}(X)$ . On note alors :

**Définition 9.2.3** Dans la suite, on note  $X \sim \mathcal{N}_d(m, K)$  où  $m \in \mathbb{R}^d$  et  $K$  est une matrice carrée  $d \times d$  symétrique positive pour signifier que  $X$  est un vecteur gaussien de moyenne  $m \in \mathbb{R}^d$  et de matrice de covariance  $K$ .

**Remarque 9.2.1** — On identifie les paramètres  $m$  et  $K$  d'une loi  $\mathcal{N}_d(m, K)$  à l'aide de l'expression de la fonction caractéristique par exemple si

$$\varphi_X(s, t) = \exp\left(2is + 3it - \frac{1}{2}s^2 + st - t^2\right)$$

alors  $X \sim \mathcal{N}_2(m, K)$  avec

$$m = (2, 3)^t \quad \text{et} \quad K = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

- On parle du vecteur gaussien standard en dimension  $d$  lorsque  $\mathbb{E}[X] = 0$  et  $\text{Cov}(X) = I_d$ . Sa fonction caractéristique est alors

$$\varphi_X(x) = \exp(-\langle x, x \rangle / 2) = \exp(-\|x\|^2 / 2).$$

- Pour un vecteur gaussien centré,  $\mathbb{E}[X] = 0$  et on montre que

$$\langle x, \text{Cov}(X)x \rangle = \mathbb{E}[\langle x, X \rangle^2],$$

la fonction caractéristique devient alors

$$\varphi_X(x) = \exp\left(-\frac{1}{2}\langle x, \text{Cov}(X)x \rangle\right) = \exp\left(-\frac{1}{2}\mathbb{E}[\langle x, X \rangle^2]\right).$$

- En prenant  $x = (x_1, 0, \dots, 0)^t$ , on a

$$\varphi_X(x) = \varphi_{X_1}(x_1) = \exp(i\mathbb{E}[X_1]x_1 - \text{Var}(X_1)x_1^2/2).$$

On retrouve que  $X_1 \sim \mathcal{N}(\mathbb{E}[X_1], \text{Var}(X_1))$ . De même, pour tout  $1 \leq i \leq d$ , on a  $X_i \sim \mathcal{N}(\mathbb{E}[X_i], \text{Var}(X_i))$ .

**Proposition 9.2.3** Soit  $X \sim \mathcal{N}_d(m, K)$  un vecteur gaussien de dimension  $d$  et  $A \in \mathcal{M}_{p,d}(\mathbb{R})$  alors  $AX \sim \mathcal{N}_p(Am, AKA^t)$ .

**Démonstration :** D'après la Proposition 9.2.1,  $AX$  est gaussien. Si  $a \in \mathbb{R}^p$ , on a

- pour la moyenne :

$$\mathbb{E}[\langle a, AX \rangle] = \mathbb{E}[\langle A^t a, X \rangle] = \langle A^t a, m \rangle = \langle a, Am \rangle$$

ce qui assure que la moyenne de  $AX$  est  $\mathbb{E}[AX] = Am$ ;

- pour la covariance, on a :

$$\text{Var}(\langle a, AX \rangle) = \text{Var}(\langle A^t a, X \rangle) = (A^t a)^t K (A^t a) = a^t (AKA^t) a$$

ce qui garantit que la covariance de  $AX$  est  $AKA^t$ .

□

**Corollaire 9.2.1** *Un vecteur gaussien standard  $X \sim \mathcal{N}(0, I_d)$  est de loi invariante par rotation.*

**Démonstration :** La matrice d'une rotation est  $U \in O_d(\mathbb{R})$  (matrice orthogonale :  $UU^t = U^tU = I_d$ ). D'après la Prop. 9.2.1, si  $X \sim \mathcal{N}(0, I_d)$  alors  $UX \sim \mathcal{N}(0, UI_dU^t) = \mathcal{N}(0, UU^t) = \mathcal{N}(0, I_d)$ .  $\square$

**Définition 9.2.4 (Vecteur gaussien dégénéré)** *Un vecteur gaussien  $X \sim \mathcal{N}_d(m, K)$  est dit dégénéré si sa matrice de covariance est non inversible, ie.  $\det K = 0$ .*

En fait si  $X \sim \mathcal{N}_d(m, K)$  est dégénéré en dimension  $d$  si et seulement s'il existe  $a \in \mathbb{R}^d \setminus \{0\}$  tel que  $Ka = 0$ . On a alors  $\text{Var}(\langle a, X \rangle) = a^t K a = 0$  et donc  $\langle a, X \rangle = \mathbb{E}[\langle a, X \rangle]$ . Par conséquent, le vecteur  $X$  prend ses valeurs dans l'hyperplan d'équation  $\langle a, x \rangle = b$  où on a noté  $b = \mathbb{E}[\langle a, X \rangle]$ . Le vecteur  $X$  vit donc dans un sous-espace (affine) de dimension strictement plus petite que  $d$  (il vit peut être même dans un espace affine encore plus petit que l'hyperplan affine exhibé, en fait il vit dans un espace de dimension le rang de  $K$ ). Des exemples simples de vecteurs gaussiens dégénérés sont fournis par

- $X = (N_1, N_2, \dots, N_{n-1}, -N_1 - \dots - N_{n-1})$  où  $N_1, \dots, N_{n-1}$  sont des variables aléatoires normales indépendantes : il est clair que  $X$  est gaussien car

$$\langle A, X \rangle = \sum_{i=1}^{n_1} (a_i - a_n) N_i \text{ est normale}$$

et  $X$  vit dans l'hyperplan d'équation  $x_1 + \dots + x_n = 0$ .

- $X = (N, \dots, N)$  où  $N \sim \mathcal{N}(0, 1)$  : c'est bien un vecteur gaussien car  $\langle a, X \rangle = \left(\sum_{i=1}^d a_i\right) N$  suit une loi normale et  $X$  vit sur la droite d'équation  $x_1 = \dots = x_d$ .

