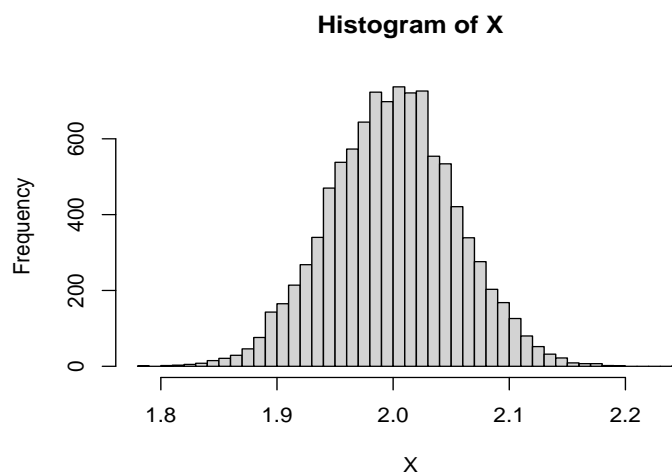


*Université Paris I, Panthéon - Sorbonne*

LICENCE M.I.A.S.H.S. TROISIÈME ANNÉE

## Cours de Statistique 2

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



## Plan du cours

### Introduction

1. Variables aléatoires et espérance
2. Vecteurs aléatoires et indépendance
3. Vecteurs gaussiens
4. Convergence et théorèmes limite
5. Estimation paramétrique
6. Tests paramétriques et non paramétriques

## References

- [1] Barbe et Ledoux, *Probabilités*, Belin.
- [2] Dacunha-Castelle et Duflo, *Probabilités et Statistiques (I)*, Masson
- [3] Dauxois, J. et Hassenforder, C. (2004). Toutes les probabilités et Statistiques. Cours et Exercices corrigés. Ellipses.
- [4] Garet, O. et Kuntzmann, A., *De l'intégration aux probabilités*, Ellipses.
- [5] Leboeuf, C., Guegand, J., Roque, J.L. et Landry, P. Cours de Probabilités et de statistiques, Ellipses
- [6] Leboeuf, C., Guegand, J., Roque, J.L. et Landry, P. Exercices corrigés de probabilités, Ellipses
- [7] Ross, S.M (2007). Initiation aux probabilités, Enseignement des Mathématiques. Presses polytechniques et universitaires romandes.
- [8] Saporta, G. Probabilités, analyse des données et statistique (2nd édition), éditions Technip.

## Introduction

Il demeure des choses inconnues à partir des connaissances antérieures en probabilités :

- Qu'est-ce qu'un événement et l'ensemble de tous les événements ?
- Que se passe-t-il pour des probabilités d'événements moins classiques (par exemple l'ensemble des décimaux) ?
- Comment traiter une variable aléatoire qui est continue et discrète à la fois (par exemple le nombre de minutes passées devant la TV) ?

## Rappels: Mesures

### Tribus

**Notation.** •  $\Omega$  est un ensemble (fini ou infini).

- $\mathcal{P}(\Omega)$  est l'ensemble de tous les sous-ensembles (parties) de  $\Omega$ .

**Rappel.** Soit  $E$  un ensemble.  $E$  est dit dénombrable s'il existe une bijection entre  $E$  et  $\mathbb{N}$  ou un sous-ensemble de  $\mathbb{N}$ . Par exemple, un ensemble fini,  $\mathbb{Z}$ ,  $\mathbb{D}$ ,  $\mathbb{Z} \times \mathbb{Z}$ ,  $\mathbb{Q}$  sont dénombrables. En revanche,  $\mathbb{R}$  n'est pas dénombrable.

**Définition.** Soit une famille  $\mathcal{F}$  de parties de  $\Omega$  (donc  $\mathcal{F} \subset \mathcal{P}(\Omega)$ ). On dit que  $\mathcal{F}$  est une algèbre si:

- $\Omega \in \mathcal{F}$ ;
- lorsque  $A \in \mathcal{F}$  alors  $(\Omega \setminus A) \in \mathcal{F}$ ;
- pour tout  $n \in \mathbb{N}^*$ , lorsque  $(A_1, \dots, A_n) \in \mathcal{F}^n$  alors  $A_1 \cup \dots \cup A_n \in \mathcal{F}$ .

**Définition.** Soit une famille  $\mathcal{A}$  de parties de  $\Omega$  (donc  $\mathcal{A} \subset \mathcal{P}(\Omega)$ ). On dit que  $\mathcal{A}$  est une tribu (ou  $\sigma$ -algèbre) sur  $\Omega$  si :

- $\Omega \in \mathcal{A}$ ;
- lorsque  $A \in \mathcal{A}$  alors  $(\Omega \setminus A) \in \mathcal{A}$ ;
- pour  $I \subset \mathbb{N}$ , lorsque  $(A_i)_{i \in I} \in \mathcal{A}^I$  alors  $\bigcup_{i \in I} A_i \in \mathcal{A}$ .

**Exemple.**

- Cas du Pile ou Face.
- Cas où  $\Omega$  est infini :  $\Omega = \mathbb{N}$  par exemple.

**Propriété.** Avec les notations précédentes :

1.  $\emptyset \in \mathcal{A}$ ;
2. si  $A$  et  $B$  sont dans la tribu  $\mathcal{A}$ , alors  $A \cap B$  est dans  $\mathcal{A}$ ;
3. si  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont deux tribus sur  $\Omega$ , alors  $\mathcal{A}_1 \cap \mathcal{A}_2$  est une tribu sur  $\Omega$ . Plus généralement, pour  $I \subset \mathbb{N}$ , si  $(\mathcal{A}_i)_{i \in I}$  ensemble de tribus sur  $\Omega$ , alors  $\bigcap_{i \in I} \mathcal{A}_i$  est une tribu sur  $\Omega$ ;

4. si  $\mathcal{A}_1$  et  $\mathcal{A}_2$  sont deux tribus sur  $\Omega$ , alors  $\mathcal{A}_1 \cup \mathcal{A}_2$  n'est pas forcément une tribu sur  $\Omega$ .

**Définition.** Si  $\mathcal{E}$  est une famille de parties de  $\Omega$  (donc  $\mathcal{E} \subset \mathcal{P}(\Omega)$ ), alors on appelle tribu engendrée par  $\mathcal{E}$ , notée  $\sigma(\mathcal{E})$ , la tribu engendrée par l'intersection de toutes les tribus contenant  $\mathcal{E}$  (on peut faire la même chose avec des algèbres).

**Remarque.**

La tribu engendrée est la "plus petite" tribu (au sens de l'inclusion) contenant la famille  $\mathcal{E}$ .

**Rappel.** • Un ensemble ouvert  $U$  dans un espace métrique  $X$  est telle que pour tout  $x \in U$ , il existe  $r > 0$  tel que  $B(x, r) \subset U$ .

- On dit qu'un ensemble dans un espace métrique  $X$  est fermé si son complémentaire dans  $X$  est ouvert.

**Définition.** Soit  $\Omega$  un espace métrique. On appelle tribu borélienne sur  $\Omega$ , notée,  $\mathcal{B}(\Omega)$ , la tribu engendrée par les ouverts de  $\Omega$ . Un ensemble de  $\mathcal{B}(\Omega)$  est appelé borélien.

**Exemple.**

- Boréliens sur  $\mathbb{R}$ , sur  $]0, 1[$ .
- Boréliens sur  $\mathbb{R}^2$ .

## Espace mesurable

**Définition.** Soit  $\Omega$  un ensemble et soit  $\mathcal{A}$  une tribu sur  $\Omega$ . On dit que  $(\Omega, \mathcal{A})$  est un espace mesurable.

**Corollaire.** Quand on s'intéressera aux probabilités, on dira que  $(\Omega, \mathcal{A})$  est un espace probabilisable.