Dans le cas non dégénéré, comme pour les variables aléatoires gaussiennes, on peut se ramener à un vecteur gaussien standard en centrant et en réduisant un vecteur gaussien quelconque non dégénéré :

**Proposition 9.2.4** *Soit  $X \sim \mathcal{N}_d(m, K)$  un vecteur gaussien non dégénéré de moyenne  $m \in \mathbb{R}^d$  et de matrice de covariance  $K$ . Alors*

$$\sqrt{K}^{-1}(X - m) \sim \mathcal{N}_d(0, I_d). \quad (9.3)$$

**Démonstration :** Comme le vecteur  $X$  est non dégénéré, sa matrice de covariance  $K$  est définie (c'est à dire inversible). Il existe donc une matrice  $A = \sqrt{K}$  inversible telle que  $K = AA^t$ . Il est donc légitime d'utiliser  $\sqrt{K}^{-1}$  dans (9.3).

On montre maintenant que  $\tilde{X} = \sqrt{K}^{-1}(X - m)$  est gaussien, standard :

$$\begin{aligned} \varphi_{\tilde{X}}(x) &= \mathbb{E} \left[ \exp(i\langle x, \tilde{X} \rangle) \right] \\ &= \mathbb{E} \left[ \exp(i\langle x, \sqrt{K}^{-1}(X - m) \rangle) \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ \exp(i \langle (\sqrt{K}^{-1})^t x, X - m \rangle) \right] \\
&= \mathbb{E} \left[ \exp(i \langle (\sqrt{K}^{-1})^t x, X \rangle) \right] \times \exp \left( -i \langle (\sqrt{K}^{-1})^t x, m \rangle \right) \\
&= \varphi_X \left( (\sqrt{K}^{-1})^t x \right) \times \exp \left( -i \langle (\sqrt{K}^{-1})^t x, m \rangle \right) \\
&= \exp \left( i \langle m, (\sqrt{K}^{-1})^t x \rangle - \frac{1}{2} \langle (\sqrt{K}^{-1})^t x, K(\sqrt{K}^{-1})^t x \rangle \right) \times \exp \left( -i \langle (\sqrt{K}^{-1})^t x, m \rangle \right) \\
&= \exp \left( -\frac{1}{2} \langle (\sqrt{K}^{-1})^t x, K(\sqrt{K}^{-1})^t x \rangle \right) \\
&= \exp \left( -\frac{1}{2} \langle x, \sqrt{K}^{-1} K(\sqrt{K}^{-1})^t x \rangle \right) = \exp \left( -\frac{1}{2} \langle x, \sqrt{K}^{-1} \sqrt{K} (\sqrt{K})^t (\sqrt{K}^{-1})^t x \rangle \right) \\
&= \exp \left( -\frac{1}{2} \langle x, x \rangle \right) = \exp \left( -\frac{1}{2} \|x\|^2 \right).
\end{aligned}$$

On a donc bien  $\tilde{X} \sim \mathcal{N}_d(0, I_d)$ .

On peut aussi utiliser la Proposition 9.2.3 : d'abord  $X - m$  est de loi  $\mathcal{N}(0, K)$  puis  $\tilde{X}$  est l'image de  $X - m$  par  $f(x) = \sqrt{K}^{-1} x$  donc  $\tilde{X}$  a pour loi  $\mathcal{N}(0, \sqrt{K}^{-1} K (\sqrt{K}^{-1})^t) = \mathcal{N}(0, I_d)$ .  $\square$

**Remarque 9.2.2 (Centrer-réduire un vecteur gaussien)** Comme en dimension 1, on peut centrer-réduire une loi gaussienne : une variable aléatoire  $X \sim \mathcal{N}_d(m, K)$  avec  $K$  inversible peut se voir comme une translatée et dilatée du vecteur gaussien standard  $N \sim \mathcal{N}_d(0, I_d)$  :

$$X \sim \sqrt{K}N + m.$$

### Indépendance de variables aléatoires gaussiennes

**Proposition 9.2.5** Soit  $(X, Y)$  un couple gaussien. Alors  $X$  et  $Y$  sont indépendantes si et seulement si  $\text{Cov}(X, Y) = 0$ .

**Démonstration :** Le sens direct est vrai quelque soit les lois de  $X$  et de  $Y$ , carrés intégrables. Pour la réciproque, on sait que  $X$  et  $Y$  sont indépendantes si et seulement si  $\varphi_{(X,Y)}(t_1, t_2) = \varphi_X(t_1)\varphi_Y(t_2)$ . Or  $(X, Y)$  est un vecteur gaussien de matrice de covariance diagonale

$$\begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$$

car  $\text{Cov}(X, Y) = \text{Cov}(Y, X) = 0$ . On déduit de (9.2) que

$$\begin{aligned}
\varphi_{(X,Y)}(t_1, t_2) &= \exp \left( im_1 t_1 + im_2 t_2 - \frac{1}{2} (t_1^2 \sigma_X^2 + t_2^2 \sigma_Y^2) \right) \\
&= \exp \left( im_1 t_1 - \frac{1}{2} t_1^2 \sigma_X^2 \right) \times \exp \left( -im_2 t_2 - \frac{1}{2} t_2^2 \sigma_Y^2 \right)
\end{aligned}$$

$$= \varphi_X(t_1)\varphi_Y(t_2),$$

ce qui justifie l'indépendance.  $\square$

De la même façon, pour des vecteurs gaussiens, on a :

**Proposition 9.2.6** Soit  $(X_1, \dots, X_{d_1}, Y_1, \dots, Y_{d_2})^t$  un vecteur gaussien de dimension  $d_1 + d_2$ . Les deux vecteurs aléatoires gaussiens  $X = (X_1, \dots, X_{d_1})^t$  et  $Y = (Y_1, \dots, Y_{d_2})^t$  sont indépendants si et seulement si les covariances  $\text{Cov}(X_i, Y_j)$ ,  $1 \leq i \leq d_1, 1 \leq j \leq d_2$ , sont toutes nulles.

### Densité gaussienne en dimension $d$

Soit  $X \sim \mathcal{N}_d(0, I_d)$  un vecteur gaussien standard en dimension  $d$ . Comme  $\text{Cov}(X) = I_d$ , les marginales  $X_1, \dots, X_d$  sont toutes indépendantes. La loi du vecteur  $X = (X_1, \dots, X_d)^t$  est donc la loi produit de ses lois marginales

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}.$$

On a donc pour  $A \in \mathcal{B}(\mathbb{R}^d)$  :

$$\begin{aligned} \mathbb{P}(X \in A) &= \int_A \mathbb{P}_X(dx) = \int_A \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}(dx) \\ &= \int_A \mathbb{P}_{X_1}(dx_1) \dots \mathbb{P}_{X_d}(dx_d) \\ &= \int_A f_{X_1}(x_1) dx_1 \dots f_{X_d}(x_d) dx_d \\ &= \int_A f_{X_1}(x_1) \dots f_{X_d}(x_d) dx_1 \dots dx_d. \end{aligned}$$

La densité de  $X$  est donc donnée par

$$\begin{aligned} f_X(x_1, \dots, x_d) &= f_{X_1}(x_1) \times \cdots \times f_{X_d}(x_d) \\ &= \left( \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \right) \times \cdots \times \left( \frac{1}{\sqrt{2\pi}} e^{-x_d^2/2} \right) \\ &= \frac{1}{\sqrt{2\pi}^d} \exp(- (x_1^2 + \cdots + x_d^2)/2). \end{aligned}$$

On peut énoncer :

**Proposition 9.2.7 (Densité gaussienne standard)** La densité d'un vecteur gaussien standard  $\mathcal{N}_d(0, I_d)$  en dimension  $d$  est

$$f_X(x) = \frac{1}{\sqrt{2\pi}^d} \exp(- (x_1^2 + \cdots + x_d^2)/2) = \frac{1}{\sqrt{2\pi}^d} \exp(- \|x\|^2/2).$$

Pour passer au cas général d'un vecteur gaussien  $X \sim \mathcal{N}_d(m, K)$ , on suppose  $X$  non dégénéré sinon, c'est que  $X$  vit dans un sous-espace affine strict de  $\mathbb{R}^d$ , de mesure nulle pour  $\lambda^d$  et il n'y a pas de densité dans  $\mathbb{R}^d$ . Il faut alors étudier un tel vecteur en dimension inférieure.

On considère donc  $X \sim \mathcal{N}_d(m, K)$  non dégénéré, c'est à dire avec  $K = \text{Cov}(X)$  inversible. On utilise le vecteur centré-réduit donné en (9.3) : on a  $X \sim \sqrt{K}N + m$  avec  $N \sim \mathcal{N}_d(0, I_d)$ .

$$\begin{aligned}\mathbb{P}(X \in A) &= \mathbb{P}(\sqrt{K}N + m \in A) \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\{\sqrt{K}y+m \in A\}} f(y) dy \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\{x \in A\}} f(\sqrt{K}^{-1}(x - m)) \frac{dx}{\det \sqrt{K}} \\ &= \int_A \frac{\exp(-\|\sqrt{K}^{-1}(x - m)\|^2/2)}{((2\pi)^d \det K)^{1/2}} dx \\ &= \int_A \frac{\exp(-\langle(x - m), K^{-1}(x - m)\rangle/2)}{((2\pi)^d \det K)^{1/2}} dx\end{aligned}$$

où on a fait à la 4ème ligne le changement de variable  $y = \sqrt{K}x + m$ . On a prouvé :

**Proposition 9.2.8 (Densité gaussienne général)** *La densité d'un vecteur gaussien  $X \sim \mathcal{N}_d(m, K)$  non dégénéré est*

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

### Vecteurs non gaussiens avec des variables marginales gaussiennes

On a déjà vu que si un vecteur  $X = (X_1, \dots, X_d)^t$  est gaussien alors ses marginales  $X_i$  le sont aussi, de même les combinaisons linéaires de ses marginales le sont. **La réciproque est fausse :** si des variables aléatoires sont gaussiennes alors le vecteur formé par ces variables n'est pas nécessairement gaussien. En effet, prenons  $X$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$  et  $Y$  de loi donnée, pour  $a > 0$  fixé, par

$$Y = \begin{cases} X & \text{si } |X| \leq a, \\ -X & \text{si } |X| > a. \end{cases}$$

Alors  $Y$  est de loi  $\mathcal{N}(0, 1)$  en effet

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}[e^{itY}] = \mathbb{E}[e^{itX} \mathbf{1}_{|X| \leq a}] + \mathbb{E}[e^{-itX} \mathbf{1}_{|X| > a}] \\ &= \mathbb{E}[e^{itX} \mathbf{1}_{\{|X| \leq a\}}] + \mathbb{E}[e^{-itX} \mathbf{1}_{\{|X| > a\}}] = \mathbb{E}[e^{itX} \mathbf{1}_{\{|X| \leq a\}}] + \mathbb{E}[e^{itX} \mathbf{1}_{\{|X| > a\}}] \\ &= \mathbb{E}[e^{itX} (\mathbf{1}_{\{|X| \leq a\}} + \mathbf{1}_{\{|X| > a\}})] \\ &= \mathbb{E}[e^{itX}] = e^{-t^2/2}\end{aligned}$$

car la loi de  $X$  est symétrique :  $\mathcal{L}(X) = \mathcal{L}(-X)$ . Puis, la variable  $X + Y$  est donnée par

$$\begin{aligned} X + Y &= \begin{cases} X + X = 2X & \text{si } |X| \leq a \\ X - X = 0 & \text{si } |X| > a \end{cases} \\ &= 2X\mathbf{1}_{\{|X| \leq a\}}. \end{aligned}$$

La combinaison linéaire  $X + Y$  a un atome en 0 car  $\mathbb{P}(X + Y = 0) \geq \mathbb{P}(|X| > a) > 0$ . Elle ne suit donc pas une loi gaussienne. Le couple aléatoire  $(X, Y)$  n'est donc pas gaussien (sinon on devrait avoir  $X + Y$  de loi gaussienne).