**Propriété.** Si  $(\Omega_i, \mathcal{A}_i)_i$  sont  $n$  espaces mesurables, alors un ensemble élémentaire de  $\Omega = \Omega_1 \times \cdots \times \Omega_n$  est une réunion finie d'ensembles  $A_1 \times \cdots \times A_n$  où chaque  $A_i \in \mathcal{A}_i$ . L'ensemble des ensembles élémentaires est une algèbre et on note  $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$  (on dit  $\mathcal{A}_1$  tensoriel  $\mathcal{A}_2$  ... tensoriel  $\mathcal{A}_n$ ) la tribu sur  $\Omega$  engendrée par ces ensembles élémentaires.

**Exemple.**

Pavés de  $\mathbb{R}^d$ .

**Définition.** On appelle espace mesurable produit des  $(\Omega_i, \mathcal{A}_i)_i$  l'espace mesurable  $\left( \prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \right)$ .

**Exemple.**

Pile / Face 2 fois.

## Définitions et Propriétés d'une mesure

**Définition.** Soit  $(\Omega, \mathcal{A})$  un espace mesurable. L'application  $\mu : \mathcal{A} \rightarrow [0, +\infty]$  est une mesure si :

- $\mu(\emptyset) = 0$ .
- Pour tout  $I \subset \mathbb{N}$  et pour  $(A_i)_{i \in I}$  famille disjointe de  $\mathcal{A}$  (telle que  $A_i \cap A_j = \emptyset$  pour  $i \neq j$ ), alors  $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$  (propriété dite de  $\sigma$ -additivité).

**Définition.** Avec les notations précédentes :

- Si  $\mu(\Omega) < +\infty$ , on dit que  $\mu$  est finie.
- Si  $\mu(\Omega) < M$  avec  $M < +\infty$ , on dit que  $\mu$  est bornée.
- Si  $\mu(\Omega) = 1$ , on dit que  $\mu$  est une mesure de probabilité.

**Exemple.**

Cas de  $\Omega = \mathbb{R}$ , de  $\mathbb{N}$ , ou  $\mathbb{R}^2$ .

**Définition.** Si  $(\Omega, \mathcal{A})$  est un espace mesurable (resp. probabilisable) alors  $(\Omega, \mathcal{A}, \mu)$  est un espace mesuré (resp. probabilisé quand  $\mu$  est une probabilité).

**Remarque.**

Sur  $(\Omega, \mathcal{A})$ , on peut définir une infinité de mesures.

**Propriété.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré et  $(A_i)_{i \in \mathbb{N}}$ , une famille de  $\mathcal{A}$ .

1. Si  $A_1 \subset A_2$ , alors  $\mu(A_1) \leq \mu(A_2)$ .
2. Si  $\mu(A_1) < +\infty$  et  $\mu(A_2) < +\infty$ , alors  $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$ .
3. Pour tout  $I \subset \mathbb{N}$ , on a  $\mu\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mu(A_i)$ .
4. Si  $A_i \subset A_{i+1}$  pour tout  $i \in \mathbb{N}$  (suite croissante en sens de l'inclusion), alors  $(\mu(A_n))_{n \in \mathbb{N}}$  est une suite croissante convergente telle que  $\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{i \rightarrow +\infty} \mu(A_i)$  (même si cette limite est  $+\infty$ ).
5. Si  $A_{i+1} \subset A_i$  pour tout  $i \in \mathbb{N}$  (suite décroissante en sens de l'inclusion) et  $\mu(A_0) < +\infty$ , alors  $(\mu(A_n))_{n \in \mathbb{N}}$  est une suite décroissante convergente telle que  $\mu\left(\bigcap_{i \in \mathbb{N}} A_i\right) = \lim_{i \rightarrow +\infty} \mu(A_i)$ .

**Exemple.**

1. Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré. On définit  $\nu(A) = \mu(A \cap B)$  où  $B \in \mathcal{A}$ .  $\nu$  mesure ?
2. Si  $\mu_1$  et  $\mu_2$  mesures sur  $(\Omega, \mathcal{A})$ ,  $\mu_1 + \mu_2$  et  $\alpha\mu$  sont-elles des mesures ?

**Définition.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré et  $(A_i)_{i \in \mathbb{N}}$  une famille de  $\mathcal{A}$ .

1. On définit  $\limsup(A_n)_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m$  (intuitivement,  $\limsup(A_n)_n$  est l'ensemble des  $\omega \in \Omega$  tels que  $\omega$  appartienne à une infinité de  $A_n$ ).
2. On définit  $\liminf(A_n)_n = \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m$  (intuitivement,  $\liminf(A_n)_n$  est l'ensemble des  $\omega \in \Omega$  tels que  $\omega$  appartienne à tous les  $A_n$  sauf à un nombre fini d'entre eux).

**Exemple.**

Cas des suites croissantes et décroissantes d'ensembles.

**Théorème** (Théorème d'extension de Hahn - Caratheodory). *Si  $\Omega$  est un ensemble,  $\mathcal{F}$  une algèbre sur  $\Omega$ , et  $\nu$  une application de  $\mathcal{F}$  dans  $[0, +\infty]$  additive (telle que  $\nu(A \cup B) = \nu(A) + \nu(B)$  pour  $A \cup B = \emptyset$ ), alors si  $\mathcal{A}$  est la tribu engendrée par  $\mathcal{F}$ , il existe une mesure  $\widehat{\nu}$  sur la tribu  $\mathcal{A}$  qui coïncide avec  $\nu$  sur  $\mathcal{F}$  (c'est-à-dire que pour tout  $F \in \mathcal{F}$ ,  $\widehat{\nu}(F) = \nu(F)$ ). On dit que  $\widehat{\nu}$  prolonge  $\nu$  sur la tribu  $\mathcal{A}$ .*

**Exemple.**

Définition de la mesure de Lebesgue sur  $\mathbb{R}$ ,  $\mathbb{R}^n$ , ...

**Définition.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré.

1. Pour  $A \in \mathcal{A}$ , on dit que  $A$  est  $\mu$ -négligeable si  $\mu(A) = 0$ .
2. Soit une propriété  $\mathcal{P}$  dépendant des éléments  $\omega$  de  $\Omega$ . On dit que  $\mathcal{P}$  est vraie  $\mu$ -presque partout ( $\mu$ -presque sûrement sur un espace probabilisé) si l'ensemble des  $\omega$  pour laquelle elle n'est pas vérifiée est  $\mu$ -négligeable.

**Exemple.**

- Mesure de Lebesgue sur  $\mathbb{N}$  ou  $\mathbb{Q}$ .
- La propriété " la suite de fonction  $f_n(x) = x^n$  converge vers la fonction  $f(x) = 0$ " est vraie  $\lambda$ -presque partout sur  $[0, 1]$ .
- Soit  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$  et soit  $F$  la fonction définie par  $F(x) = \mu(] - \infty, x])$  pour  $x \in \mathbb{R}$ .

**Fonctions mesurables**

**Rappel.** Soit  $f : E \mapsto F$ , où  $E$  et  $F$  sont 2 espaces métriques.

- Pour  $I \subset F$ , on appelle ensemble réciproque de  $I$  par  $f$ , l'ensemble  $f^{-1}(I) = \{x \in E, f(x) \in I\}$ .
- ( $f$  continue)  $\iff$  (pour tout ouvert  $U$  de  $F$  alors  $f^{-1}(U)$  est un ouvert de  $E$ ).

**Définition.** Soit  $f : E \mapsto F$  et soit  $\mathcal{I}$  une tribu sur  $F$ . On note  $f^{-1}(\mathcal{I})$  l'ensemble de sous-ensembles de  $E$  tel que  $f^{-1}(\mathcal{I}) = \{f^{-1}(I), I \in \mathcal{I}\}$ .

**Propriété.** Soit  $(\Omega', \mathcal{A}')$  un espace mesurable et soit  $f : \Omega \mapsto \Omega'$ . Alors  $f^{-1}(\mathcal{A}')$  est une tribu sur  $\Omega$  appelée tribu engendrée par  $f$ .

**Définition.** Soit  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  deux espaces mesurables. Une fonction  $f : \Omega \mapsto \Omega'$  est dite mesurable pour les tribus  $\mathcal{A}$  et  $\mathcal{A}'$  si et seulement si  $f^{-1}(A') \in \mathcal{A}$  (donc si et seulement si  $\forall A' \in \mathcal{A}'$ , alors  $f^{-1}(A') \in \mathcal{A}$ ).

**Exemple.**

- Fonction indicatrice.
- Combinaison linéaire de fonctions indicatrices.

**Remarque.**

Dans le cas où  $(\Omega, \mathcal{A})$  est un espace probablisable, et si  $f : \Omega \mapsto \mathbb{R}$ , alors si  $f$  est une fonction mesurable sur  $\mathcal{A}$  et  $\mathcal{B}(\mathbb{R})$ , alors  $f$  est une variable aléatoire.

**Exemple.**

Nombre de Piles dans un jeu de Pile/Face.

**Remarque.**

Dans le cas où  $(\Omega, \mathcal{A})$  est un espace mesurable, et si  $f : \Omega \mapsto (\Omega', \mathcal{B}(\Omega'))$ , où  $\Omega'$  est un espace métrique et  $\mathcal{B}(\Omega')$  l'ensemble des boréliens de  $\Omega'$ , si  $f$  est une fonction mesurable sur  $\mathcal{A}$  et  $\mathcal{B}(\Omega')$ , alors  $f$  est dite fonction borélienne.

**Proposition.** Soit  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  deux espaces mesurables et  $f : \Omega \mapsto \Omega'$ . Soit  $\mathcal{F}$  une famille de sous-ensembles de  $\Omega'$  telle que  $\sigma(\mathcal{F}) = \mathcal{A}'$ . Alors

1.  $f^{-1}(\mathcal{F})$  engendre la tribu  $f^{-1}(\mathcal{A}')$ .
2.  $(f \text{ mesurable}) \iff (f^{-1}(\mathcal{F}) \subset \mathcal{A})$

**Conséquence.** • Si  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  sont deux espaces mesurables boréliens, alors toute application continue de  $\Omega \mapsto \Omega'$  est mesurable.

- Pour montrer qu'une fonction  $f : \Omega \mapsto \mathbb{R}$  est mesurable, il suffit de montrer que la famille d'ensemble  $(\{\omega \in \Omega, f(\omega) \leq a\})_{a \in \mathbb{R}} \in \mathcal{A}$ .

**Propriété.** • Soit  $f$  mesurable de  $(\Omega, \mathcal{A})$  dans  $(\Omega', \mathcal{A}')$  et  $g$  mesurable de  $(\Omega', \mathcal{A}')$  dans  $(\Omega'', \mathcal{A}'')$ . Alors  $g \circ f$  est mesurable dans  $\mathcal{A}$  et  $\mathcal{A}'$ .

- Soit  $f_1$  mesurable de  $(\Omega, \mathcal{A})$  dans  $(\Omega_1, \mathcal{A}_1)$  et  $f_2$  mesurable de  $(\Omega, \mathcal{A})$  dans  $(\Omega_2, \mathcal{A}_2)$ . Alors  $h : \Omega \mapsto \Omega_1 \times \Omega_2$  telle que  $h(\omega) = (f_1(\omega), f_2(\omega))$  est mesurable dans  $\mathcal{A}$  et  $\mathcal{A}_1 \otimes \mathcal{A}_2$ .
- Soit  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions mesurables de  $(\Omega, \mathcal{A})$  dans  $(\Omega', \mathcal{B}(\Omega'))$ , où  $\Omega'$  est un espace métrique, telle qu'il existe une fonction  $f$  limite simple de  $(f_n)$  (donc  $\forall \omega \in \Omega, \lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$ ). Alors  $f$  est mesurable dans  $\mathcal{A}$  et  $\mathcal{B}(\Omega')$ .

**Définition.** Soit  $f$  mesurable de  $(\Omega, \mathcal{A}, \mu)$  dans  $(\Omega', \mathcal{A}')$  et soit  $\mu_f : \mathcal{A}' \mapsto [0, +\infty]$  telle que pour tout  $A' \in \mathcal{A}'$ , on ait  $\mu_f(A') = \mu(f^{-1}(A'))$ . Alors  $\mu_f$  est une mesure sur  $(\Omega', \mathcal{A}')$  appelée mesure image de  $\mu$  par  $f$ .

**Cas particulier.**

Si  $\mu$  est une mesure de probabilité et si  $X$  est une variable aléatoire alors  $\mu_X$  est la mesure (loi) de probabilité de la variable aléatoire  $X$ .

## Cas des fonctions réelles mesurables

**Propriété.** Soit  $f$  et  $g$  deux fonctions réelles mesurables (de  $(\Omega, \mathcal{A}, \mu)$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ). Alors  $\alpha.f$ ,  $f + g$ ,  $\min(f, g)$  et  $\max(f, g)$  sont des fonctions réelles mesurables.

**Propriété.** Soit  $(f_n)_{n \in \mathbb{N}}$  une suite de fonctions réelles mesurables. Alors  $\inf(f_n)$  et  $\sup(f_n)$  sont des fonctions réelles mesurables.

**Définition.** Soit  $f : \Omega \rightarrow \mathbb{R}$ . Alors  $f$  est dite étagée s'il existe une famille d'ensembles disjoints  $(A_i)_{1 \leq i \leq n}$  de  $\Omega$  et une famille de réels  $(\alpha_i)_{1 \leq i \leq n}$  telles que pour tout  $\omega \in \Omega$ , on ait  $f(\omega) = \sum_{i=1}^n \alpha_i \mathbb{I}_{A_i}(\omega)$ .

### Remarque.

Si les  $A_i$  sont tous dans  $\mathcal{A}$  tribu sur  $\Omega$ , alors  $f$  est  $\mathcal{A}$ -mesurable.

**Théorème.** Toute fonction réelle mesurable à valeurs dans  $[0, +\infty]$  est limite simple d'une suite croissante de fonctions étagées.

**Conséquence.** Soit  $f$  une fonction réelle mesurable. Alors  $f$  est limite simple de fonctions étagées.

## Intégration de Lebesgue

Dans toute la suite, on considère  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré.

### Intégrale de Lebesgue d'une fonction positive

**Définition.** 1. Soit  $f = \mathbb{I}_A$ , où  $A \in \mathcal{A}$ . Alors :

$$\int f d\mu = \int_{\omega} f(\omega) d\mu(\omega) = \mu(A).$$

2. Soit  $f = \mathbb{I}_A$ , où  $A \in \mathcal{A}$  et soit  $B \in \mathcal{A}$ . Alors :

$$\int_B f d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B \mu(A)(\omega) f(\omega) d\mu(\omega) = \mu(A \cap B).$$

3. Soit  $f$  une fonction étagée positive telle que  $f = \sum_{i=1}^n \alpha_i \mathbb{I}_{A_i}$ , où les  $A_i \in \mathcal{A}$  et  $\alpha_i > 0$  et soit  $B \in \mathcal{A}$ . Alors :

$$\int_B f d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sum_{i=1}^n \alpha_i \mu(A_i \cap B).$$

### Exemple.

Fonction  $\mathbb{I}_{\mathbb{Q}}$ , fonctions en escalier,...