De plus, cet exemple montre aussi que dans la Proposition 9.2.5, l'hypothèse  $(X, Y)$  gaussien est nécessaire et il ne suffit pas de supposer que  $X$  et  $Y$  sont des variables aléatoires gaussiennes. En effet,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY] = \mathbb{E}[X^2\mathbf{1}_{\{|X| \leq a\}}] - \mathbb{E}[X^2\mathbf{1}_{\{|X| > a\}}] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X^2\mathbf{1}_{\{|X| > a\}}] - \mathbb{E}[X^2\mathbf{1}_{\{|X| > a\}}] \\ &= 1 - 2\mathbb{E}[X^2\mathbf{1}_{\{|X| > a\}}]. \end{aligned}$$

La fonction  $u(a) = \mathbb{E}[X^2\mathbf{1}_{\{|X| > a\}}]$  tend vers 0 en  $+\infty$  par convergence dominée, est continue et vaut  $\mathbb{E}[X^2] = 1$  en 0. Par le théorème des valeurs intermédiaires, il existe donc  $a$  tel que  $u(a) = 1/2$  et  $\text{Cov}(X, Y) = 0$ . Pourtant,  $X$  et  $Y$  sont non indépendantes sinon la loi du couple  $(X, Y)$  serait

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y = \mathcal{N}(0, 1) \otimes \mathcal{N}(0, 1) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

qui est gaussienne, ce qui est faux. Pour cette valeur de  $a$ , on a donc des variables aléatoires gaussiennes  $X$  et  $Y$  non corrélées mais non indépendantes.

Un autre exemple simple est donné par  $X = (N, \varepsilon N)$  où  $N \sim \mathcal{N}(0, 1)$  et  $\varepsilon$  est indépendante de  $N$  et de loi  $\mathbb{P}(\varepsilon = -1) = \mathbb{P}(\varepsilon = 1) = \frac{1}{2}$ . On a  $\varepsilon N \sim \mathcal{N}(0, 1)$  car

$$\varphi_{\varepsilon N}(t) = \mathbb{E}[e^{it\varepsilon N}] = \mathbb{E}[\mathbb{E}[e^{it\varepsilon N}|\varepsilon]] = \mathbb{E}[\varphi_N(t\varepsilon)] = \mathbb{E}[\exp(-(t\varepsilon)^2/2)] = \exp(-t^2/2).$$

Mais  $X$  n'est pas gaussien car par exemple la combinaison linéaire de ses marginales  $N + \varepsilon N$  ne suit pas une loi normale (il y a un atome en 0). Pourtant,

$$\text{Cov}(N, \varepsilon N) = \mathbb{E}[N^2\varepsilon] = \mathbb{E}[N^2]\mathbb{E}[\varepsilon] = 0.$$

## 9.3 Applications

### 9.3.1 TCL multidimensionnel

Le théorème central limite se généralise pour des vecteurs aléatoires *iid*  $L^2$  non dégénérés.

**Théorème 9.3.1** Soit  $(X^n)_{n \geq 1}$  une suite de vecteurs aléatoires iid  $L^2$  de vecteur espérance  $m = \mathbb{E}[X^1] \in \mathbb{R}^d$  et de matrice de covariance  $K = \text{Cov}(X^1)$  non dégénérée. Alors en notant  $S_n = \sum_{i=1}^n X^i$  la somme partielle des vecteurs, on a

$$n^{-1/2} \sqrt{K}^{-1} (S_n - nm) \Rightarrow \mathcal{N}_d(0, I_d) \quad (9.4)$$

$$\text{ou} \quad n^{-1/2} (S_n - nm) \Rightarrow \mathcal{N}_d(0, K). \quad (9.5)$$

**Démonstration :** On note  $Z_n = n^{-1/2} \sqrt{K}^{-1} (S_n - nm)$ . D'après le théorème de Paul Lévy (Théorème 7.4.4), il s'agit de montrer que pour tout  $x \in \mathbb{R}^d$ , on a

$$\lim_{n \rightarrow +\infty} \varphi_{Z_n}(x) = \varphi_{\mathcal{N}_d(0, I_d)}(x) = \exp(-\|x\|^2/2).$$

Pour cela, on observe que

$$\langle Z_n, x \rangle = n^{-1/2} \left\langle \sqrt{K}^{-1} (S_n - nm), x \right\rangle = n^{-1/2} \langle (S_n - nm), y \rangle = n^{-1/2} \left( \sum_{i=1}^n Y_i - n \langle m, y \rangle \right)$$

où on a noté  $y = (\sqrt{K}^{-1})^t x$  et  $Y_i = \langle X^i, y \rangle$ . On a

$$\begin{aligned} \mathbb{E}[Y_i] &= \mathbb{E}[\langle X^i, y \rangle] = \langle m, y \rangle \\ \text{Var}(Y_i) &= \text{Var}(\langle X^i, y \rangle) = y^t K y = x^t \sqrt{K}^{-1} K (\sqrt{K}^{-1})^t x = x^t x = \|x\|^2. \end{aligned}$$

Le TCL usuel en dimension 1 (Théorème 8.2.1) appliqué aux variables aléatoires iid  $Y_i = \langle X^i, y \rangle$  donne alors

$$\|x\|^{-1} \langle Z_n, x \rangle = n^{-1/2} \|x\|^{-1} \left( \sum_{i=1}^n Y_i - n \mathbb{E}[Y_1] \right) \Rightarrow \mathcal{N}(0, 1).$$