**Définition.** Soit  $f$  une fonction  $\mathcal{A}$ -mesurable positive et soit  $B \in \mathcal{A}$ . Alors l'intégrale de Lebesgue de  $f$  par rapport à  $\mu$  sur  $B$  est :

$$\int_B f d\mu = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sup \left\{ \int_B g d\mu, \text{ pour } g \text{ étagée positive telle que } g \leq f \right\}.$$

**Propriété.** Soit  $f$  une fonction  $\mathcal{A}$ -mesurable positive et soit  $A$  et  $B \in \mathcal{A}$ . Alors :

1. Pour  $c \geq 0$ ,  $\int_B cf \, d\mu = c \int_B f \, d\mu$ .
2. Si  $A \subset B$ , alors  $\int_A f \, d\mu \leq \int_B f \, d\mu$ .
3. Si  $g$  est une fonction  $\mathcal{A}$ -mesurable positive telle que  $0 \leq f \leq g$  alors  $0 \leq \int_B f \, d\mu \leq \int_B g \, d\mu$ .
4. Si  $\mu(B) = 0$  alors  $\int_B f \, d\mu = 0$ .

**Théorème** (Théorème de convergence monotone (Beppo-Lévi)). Si  $(f_n)_n$  est une suite croissante de fonctions mesurables positives convergeant simplement vers  $f$  sur  $\Omega$ , alors :

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

**Conséquence.** Pour les séries de fonctions mesurables positives, on peut toujours appliquer le Théorème de convergence monotone et donc inverser la somme et l'intégrale.

**Lemme** (Lemme de Fatou). Soit  $(f_n)_n$  est une suite de fonctions mesurables positives alors :

$$\int \left( \liminf_{n \rightarrow \infty} f_n \right) \, d\mu \leq \liminf_{n \rightarrow \infty} \int f_n \, d\mu.$$

**Exemple.**

Appliquer Fatou à  $(f_n)$  telle que  $f_{2n} = \mathbb{I}_A$  et  $f_{2n+1} = \mathbb{I}_B$ .

### Intégrale de Lebesgue d'une fonction réelle et propriétés

**Définition.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré,  $B \in \mathcal{A}$  et soit  $f$  une fonction  $\mathcal{A}$ -mesurable à valeurs réelles telle que  $f = f^+ - f^-$  avec  $f^+ = \max(f, 0)$  et  $f^- = \max(-f, 0)$ . On dit que  $f$  est  $\mu$ -intégrable sur  $B$  si  $\int_B |f| \, d\mu < +\infty$ . On a alors

$$\int_B f \, d\mu = \int_B f^+ \, d\mu - \int_B f^- \, d\mu.$$

**Notation.** Lorsque  $f$  est  $\mu$ -intégrable sur  $B$ , soit  $\int |f| \, d\mu < +\infty$ , on note  $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$  (on dit que  $f$  est  $\mathcal{L}^1$ ).

**Exemple.**

Intégrale de Riemann et intégrale de Lebesgue.

Cas de la masse de Dirac.

**Propriété.** On suppose que  $f$  et  $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ . Alors :

1.  $\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu$  pour  $(\alpha, \beta) \in \mathbb{R}^2$ .
2. Si  $f \leq g$  alors  $\int f \, d\mu \leq \int g \, d\mu$ .

**Théorème** (Théorème de convergence dominée de Lebesgue). Soit  $(f_n)_n$  est une suite de fonctions de  $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$  telles que pour tout  $n \in \mathbb{N}$ ,  $|f_n| \leq g$  avec  $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ . Si on suppose que  $(f_n)$  converge simplement vers  $f$  sur  $\Omega$  alors :

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

### Extension.

Le Théorème de Lebesgue s'applique également dans le cas où  $(f_n)_n$  converge presque partout vers  $f$ .

### Exemple.

Convergence d'intégrale dépendant d'un paramètre : par exemple  $\int_0^\infty \frac{f(x)}{1+x^n} dx$ .

**Théorème** (Inégalité de Jensen). Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé, soit  $\phi : \mathbb{R} \mapsto \mathbb{R}$  une fonction convexe et soit  $f : \Omega \mapsto \mathbb{R}$  mesurable telle que  $\phi(f)$  soit une fonction intégrable par rapport à  $P$ . Alors :

$$\phi\left(\int f dP\right) \leq \int \phi(f) dP.$$

### Exemple.

Soit  $X$  une v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors  $\phi(\mathbb{E}[X]) \leq \mathbb{E}(\phi(X))$ .

## Mesures induites et densités

**Théorème** (Théorème du Transport). Soit  $f$  une fonction mesurable de  $(\Omega, \mathcal{A}, \mu)$  dans  $(\Omega', \mathcal{A}')$  telle que  $\mu_f$  soit la mesure induite par  $f$  (donc  $\mu_f(A') = \mu(f^{-1}(A'))$  pour  $A' \in \mathcal{A}'$ ) et soit  $\phi$  une fonction mesurable de  $(\Omega', \mathcal{A}')$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Alors, si  $\phi \circ f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ ,

$$\int_{\Omega'} \phi d\mu_f = \int_{\Omega} \phi \circ f d\mu.$$

**Définition.** Soit  $\mu$  et  $\nu$  deux mesures sur  $(\Omega, \mathcal{A})$ . On dit que  $\mu$  domine  $\nu$  (ou  $\nu$  est dominée par  $\mu$ ) et que  $\nu$  est absolument continue par rapport à  $\mu$  lorsque pour tout  $A \in \mathcal{A}$ ,  $\mu(A) = 0 \implies \nu(A) = 0$ .

**Propriété.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré et  $f$  une fonction définie sur  $(\Omega, \mathcal{A})$  mesurable et positive. On suppose que pour  $A \in \mathcal{A}$ ,  $\nu(A) = \int_A f d\mu$ . Alors,  $\nu$  est une mesure sur  $(\Omega, \mathcal{A})$ , dominée par  $\mu$ . De plus, pour toute fonction  $g$  définie sur  $(\Omega, \mathcal{A})$  mesurable et positive,

$$\int g d\nu = \int g \cdot f d\mu.$$

Enfin,  $g$  est  $\nu$  intégrable si et seulement si  $g \cdot f$  est  $\mu$  intégrable.

**Définition.** On dit que  $\mu$  mesure sur  $(\Omega, \mathcal{A})$  est  $\sigma$ -finie lorsqu'il existe une famille  $(A_i)_{i \in I}$ , avec  $I$  dénombrable, d'ensembles de  $\mathcal{A}$  telle que  $\bigcup A_i = \Omega$  et  $\mu(A_i) < +\infty$  pour tout  $i \in I$ .

**Théorème** (Théorème de Radon-Nikodym). On suppose que  $\mu$  et  $\nu$  sont deux mesures  $\sigma$ -finies sur  $(\Omega, \mathcal{A})$  telles que  $\mu$  domine  $\nu$ . Alors il existe une fonction  $f$  définie sur  $(\Omega, \mathcal{A})$  mesurable et positive, appelée densité de  $\nu$  par rapport à  $\mu$ , telle que pour tout  $A \in \mathcal{A}$ ,  $\nu(A) = \int_A f d\mu$ .

**Théorème** (Théorème de Fubini). Soit  $\Omega = \Omega_1 \times \Omega_2$ ,  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$  et  $\mu = \mu_1 \otimes \mu_2$  (mesures  $\sigma$  finies), où  $(\Omega_1, \mathcal{A}_1, \mu_1)$  et  $(\Omega_2, \mathcal{A}_2, \mu_2)$  sont des espaces mesurés. Soit une fonction  $f : \Omega \mapsto \mathbb{R}$ ,  $\mathcal{A}$ -mesurable et  $\mu$ -intégrable. alors :

$$\int_{\Omega} f d\mu = \int_{\Omega_1} \left( \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left( \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).$$

## Espaces $\mathcal{L}^p$

**Définition.** Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré. On appelle espace  $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ , où  $p > 0$ , l'ensemble des fonctions  $f : \Omega \mapsto \mathbb{R}$ , mesurables et telles que  $\int |f|^p d\mu < +\infty$ .