D'où

$$\begin{aligned} \lim_{n \rightarrow +\infty} \varphi_{Z_n}(x) &= \lim_{n \rightarrow +\infty} \mathbb{E}[\exp(i \langle Z_n, x \rangle)] = \lim_{n \rightarrow +\infty} \mathbb{E}[\exp(i \|x\| (\langle Z_n, x \rangle / \|x\|))] \\ &= \lim_{n \rightarrow +\infty} \varphi_{\|x\|^{-1} \langle Z_n, x \rangle}(\|x\|) = \varphi_{\mathcal{N}(0, 1)}(\|x\|) = \exp(-\|x\|^2/2) \end{aligned}$$

comme souhaité pour (9.4). Pour (9.5), on applique le continuous mapping theorem (Th. 7.4.5) avec l'application linéaire  $x \in \mathbb{R}^d \mapsto \sqrt{K}x$  où  $\sqrt{K}$  est la matrice telle que  $K = \sqrt{K}\sqrt{K}^t$ . D'après la Proposition 9.2.3, l'image par cette application linéaire de  $\mathcal{N}_d(0, I_n)$  est  $\mathcal{N}_d(0, \sqrt{K}\sqrt{K}^t) = \mathcal{N}_d(0, K)$  ce qui prouve (9.5).  $\square$

### 9.3.2 Estimations : moyenne, variance

157

#### empiriques

D'abord, on précise avec le TCL le comportement asymptotique de la moyenne empirique  $\bar{X}_n$  et de la variance empirique  $S_n^2$ .

**Proposition 9.3.1** Soit  $(X_i)_{i \geq 1}$  une suite de variables aléatoires iid  $L^2$  de moyenne  $m$ , de variance  $\sigma^2 < +\infty$ .

1. On a  $\sqrt{n}(\bar{X}_n - m) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ .
2. Si  $\mathbb{E}[X_1^4] < +\infty$  alors en notant  $\mu_4 = \mathbb{E}[(X_1 - m)^4]$ , on a :

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_4 - \sigma^4)$$

**Démonstration :** 1) C'est une application directe du TCL pour les variables aléatoires  $(X_i)_{i \geq 1}$  iid :

$$\sqrt{n}(\bar{X}_n - m) = \sigma \frac{S_n - nm}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} \sigma \mathcal{N}(0, 1) = \mathcal{N}(0, \sigma^2).$$

2) Pour simplifier on prend  $m = 0$ . Comme

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}_n \frac{\sum_{i=1}^n X_i}{n} + \frac{n}{n} \bar{X}_n^2 = \frac{1}{n} \left( \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2.$$

on a

$$\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} = \frac{\sum_{i=1}^n X_i^2 - n\sigma^2}{\sqrt{n(\mu_4 - \sigma^4)}} - \frac{\sqrt{n}\bar{X}_n^2}{\sqrt{\mu_4 - \sigma^4}}. \quad (9.6)$$

Le TCL appliqué aux variables aléatoires iid  $X_i^2$ ,  $i \geq 1$ , de moyenne  $\sigma^2$  et de variance  $\mathbb{E}[X_1^4] - \mathbb{E}[X_1^2]^2 = \mu_4 - \sigma^4$ , on a

$$\frac{\sum_{i=1}^n X_i^2 - n\sigma^2}{\sqrt{n(\mu_4 - \sigma^4)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Puis comme par la LGNf, on a  $\bar{X}_n \xrightarrow{\mathbb{P}} 0$  et par le TCL, on a  $\sqrt{n}\bar{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ . En les combinant, le théorème de Slutsky assure  $\sqrt{n}\bar{X}_n^2 \xrightarrow{\mathbb{P}} 0$ . Finalement, on déduit de (9.6), encore avec le théorème de Slutsky, que

$$\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

□

Pour un échantillon de variables aléatoires iid gaussiennes, on peut préciser la Proposition 8.1.1 et la loi (non asymptotique) de l'espérance et de la variance empiriques :

**Théorème 9.3.2 (Fisher sur les moyenne, variance empiriques)** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires iid de même loi gaussienne  $\mathcal{N}(m, \sigma^2)$ . Alors

1.  $\bar{X}_n \perp\!\!\!\perp S_n^2$
2.  $\bar{X}_n \sim \mathcal{N}(m, \sigma^2/n)$  et  $\frac{n}{\sigma^2} S_n^2 \sim \chi^2(n-1)$ .

**Démonstration :** La première étape consiste à vérifier qu'on peut se ramener au cas  $m = 0$  et  $\sigma^2 = 1$ . Pour cela, on pose

$$X'_i = \frac{X_i - m}{\sigma} \iff X_i = \sigma X'_i + m, \quad 1 \leq i \leq n,$$

et on observe que les variables aléatoires  $X'_1, \dots, X'_n$  sont indépendantes, de loi  $\mathcal{N}(0, 1)$ . Notons  $\bar{X}'_n$  leur moyenne empirique et  $S'^2_n$  leur variance empirique. On a alors  $\bar{X}_n = \sigma \bar{X}'_n + m$  et comme  $(X_i - \bar{X}_n) = \sigma(X'_i - \bar{X}'_n)$ , on a aussi  $S_n^2 = \sigma^2 S'^2_n$ . Si le résultat est établi pour  $m = 0, \sigma = 1$ , il reste donc vrai dans le cas général.

On suppose chaque variable aléatoire  $X_i \sim \mathcal{N}(0, 1)$ . Soit  $u_1 = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}) \in \mathbb{R}^n$  de norme 1. On complète ce vecteur  $u_1$  en base orthonormée (*bon*)  $(u_1, \dots, u_n)$  de  $\mathbb{R}^n$  et on considère  $A$  la matrice carrée changement de base, ie. la matrice dont les colonnes sont  $u_1, \dots, u_n$ . Comme  $A$  est une matrice changement de base d'une *bon* (la base canonique) à une autre  $(u_1, \dots, u_n)$ ,  $A$  est orthogonale, ie.  $AA^t = A^t A = I_n$ . Notons  $Y = A^t X$  et  $Y^t = (Y_1, \dots, Y_n)$ . Le vecteur  $Y$  est gaussien de covariance  $K_Y = A^t K_X (A^t)^t = A^t I_n A = A^t A = I_n$  puisque, les variables  $X_i$  étant indépendantes centrées-réduites,  $K_X = I_n$ . En particulier, les variables aléatoires  $Y_1, \dots, Y_n$  sont indépendantes et de loi  $\mathcal{N}(0, 1)$ . La première ligne de  $A^t$  étant  $u_1^t = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ , on a

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n) = \sqrt{n}\bar{X}_n.$$

On exprime de la même façon  $S_n^2$  en terme de  $Y$  :

$$\begin{aligned} nS_n^2 &= \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=1}^n (X_k^2 - 2X_k \bar{X}_n + \bar{X}_n^2) \\ &= \left( \sum_{k=1}^n X_k^2 \right) - 2\bar{X}_n \left( \sum_{k=1}^n X_k \right) + n\bar{X}_n^2. \end{aligned}$$

Par conséquent,

$$nS_n^2 = \left( \sum_{k=1}^n X_k^2 \right) - 2\bar{X}_n(n\bar{X}_n) + n\bar{X}_n^2 = \left( \sum_{k=1}^n X_k^2 \right) - n\bar{X}_n^2.$$

Mais comme  $A^t$  est orthogonale, elle conserve la norme si bien que  $Y = A^t X$  a pour norme (au carré)  $\sum_{k=1}^n Y_k^2 = \sum_{k=1}^n X_k^2$ . On déduit alors  $nS_n^2 = \sum_{k=2}^n Y_k^2$ . Puis comme  $Y$  est un vecteur gaussien de matrice de covariance identité, les variables  $Y_1, \dots, Y_n$  sont indépendantes et de loi  $\mathcal{N}(0, 1)$ . On en déduit

- l'indépendance de  $\bar{X}_n = Y_1/\sqrt{n}$  et de  $S_n^2 = \frac{1}{n} \sum_{k=2}^n Y_k^2$  ;
- $\bar{X}_n \sim \mathcal{N}(0, \frac{1}{n})$  et  $S_n^2 \sim \chi^2(n-1)$  car somme de  $n-1$  carrés de variables aléatoires  $\mathcal{N}(0, 1)$  indépendantes.

□

### 9.3.3 Décomposition de vecteurs gaussiens et test du $\chi^2$

Le théorème de Cochran donne les lois des projections d'un vecteur aléatoire gaussien.

**Théorème 9.3.3 (Cochran)** Soit  $X$  un vecteur gaussien de loi  $\mathcal{N}_d(0, \sigma^2 I_d)$ ,  $\sigma^2 > 0$ . Soit une décomposition orthogonale  $\mathbb{R}^d = V_1 \oplus \dots \oplus V_k$  avec  $\dim(V_i) = d_i$ ,  $i = 1, \dots, k$  et  $V_i \perp V_j$ ,  $i \neq j$ . Alors les vecteurs  $\Pi_{V_1}(X), \dots, \Pi_{V_k}(X)$ , où  $\Pi_{V_i}$  est la projection orthogonale sur  $V_i$ , sont gaussiens indépendants et

$$\frac{\|\Pi_{V_i}(X)\|^2}{\sigma^2} \sim \chi^2(d_i), \quad i = 1, \dots, k.$$

Pour avoir un résultat analogue pour un vecteur gaussiens général  $X \sim \mathcal{N}_d(m, \Gamma)$  non dégénéré, il faut considérer des sous-espaces  $V_i$  orthogonaux pour un produit scalaire lié à  $\Gamma$  :  $\langle x, y \rangle_\Gamma = \langle x, \Gamma y \rangle = x^t \Gamma y$ ,  $x, y \in \mathbb{R}^d$ .

**Démonstration :** Soit  $(e_{i,j})_{i,j}$  une base orthonormée de  $\mathbb{R}^d$  adaptée à la décomposition  $\mathbb{R}^d = V_1 \oplus \dots \oplus V_k$ , ie.  $\{e_{i,j} : j = 1, \dots, d_i\}$  est une base orthonormée de  $V_i$ . Soit  $X = \sum_{i,j} \langle X, e_{i,j} \rangle e_{i,j}$  la décomposition de  $X$  dans cette base orthonormée. On pose  $\eta_{i,j} = \langle X, e_{i,j} \rangle / \sigma$  et on obtient un vecteur gaussien  $\eta = (\eta_{i,j})_{i,j} = X / \sigma \sim \mathcal{N}_d(0, I_d)$ . Mais

$$\Pi_{V_i}(X) = \sum_{j=1}^{d_i} \langle X, e_{i,j} \rangle e_{i,j} = \sigma \sum_{j=1}^{d_i} \eta_{i,j} e_{i,j}$$

ne dépend que des  $\eta_{i,j}$  pour  $j \in \{1, \dots, d_i\}$ . On en déduit que les vecteurs  $\Pi_{V_i}(X)$  sont indépendants et que pour  $i = 1, \dots, k$

$$\frac{\|\Pi_{V_i}(X)\|^2}{\sigma^2} = \sum_{j=1}^{d_i} \eta_{i,j}^2 \sim \chi^2(d_i).$$

□

#### Test d'adéquation du $\chi^2$

On observe une variable aléatoire discrète  $X$ , de support connu  $\mathcal{S}(X) = \{a_1, \dots, a_r\} \subset \mathbb{R}^r$  avec les probabilités ponctuelles  $p_j = \mathbb{P}(X = a_j) = Q(\{a_j\}) \in ]0, 1[$ ,  $j = 1, \dots, r$  inconnues. On note  $p = (p_1, \dots, p_r)$  le vecteur de ses probabilités.