**Définition.** Pour  $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ , où  $p > 0$ , on note  $\|f\|_p = \left( \int |f|^p d\mu \right)^{1/p}$ .

**Propriété** (Inégalité de Hölder). Soit  $p > 1$  et  $q > 1$  tels que  $\frac{1}{p} + \frac{1}{q} = 1$ , et soit  $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$  et  $g \in \mathcal{L}^q(\Omega, \mathcal{A}, \mu)$ . Alors,  $fg \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$  et

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

**Propriété** (Inégalité de Minkowski). Soit  $p > 1$  et soit  $f$  et  $g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ . Alors,  $f + g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$  et

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

### Remarque.

Pour  $p > 1$ ,  $\|\cdot\|_p$  définie ainsi sur une semi-norme sur  $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ . Pour obtenir une norme, il faut se placer dans l'espace  $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$  obtenu en "quotientant"  $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$  par la relation d'équivalence  $f = g$   $\mu$ -presque partout (c'est-à-dire que dans  $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$  on dira que  $f = g$  lorsque  $f = g$   $\mu$ -presque partout).

**Définition.** Pour  $f$  et  $g \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ , on définit le produit scalaire  $\langle f, g \rangle = \int f \cdot g d\mu$ . On muni ainsi  $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$  d'une structure d'espace de Hilbert. On dira que  $f$  est orthogonale à  $g$  lorsque  $\langle f, g \rangle = 0$ .

**Conséquence.** Si  $A$  est un sous-espace vectoriel fermé de  $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$  (par exemple un sous-espace de dimension finie), alors pour tout  $f \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ , il existe un unique projeté orthogonal de  $f$  sur  $A$ , noté  $f_A$ , qui vérifie  $f_A = \underset{g \in A}{\text{Arginf}} \|g - f\|_2$ .

## 1 Variables aléatoires et espérance

### 1.1 Variables aléatoires

**Définition.** On dit que  $X$  est une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité si:  $X : \Omega \rightarrow \mathbb{R}$  est une application mesurable de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Pourquoi demander à  $X$  d'être mesurable? Parce que l'on veut pouvoir définir une fonction de répartition pour  $X$ , soit  $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega, X(\omega) \leq x\}) = \mathbb{P}(X^{-1}([-\infty, x]))$ : il faut donc que l'ensemble  $X^{-1}([-\infty, x])$  soit un événement de  $\mathcal{A}$  pour tout  $x \in \mathbb{R}$ .

Dans la suite, pour une variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , on n'utilisera presque jamais la forme explicite de la fonction  $\omega \rightarrow X(\omega)$  qui restera inconnue. En revanche, on préférera travailler avec la **loi** de  $X$ .

Qu'appelle-t-on loi de  $X$ ? Il y a en réalité plusieurs moyens de la définir:

1. On définit la loi par  $F_X(x) = \mathbb{P}(X \leq x)$  pour  $x \in \mathbb{R}$ , la **fonction de répartition** de  $X$ ;
2. On définit la loi par la **mesure de probabilité  $\mathbb{P}_X$  induite par  $X$** : pour  $B \in \mathcal{B}(\mathbb{R})$ ,  $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B))$ . On notera que  $\mathbb{P}_X$  est une mesure de probabilité définie sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  ou bien sur  $(X(\Omega), \mathcal{T})$ , avec  $\mathcal{T}$  une tribu sur  $X(\Omega) \subset \mathbb{R}$ .

3. Dans 2 cas particuliers, on préférera travailler avec:

- Si  $X(\Omega) = \{x_i\}_{i \in I}$  avec  $I \subset \mathbb{N}$  et  $x_i \in \mathbb{R}$  for all  $i \in I$ , on parlera de variable aléatoire **discrète** et la loi sera donnée par les  $\mathbb{P}(X = x_i) = \mathbb{P}_X(\{x_i\})$  pour  $i \in I$ . On remarque alors que pour tout  $B \in \mathcal{B}(\mathbb{R})$ ,  $\mathbb{P}_X(B) = \sum_{i \in I} \mathbb{P}(X = x_i) \delta_{\{x_i\}}(B)$ , avec  $\delta$  masse (mesure) de Dirac.
- Si  $X(\Omega)$  est une union finie ou dénombrable d'intervalles de  $I$ , et si  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , alors on peut définir la **densité de probabilité**  $f_X$  de  $X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , et l'on a:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{pour tout } x \in \mathbb{R}.$$

La fonction  $F_X$  est une fonction **absolument continue** sur  $\mathbb{R}$ , elle est même dérivable  $\lambda$ -presque partout sur  $\mathbb{R}$  et  $F'_X(x) = f_X(x)$  pour presque tout  $x$  dans  $\mathbb{R}$ . On dira souvent pour simplifier que  $X$  est une variable aléatoire "**absolument continue**" et même parfois que  $X$  est une variable aléatoire "**continue**".

4. La loi est donnée par fonction caractéristique de  $X$  soit  $\phi_X(u) = \mathbb{E}[e^{iuX}]$  pour  $u \in \mathbb{R}$  (voir plus loin).

### Remarque.

On peut connaître le "type" de variable aléatoire qu'est la variable  $X$  grâce au graphe de la fonction de répartition  $F_X$ :

1. Si  $F_X$  est une fonction en escalier avec un nombre fini ou dénombrable de sauts en les  $(x_i)_{i \in I}$  où  $I \subset \mathbb{N}$ , alors  $X$  est une variable discrète de loi  $\mathbb{P}(X = x_i) = F_X(x_i) - \lim_{x \rightarrow x_i^-} F_X(x)$ .
2. Si  $F_X$  est une fonction continue sur  $\mathbb{R}$  et dérivable sauf en un nombre fini ou dénombrable de points, alors  $X$  est une variable (absolument) continue, et la dérivée de  $F_X$  lorsqu'elle existe est la densité de probabilité  $f_X$ . Pour  $x_0$  tel que  $F_X$  n'est pas dérivable en  $x_0$ , on pourra choisir que  $f_X(x_0) = 0$  ou tout autre réel.
3. Si aucun des 2 cas précédents n'est vérifié,  $X$  pourra être un "mélange" de loi discrète et continue, soit  $\mathbb{P}_X(A) = \sum_{i \in I} p_i \delta_{x_i}(A) + \int_A g(x) d\lambda(x)$ , ou bien une loi diffuse non absolument continue par rapport à la mesure de Lebesgue comme la loi de Cantor.

Dans la propriété suivante, on va utiliser pleinement le fait qu'une variable aléatoire est une fonction mesurable:

**Propriété.** Soit  $X$  et  $Y$  deux v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  une fonction borélienne. Alors  $Z = g(X, Y)$  est une v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$ .

*Proof.* Voir cours Intégration et Probabilités: il s'agit de montrer que la composition d'une fonction mesurable par une autre fonction mesurable est une fonction mesurable. Et également que le vecteur composé de fonctions mesurables est une fonction mesurable.  $\square$

Par itération du procédé, la propriété est aussi vraie pour une fonction  $g$  à  $n$  variables. Mais aussi bien-sûr pour une fonction à 1 variable.

De la même manière, toujours du fait qu'une variable aléatoire est une fonction mesurable:

**Propriété.** Soit  $X$  et  $Y$  deux v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telles que  $\mathbb{P}_X = \mathbb{P}_Y$ , ce que l'on peut aussi noter  $X \stackrel{\mathcal{L}}{\sim} Y$ . Alors pour et  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction borélienne,  $g(X) \stackrel{\mathcal{L}}{\sim} g(Y)$ .