Soit, par ailleurs, une probabilité  $Q_0 = \sum_{j=1}^r \pi_j \delta_{a_j}$  de même support mais avec les probabilités  $\pi = (\pi_1, \dots, \pi_r)$  connues ( $\pi_i > 0$ ,  $1 \leq i \leq r$ ). On souhaite étudier la possibilité que  $Q$  soit égal à  $Q_0$ , c'est à dire tester l'égalité de vecteurs de  $\mathbb{R}^r$  :  $p = \pi$ . L'hypothèse à tester est alors

$$(H) : Q = Q_0 \text{ contre } (A) : Q \neq Q_0.$$

On observe un échantillon  $X = (X_1, \dots, X_n)^t$  de la loi  $Q$  de taille  $n$ , ie.  $n$  variables aléatoires iid de loi  $Q$  et on compte le nombre de fois que la valeur  $a_j$ ,  $j = 1, \dots, r$ , est prise dans l'échantillon :

$$N_j = \sum_{i=1}^n \mathbf{1}_{\{X_i=a_j\}} = \text{le nombre des } X_i \text{ égaux à } a_j.$$

On peut voir que  $N = (N_1, \dots, N_r)^t$  suit la loi multinomiale  $\mathcal{M}(n, p_1, \dots, p_r)$  donnée par

$$\mathbb{P}((N_1, \dots, N_r) = (n_1, \dots, n_r)) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \dots p_r^{n_r}, \quad \text{avec } n_1 + \dots + n_r = n.$$

Soient  $\hat{p}_j = N_j/n$ ,  $j = 1, \dots, r$ , les fréquences empiriques. Par la loi forte des grands nombres, lorsque  $n \rightarrow +\infty$ , on a  $\hat{p}_j \rightarrow p_j = \mathbb{E}[\mathbf{1}_{\{X=a_j\}}] = \mathbb{P}(X = a_j)$  ps. Pour déterminer si  $p = \pi$ , on considère la distance entre  $\hat{p}$ , qui approche  $p$  lorsque  $n$  est grand, et  $\pi$ . En pratique, pour des fréquences d'apparition  $f_i$  de chaque atome  $a_i$ , on note  $f = (f_1, \dots, f_r) = (\frac{n_1}{n}, \dots, \frac{n_r}{n})$  le vecteur des fréquences empiriques et on définit la « distance » de Karl Pearson entre  $f$  et  $\pi = (\pi_1, \dots, \pi_r)$  par

$$\rho(f, \pi) = \sum_{j=1}^r \frac{(f_j - \pi_j)^2}{\pi_j}$$

(ce n'est pas vraiment une distance puisque  $\rho(f, \pi)$  n'est pas symétrique). Avec les fréquences empiriques  $\hat{p}$ , on considère alors la quantité :

$$T_n = n\rho(\hat{p}, \pi) = n \sum_{j=1}^r \frac{(\hat{p}_j - \pi_j)^2}{\pi_j} = \sum_{j=1}^r \frac{(N_j - n\pi_j)^2}{n\pi_j}$$

dont le comportement permet de décider laquelle des deux hypothèses  $(H)$  ou  $(A)$  est la plus vraisemblable :

- Sous l'hypothèse  $(H)$  :  $T_n$  est proche de 0 ;
- par contre sous l'hypothèse  $(A)$  ( $\hat{p} \neq \pi$ ), on a  $\rho(\hat{p}, \pi) \rightarrow \rho(p, \pi) > 0$  ps et donc  $T_n \rightarrow +\infty$  ps.

Avec ces observations, on peut alors formaliser le test du  $\chi^2$  (chi-deux) :

**Théorème 9.3.4 (Test d'adéquation du  $\chi^2$ )** Soient  $\pi_1, \dots, \pi_r > 0$ . Alors :

- sous l'hypothèse  $(H)$  :  $T_n \xrightarrow{} \chi^2(r-1)$ ,  $n \rightarrow +\infty$  ;
- sous l'hypothèse  $(A)$  :  $T_n \xrightarrow{} +\infty$  ps,  $n \rightarrow +\infty$ .

**Démonstration :** D'abord, sous  $(A)$ , il existe  $j$  tel que  $\pi_j \neq p_j$  et donc par la LGN forte (Th. 8.1.3)  $\liminf_{n \rightarrow +\infty} (\hat{p}_j - \pi_j)^2 > 0$  ps. On a alors  $\liminf_{n \rightarrow +\infty} T_n \geq \liminf_{n \rightarrow +\infty} n \frac{(\hat{p}_j - \pi_j)^2}{\pi_j} = +\infty$  ps, ce qui prouve le résultat sous l'hypothèse  $(A)$ .

On suppose maintenant l'hypothèse  $(H)$  en vigueur. On voit que  $N$  s'écrit  $N = \sum_{i=1}^n Z_i$  où

$$Z_i = \begin{pmatrix} Z_{i,1} \\ \vdots \\ Z_{i,r} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{\{X_i=a_1\}} \\ \vdots \\ \mathbf{1}_{\{X_i=a_r\}} \end{pmatrix}$$

sont des vecteurs *iid*. Sous  $(H)$ , l'espérance commune des  $Z_i$  est le vecteur  $\pi$  et la matrice de covariance de leur loi est  $\Gamma$  donnée par

$$\Gamma_{j,k} = \text{Cov}(Z_1)_{j,k} = \text{Cov}(Z_{1,j}, Z_{1,k}) = \delta_{j,k}\pi_j - \pi_j\pi_k = \begin{cases} -\pi_j\pi_k & \text{si } j \neq k \\ \pi_j(1 - \pi_j) & \text{si } j = k. \end{cases}$$

Le TCL multidimensionnel (Théorème 9.3.1) donne

161

$$\frac{1}{\sqrt{n}}(N - n\pi) \implies X \sim \mathcal{N}_r(0, \Gamma), \quad n \rightarrow +\infty$$

et, en notant  $D = \text{diag}(1/\sqrt{\pi_1}, \dots, 1/\sqrt{\pi_r})$ , par le continuous mapping theorem (Th. 7.4.5), on a

$$D \frac{1}{\sqrt{n}}(N - n\pi) \implies_{n \rightarrow +\infty} DX.$$

Mais

$$D \frac{1}{\sqrt{n}}(N - n\pi) = \begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix}$$

et, par la Proposition 9.2.3,  $DX \sim \mathcal{N}_r(0, D\Gamma D^t) = \mathcal{N}_r(0, D\Gamma D)$  où la matrice

$$\Sigma = D\Gamma D = (\delta_{j,k} - \sqrt{\pi_j \pi_k})_{j,k} = I_r - (\sqrt{\pi})(\sqrt{\pi})^t \quad \text{où} \quad (\sqrt{\pi}) = \begin{pmatrix} \sqrt{\pi_1} \\ \vdots \\ \sqrt{\pi_r} \end{pmatrix}.$$

On peut constater que  $\|\sqrt{\pi}\|^2 = 1$  et

$$\Sigma^2 = I_r - (\sqrt{\pi})(\sqrt{\pi})^t - (\sqrt{\pi})(\sqrt{\pi})^t + (\sqrt{\pi})(\sqrt{\pi})^t(\sqrt{\pi})(\sqrt{\pi})^t = I_r - (\sqrt{\pi})(\sqrt{\pi})^t = \Sigma$$

et  $\Sigma^t = \Sigma$ . La matrice  $\Sigma$  est la matrice de la projection orthogonale sur  $V = (\text{Vect } \sqrt{\pi})^\perp$ . En effet, on a  $\text{Im } \Sigma \perp \sqrt{\pi}$  car pour  $x \in \mathbb{R}^r$  :

$$\langle \Sigma x, \sqrt{\pi} \rangle = \langle x - (\sqrt{\pi})(\sqrt{\pi})^t x, \sqrt{\pi} \rangle = \langle x, \sqrt{\pi} \rangle - \langle (\sqrt{\pi})(\sqrt{\pi})^t x, \sqrt{\pi} \rangle$$

avec

$$\langle (\sqrt{\pi})(\sqrt{\pi})^t x, \sqrt{\pi} \rangle = \langle x, (\sqrt{\pi})(\sqrt{\pi})^t(\sqrt{\pi}) \rangle = \langle x, (\sqrt{\pi})I_r \rangle = \langle x, \sqrt{\pi} \rangle$$

soit  $\langle \Sigma x, \sqrt{\pi} \rangle = 0$  et  $\text{Ker } \Sigma = \text{Vect}(\sqrt{\pi})$  puisque  $\Sigma x = 0$  si et seulement si  $x = (\sqrt{\pi})(\sqrt{\pi})^t x \in \text{Vect } (\sqrt{\pi})$  car  $(\sqrt{\pi})^t x \in \mathbb{R}$ .

De plus si  $Y \sim \mathcal{N}_r(0, I_r)$  alors  $\Sigma Y \sim \mathcal{N}_r(0, \Sigma \Sigma^t) = \mathcal{N}_r(0, \Sigma) \sim DX$  (Prop. 9.2.3).

Ainsi, sous (H), lorsque  $n \rightarrow +\infty$ ,

$$\begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix} \implies \Sigma Y. \tag{9.7}$$

Mais

$$T_n = n\rho(\hat{p}, \pi) = \left\| \begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix} \right\|^2$$

et par le théorème de Cochran (Th. 9.3.3)  $\|\Sigma Y\|^2 = \|\Pi_V(Y)\|^2 \sim \chi^2(r - 1)$  car  $\dim(V) = r - 1$ . Finalement, (9.7), combiné avec le continuous mapping theorem (Th. 7.4.5) pour l'application continue  $x \in \mathbb{R}^r \mapsto \|x\|^2$ , donne

$$n\rho(\hat{p}, \pi) \implies \chi^2(r - 1)$$

ce qui prouve le test d'adéquation du  $\chi^2$ . □

Le test du  $\chi^2$  se décline sous d'autres formes : test d'homogénéité (pour tester que deux échantillons sont gouvernés par la même loi), test d'indépendance (pour tester que deux quantités issues d'un même échantillon sont indépendantes). Pour des détails et des exemples simples, on renvoie à [Dress] ou à tout cours de statistique de base.

# Bibliographie

- [BL] Philippe Barbé, Michel Ledoux. *Probabilité*. EDP science, 2007.
- [Bil1] Patrick Billingsley. *Probability and measures*. 3rd Edition, Wiley series in probabilities and mathematical statistics, 1995.
- [Bil2] Patrick Billingsley. *Convergence of Probability measures*. 2nd Edition, Wiley series in probabilities and mathematical statistics, 1999.
- [JCB-Leb] Jean-Christophe Breton. *Intégration et mesures*. [Notes de cours de L3 Mathématiques](#), Université de La Rochelle, 2009.
- [Chung] Kai Lai Chung. *A course in probability theory*. 3rd Edition, Academic Press, 2001.
- [Dhar] S. W. Dharmadhikari. *A simple proof of mean convergence in the law of large numbers*. The American Mathematical Monthly, Vol. 83, no. 6. pp. 474–475, 1976.
- [Dur] Rick Durrett. *Probability : theory and examples*. 4th Edition, Cambridge series in statistical and probabilistic Mathematics, 2010.
- [Dress] François Dress. *Probabilité et statistiques*. 2ème édition, Dunod, 1997.
- [Fel] William Feller. *An introduction to probability theory and its applications*. Vol. 1. 3rd Edition, Wiley series in probabilities and mathematical statistics, 1968.
- [FF] Dominique Foata, Aimé Fuchs. *Calcul des probabilités*. 2ème édition, Dunod, 2003.
- [Gra] Mihai Gradinaru. *Probabilités de base*. [Notes de cours de magistère](#), Université de Rennes 1, 2011.
- [Kal] Olav Kallenberg. *Foundations of modern probability*. 2nd Edition, Springer series in statistics. Probability and its applications, 2002.
- [Neveu] Jacques Neveu. *Manuel de probabilités de l'école Polytechnique*. Département de mathématiques appliquées, École Polytechnique, 1996.
- [Ouv] Jean-Yves Ouvrard. *Probabilités*. Tomes 1 et 2. Cassini, 2008.
- [Rud] Walter Rudin. *Analyse réelle et complexe*. 3ème édition, Dunod, 1998.
- [Suq] Charles Suquet. *Intégration, Fourier, Probabilités*. [Note de cours de L3](#), Université Lille 1, 2004.