*Proof.* Pour tout  $B \in \mathcal{B}(\mathbb{R})$ ,  $\mathbb{P}_{g(X)}(B) = \mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$ . Mais on a aussi  $\mathbb{P}_{g(Y)}(B) = \mathbb{P}_Y(g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$  puisque  $\mathbb{P}_X = \mathbb{P}_Y$ . D'où le résultat.  $\square$

**Remarque:** Si  $X \stackrel{\mathcal{L}}{\sim} \mathcal{U}([0, 1])$  et  $Y = 1 - X$ , alors on a aussi  $Y \stackrel{\mathcal{L}}{\sim} \mathcal{U}([0, 1])$ , donc  $\mathbb{P}_X = \mathbb{P}_Y$  (on le montre en utilisant par exemple la fonction de répartition). Mais pourtant on n'a pas du tout  $X = Y \dots$

**Définition.** Si  $X$  est une v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $0 < p < 1$ , le quantile d'ordre  $p$  de  $X$  est :  $q_X(p) = \inf \{y \in \mathbb{R}, F_X(y) \geq p\}$ .

**Cas particulier:** Si  $p = 1/2$ , alors  $q_X(p)$  est la médiane (théorique) de  $X$ .

**Propriété.** Si  $X$  est une v.a. continue,  $q_X(p) = \tilde{F}_X^{-1}(p)$  où  $\tilde{F}_X : x \in X(\Omega) \mapsto F_X(x)$ .

**Exemple:** Si  $X$  suit une distribution  $\mathcal{E}(\lambda)$  (exponentielle),

$$F_X(x) = (1 - e^{-\lambda x})\mathbb{1}_{x \geq 0} \quad \text{and} \quad q_X(1/2) = \frac{\ln(2)}{\lambda} \neq \mathbb{E}[X] = \frac{1}{\lambda}.$$

## 1.2 Espérance de variables aléatoires

**Définition.** Soit  $X$  une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. Alors si  $X \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$  (donc si  $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < \infty$ ), on définit l'espérance de  $X$  par le réel:

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

Plus généralement, si  $\phi : \mathbb{R} \mapsto \mathbb{R}$  est borélienne et si  $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ , on définit l'espérance de  $\phi(X)$  par

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

**Propriété.** Si  $X$  est une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , si  $\phi : \mathbb{R} \mapsto \mathbb{R}$  est borélienne telle que  $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ , alors :

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

*Proof.* Théorème du transport...  $\square$

**Conséquence.** • Si  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue (donc

$X$  est une v.a. dite absolument continue), de densité  $f_X$ , alors  $\mathbb{E}[\phi(X)] = \int_{\mathbb{R}} \phi(x) f_X(x) dx$ .

- Si  $\mathbb{P}_X$  est absolument continue par rapport à une mesure de comptage sur  $\{x_i\}_{i \in I}$  avec  $I \subset \mathbb{N}$  (donc  $X$  est une v.a. dite discrète), de densité  $p_X$  avec  $p_X(i) = \mathbb{P}(X = x_i)$ , alors  $\mathbb{E}[\phi(X)] = \sum_{i \in I} p_X(i) \phi(x_i)$ .

**Propriété.** 1. Soit  $X$  et  $Y$  des variables aléatoires telles que  $X$  et  $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ . Alors pour tout  $(a, b) \in \mathbb{R}^2$ ,  $aX + bY \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$  et

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2. Soit  $X$  une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , et soit  $B \in \mathcal{B}(\mathbb{R})$ . Alors  $\mathbb{E}[\mathbb{1}_B(X)] = \mathbb{P}(X \in B)$ .

3. Si  $X$  est une variable aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , si  $\phi : \mathbb{R} \mapsto \mathbb{R}$  est une fonction borélienne convexe telle que  $X$  et  $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ , alors

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]) \quad (\text{Inégalité de Jensen}).$$

4. Soit  $X$  et  $Y$  des variables aléatoires telles que  $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  et  $Y \in \mathbb{L}^q(\Omega, \mathcal{A}, \mathbb{P})$  avec  $\frac{1}{p} + \frac{1}{q} = 1$  où  $p > 1, q > 1$ . Alors  $XY \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$  et

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} (\mathbb{E}[|Y|^q])^{1/q} \quad (\text{Inégalité de Hölder}).$$

5. Soit  $X$  et  $Y$  des variables aléatoires telles que  $X$  et  $Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ , avec  $p \geq 1$ . Alors  $X + Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  et

$$(\mathbb{E}[|X + Y|^p])^{1/p} \leq (\mathbb{E}[|X|^p])^{1/p} + (\mathbb{E}[|Y|^p])^{1/p} \quad (\text{Inégalité triangulaire de Minkowski}).$$

*Proof.* 1. On a d'abord  $aX + bY$  qui est une v.a., puis  $\mathbb{E}[|aX + bY|] \leq |a|\mathbb{E}[|X|] + |b|\mathbb{E}[|Y|] < \infty$ . On utilise ensuite la linéarité de l'intégrale.

2. La v.a.  $\mathbb{I}_B(X)$  est une v.a. de Bernoulli de paramètre  $\mathbb{P}(X \in B)$ , et donc également d'espérance  $\mathbb{P}(X \in B)$ .

3. Vue en "Intégration et probabilités".

4. Plus généralement, on a  $\int |fg|d\mu \leq \|f\|_p \|g\|_q$  avec  $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ . En effet, la fonction  $x \in ]0, \infty[ \mapsto -\ln(x)$  est convexe, donc pour tout  $a > 0, b > 0$  et  $\theta \in [0, 1]$ ,  $-\log(a\theta + b(1 - \theta)) \leq -\theta \log(a) - (1 - \theta) \log(b)$ . En passant à l'exponentielle, on en déduit que  $a^\theta b^{1-\theta} \leq a\theta + b(1 - \theta)$ .

Pour tout  $x \in \Omega$ , choisissons  $a = \frac{|f(x)|^p}{\|f\|_p^p}, b = \frac{|g(x)|^q}{\|g\|_q^q}$  et  $\theta = 1/p$ . Alors pour tout  $x \in \Omega$  et avec  $1 - \theta = 1/q$ ,

$$\frac{|f(x)|}{\|f\|_p} \frac{|g(x)|}{\|g\|_q} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}.$$

Intégrons des 2 côtés (on peut et on garde le sens de l'inégalité), on obtient:

$$\frac{\int |fg|d\mu}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{\int |f|^p d\mu}{\|f\|_p^p} + \frac{1}{q} \frac{\int |g|^q d\mu}{\|g\|_q^q} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

d'où le résultat.

5. On va montrer plus généralement que  $\|f + g\|_p \leq \|f\|_p + \|g\|_p$  pour  $p \geq 1$ . Il est clair que si  $\|f + g\|_p = 0$ , le résultat est vrai. De même si  $p = 1$ . Pour  $p > 1$ ,

$$\begin{aligned} \|f + g\|_p^p &\leq \int (|f| + |g|)|f + g|^{p-1} d\mu \\ &\leq \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu \\ &\leq \|f\|_p \left( \int |f + g|^{q(p-1)} d\mu \right)^{1/q} + \|g\|_p \left( \int |f + g|^{q(p-1)} d\mu \right)^{1/q} \quad (\text{Inégalité de Hölder}) \\ &\leq (\|f\|_p + \|g\|_p) \|f + g\|_p^{p-1}, \end{aligned}$$

d'où l'inégalité. □

**Conséquence.** Soit  $X$  une variable aléatoire telle que  $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  pour  $p \geq 1$ , c'est-à-dire l'ensemble des variables aléatoires définies à la classe d'équivalence  $\mathbb{P}$ -p.s. près (donc  $X \sim Y$  si  $X - Y = 0$   $\mathbb{P}$ -p.s.) et telle que  $\mathbb{E}[|X|^p] < \infty$ . Alors  $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$  est une norme sur  $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ .

*Proof.* Pour que ce soit bien une norme, il faut que:

1.  $\|X\|_p \geq 0$  pour tout  $X$ , ce qui est évidemment vrai;
2. pour tout  $\lambda \in \mathbb{R}, \|\lambda X\|_p = |\lambda| \|X\|_p$  ce qui est vrai également;
3. l'inégalité triangulaire soit vérifiée, soit  $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ , ce qui est vrai grâce à l'Inégalité de Minkowski;

4. si  $\|X\|_p = 0$ , soit  $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) = 0$  alors  $X = 0$ , ce qui est vrai si l'on considère  $X = 0$   $\mathbb{P}$ -p.s.

La quatrième propriété implique de travailler sur  $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  et non sur  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  qui est constitué des fonctions  $f$  telles que  $\|f\|_p < \infty$  sans la classe d'équivalence: sur  $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  on a seulement une quasi-norme, la propriété 4/ n'est pas vérifiée (prendre par exemple  $\Omega = [0, 1]$ ,  $\mathcal{A} = \mathcal{B}([0, 1])$ ,  $\mathbb{P} = \lambda$  et  $f(x) = 0$  pour  $x \in ]0, 1]$  et  $f(0) = 1$ ).  $\square$

**Conséquence.** Soit  $X$  une variable aléatoire telle que  $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  pour  $p > 0$ . Alors pour tout  $1 \leq r \leq p$ ,  $X \in \mathbb{L}^r(\Omega, \mathcal{A}, \mathbb{P})$  et

$$(\mathbb{E}[|X|^r])^{1/r} \leq (\mathbb{E}[|X|^p])^{1/p} \iff \|X\|_r \leq \|X\|_p.$$

*Proof.* La fonction  $\phi(x) = x^{p/r}$  est convexe sur  $[0, \infty[$  car  $p/r \geq 1$  (il suffit de dériver 2 fois et  $\phi'' \geq 0$ ). On applique alors l'inégalité de Jensen à la variable  $|X|^r$ :  $\mathbb{E}[\phi(|X|^r)] \geq \mathbb{E}[|X|^r]^{p/r}$ , d'où le résultat.  $\square$

**Définition.** Pour  $X$  et  $Y$  des variables aléatoires telles que  $X$  et  $Y \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , on définit:

- la variance de  $X$ ,  $\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ .
- la covariance de  $X$  et  $Y$  par

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

**Remarque:** On a  $\text{var}(X) = \text{cov}(X, X)$ .

**Propriété.** Sur l'espace vectoriel  $E = \{X \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P}), \mathbb{E}[X] = 0\}$ , on définit  $\langle X, Y \rangle = \text{cov}(X, Y)$  pour  $X, Y \in E$ . Alors  $\langle \cdot, \cdot \rangle$  définit un produit scalaire.

**Conséquence.** Pour  $X$  et  $Y$  deux v.a. sur  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , alors

$$|\text{cov}(X, Y)|^2 \leq \text{var}(X)\text{var}(Y).$$

*Proof.* Soit  $X'$  et  $Y'$  telles que  $\mathbb{E}[X'] = \mathbb{E}[Y'] = 0$ : il suffit alors d'appliquer l'inégalité de Cauchy-Schwarz (cas particulier de Hölder avec  $p = q = 2$ ):  $|\mathbb{E}[X'Y']| \leq \mathbb{E}[|X'Y'|] \leq \sqrt{\mathbb{E}[X'^2]\mathbb{E}[Y'^2]}$ . Ensuite on remplace  $X'$  par  $X - \mathbb{E}[X]$  et  $Y'$  par  $Y - \mathbb{E}[Y]$ .  $\square$

**Propriété.** Pour  $X$  et  $Y$  deux v.a. sur  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ , on peut définir le coefficient de corrélation entre  $X$  et  $Y$  par:

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \quad \text{et} \quad |\text{cor}(X, Y)| \leq 1.$$

### 1.3 Fonction génératrice et fonction caractéristique

**Définition.** Si  $X$  est une v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{N}$ , on définit la fonction génératrice de  $X$  par :

$$g(z) = \mathbb{E}[z^X] \quad \text{pour tout } z \in [-1, 1].$$

**Remarque:**  $g(z) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k$  est une série entière de rayon de convergence  $R \geq 1$ .

**Propriété.** La fonction  $g$  est une fonction  $\mathcal{C}^\infty((-1, 1))$  et  $\mathbb{P}(X = k) = g^{(k)}(0)/k!$ .

*Proof.* Fonction  $\mathcal{C}^\infty((-1, 1))$ : propriété d'une série entière. De plus, par unicité du développement en série entière on a également  $g(z) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} z^k = \sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k$ , d'où le résultat.  $\square$

**Conséquence:**  $(X \stackrel{\mathcal{L}}{\sim} Y) \iff (g_X = g_Y)$ .

**Propriété.** Si  $\mathbb{E}[|X|] < \infty$ , la fonction  $g$  est telle que  $g'(1) = \mathbb{E}[X]$ . On peut également obtenir pour tout  $k \geq 1$ ,  $\mathbb{E}[X(X-1) \times \dots \times (X-k)] = g^{(k+1)}(1)$  lorsque  $\mathbb{E}[X^{k+1}] < \infty$ .

*Proof.* Pour  $|z| < 1$ ,  $g'(z) = \sum_{j=1}^{\infty} \mathbb{P}(X=j) j z^{j-1}$ . L'hypothèse  $\mathbb{E}[X] < \infty$  implique que cette série entière est normalement convergente sur  $[-1, 1]$  et ainsi  $g'(1) = \mathbb{E}[X]$ . De plus, pour  $|z| < 1$ ,  $g^{(k+1)}(z) = \sum_{j=k}^{\infty} \mathbb{P}(X=j) j(j-1) \times \dots \times (j-k) z^{j-k}$ . Donc si  $\mathbb{E}[X^{k+1}] < \infty$  pour  $k \geq 1$ , alors  $\mathbb{E}[X(X-1) \times \dots \times (X-k)] < \infty$  et  $\mathbb{E}[X(X-1) \times \dots \times (X-k)] = \sum_{j=k}^{\infty} \mathbb{P}(X=j) j(j-1) \times \dots \times (j-k) = g^{(k+1)}(1)$ .  $\square$

**Exemple:** Si  $X \sim \mathcal{P}(\lambda)$  (loi de Poisson) avec  $\lambda > 0$ . Alors  $g(z) = e^{\lambda(z-1)}$  pour  $z \in \mathbb{R}$ .

**A quoi sert la fonction génératrice?**

- A remplacer une mesure de probabilité par une fonction (dans le cas des v.a. à valeurs entières positives) puisque la fonction génératrice caractérise la loi;
- Mais surtout, à obtenir la loi ou montrer la convergence de sommes de variables indépendantes.

Une extension de la fonction génératrice à toute variable aléatoire, discrète, continue, ou autre, est obtenue par la fonction caractéristique:

**Définition.** Si  $X$  est une v.a. sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , on définit la **fonction caractéristique** de  $X$  par:

$$\phi_X(u) = \mathbb{E}[e^{iuX}] = \mathbb{E}[\cos(uX)] + i \mathbb{E}[\sin(uX)] \quad \text{pour tout } u \in \mathbb{R}.$$

**Remarque:**  $\phi_X$  est définie sur  $\mathbb{R}$  du fait que  $|e^{iuX}| \leq 1$  pour tout  $u \in \mathbb{R}$  et en appliquant ensuite le Théorème de convergence dominée de Lebesgue.

**Théorème.**  $\phi_X = \phi_Y \iff X$  et  $Y$  ont la même distribution de probabilité.

*Proof.* On peut prouver que pour tout  $a \leq b$  tels que  $\mathbb{P}_X(\{a, b\}) = 0$ , alors

$$\mathbb{P}(X \in [a, b]) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt.$$

Pour commencer, on montre que pour  $c \in \mathbb{R}$ ,  $\int_{-\infty}^{\infty} \frac{\sin(ct)}{t} dt = \pi$  si  $c > 0$  et  $-\pi$  si  $c < 0$ ,  $= 0$  si  $c = 0$ . Tous ces cas se ramène au calcul de  $I(0) = \int_0^{\infty} \frac{\sin(t)}{t} dt$ . Cette intégrale existe en tant qu'intégrale de Riemann (mais pas en tant qu'intégrale de Lebesgue!). En effet, la fonction  $t \rightarrow \frac{\sin(t)}{t}$  est prolongeable par continuité en 0 et en  $+\infty$ , on utilise une intégration par parties

$$\int_1^T \frac{\sin(t)}{t} dt = \left[ -\frac{\cos(t)}{t} \right]_1^T - \int_1^T \frac{\cos(t)}{t^2} dt = \cos(1) - \int_1^T \frac{\cos(t)}{t^2} dt.$$

Cette dernière intégrale existe car la fonction est majorée par  $\frac{1}{t^2}$  qui est intégrable en  $+\infty$ .

Par ailleurs, on considère pour  $t \geq 0$ ,  $I(x) = \int_0^{\infty} \frac{\sin(t)}{t} e^{-xt} dt$ . Cette fonction existe et pour  $x > 0$  elle est dérivable et  $I'(x) = -\int_0^{\infty} \sin(t) e^{-xt} dt$  pour  $x > 0$ . Or

$$\int_0^{\infty} \sin(t) e^{-xt} dt = \int_0^{\infty} \text{Im}(e^{it}) e^{-xt} dt = \text{Im} \left( \int_0^{\infty} e^{it-xt} dt \right) = \text{Im} \left( -\frac{1}{i-x} \right) = \frac{1}{1+x^2}.$$

On en déduit que  $I'(x) = -\frac{1}{1+x^2}$  soit  $I(x) = -\arctan(x) + K$  où  $K \in \mathbb{R}$ . Comme on sait que  $\lim_{x \rightarrow \infty} I(x) = 0$ , donc  $K = \pi/2$ . Il suffit de montrer que la fonction  $I$  est continue en 0, d'où  $I(0) = \int_0^{\infty} \frac{\sin(t)}{t} dt = \pi/2$ .

Revenons à la preuve et partons de  $\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt$ . Soit:

$$\begin{aligned} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt &= \frac{1}{2\pi} \int_{-T}^T \int_{-\infty}^{\infty} \frac{e^{it(x-a)} - e^{it(x-b)}}{it} d\mathbb{P}_X(x) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt d\mathbb{P}_X(x) \quad (\text{Fubini}) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \left( \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right) dt d\mathbb{P}_X(x) \quad (\text{Parité}). \end{aligned}$$

Mais si on s'intéresse à  $\lim_{T \rightarrow \infty} \int_{-T}^T \left( \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right) dt = \int_{-\infty}^{\infty} \left( \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right) dt = \phi_{a,b}(x)$ , alors  $\phi_{a,b}(x) = 0$  si  $x > b$  ou  $x < a$  et  $\phi_{a,b}(x) = 2\pi$  pour  $a < x < b$ . Pour  $x = a$  et  $x = b$ , on obtient  $\pm\pi$ , mais de l'hypothèse  $\mathbb{P}_X(\{a, b\}) = 0$  cela n'interviendra pas. D'où:

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt = \frac{1}{2\pi} \int_a^b 2\pi d\mathbb{P}_X(x) = \mathbb{P}(X \in [a, b]).$$

□

**Remarque:** Si  $X$  est une v.a. discrète à valeurs entières, alors  $\phi_X(u) = g(e^{iu})$ .

**Propriété.** Si  $X$  est une v.a. définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[|X|^k] < \infty$ , alors  $\phi_X$  est une fonction  $\mathcal{C}^k(\mathbb{R})$  et  $\phi_X^{(k)}(u) = i^k \mathbb{E}[X^k e^{iuX}]$  pour tout  $u \in \mathbb{R}$ .

*Proof.* Pour  $k = 1$ ,  $u \in \mathbb{R}$  et  $h \neq 0$ , nous avons  $\frac{\phi_X(u+h) - \phi_X(u)}{h} = \mathbb{E}\left[ e^{iuX} \left( \frac{e^{ihX} - 1}{h} \right) \right]$ . Il est clair que  $e^{iuX} \left( \frac{e^{ihX} - 1}{h} \right) \rightarrow iX e^{iuX}$  pour  $h \rightarrow 0$  puisque  $\frac{e^{ihx} - 1}{h} \rightarrow ix$  lorsque  $h \rightarrow 0$  pour tout  $x \in \mathbb{R}$  (on peut considérer  $g(h) = e^{ihx}$  et  $(g(h) - g(0))/h \rightarrow g'(0)$  lorsque  $h \rightarrow 0$ ). De plus,  $\left| e^{iuX} \left( \frac{e^{ihX} - 1}{h} \right) \right| \leq |X|$  pour tout  $u \in \mathbb{R}$  et  $h \neq 0$ , et  $\mathbb{E}[|X|] < \infty$  par hypothèse. Le théorème de Lebesgue implique alors que  $\frac{\phi_X(u+h) - \phi_X(u)}{h} \rightarrow \mathbb{E}[iX e^{iuX}] = \phi_X'(u)$ . Même type de preuve pour  $k \geq 2$ . □

**Remarque:**  $\phi_X'(0) = i \mathbb{E}[X]$  et  $\phi_X''(0) = -\mathbb{E}[X^2]$ .

Voici une autre propriété de la fonction caractéristique, spécifique aux v.a. "continues":

**Propriété.** Si  $X$  est une v.a. de loi continue par rapport à la mesure de Lebesgue, de densité  $f_X$  et de fonction caractéristique  $\phi_X$  telle que  $\int_{\mathbb{R}} |\phi_X(u)| du < \infty$ , alors

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(u) e^{-iux} du \quad \text{pour tout } x \in \mathbb{R}.$$

*Proof.* Dans le cas présent, on a:  $\phi_X(u) = \int_{-\infty}^{\infty} f_X(x) e^{-iux} dx$ . On peut reprendre la preuve d'identification de la loi mais en passant directement à la limite car la fonction caractéristique est alors de module intégrable. D'où pour tout  $a < b$ ,

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_a^b e^{-itx} dx \phi_X(t) dt = \int_a^b \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt dx,$$

par Fubini. D'où le résultat par identification puisqu'il est vrai pour tout  $a < b$ . □

## A quoi sert la fonction caractéristique?

- Fonction caractéristique de la somme de v.a. indépendantes;
- Fonction caractéristique pour caractériser l'indépendance (voir un peu plus loin);
- Convergence d'une suite de fonctions caractéristiques vers une fonction caractéristique  $\iff$  convergence en loi (Théorème de Lévy, voir un peu plus loin).

## 2 Vecteurs aléatoires

### 2.1 Définitions et premières propriétés

**Définition.** On dit que  $X$  est un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , un espace probabilisé, si  $X$  est une fonction mesurable de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ .

**Définition.** Soit  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . Alors la loi (ou mesure) de probabilité de  $X$ ,  $\mathbb{P}_X$ , est définie de façon univoque à partir de la fonction de répartition de  $X$ , telle que pour  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,

$$F_X(x) = \mathbb{P}_X\left(\prod_{i=1}^d ]-\infty, x_i]\right) = \mathbb{P}\left(X \in \prod_{i=1}^d ]-\infty, x_i]\right).$$

**Propriété.** Soit  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . On suppose que  $X = (X_1, \dots, X_d)$ . Alors les  $X_i$  sont des variables aléatoires sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , de fonction de répartition

$$F_{X_i}(x_i) = \lim_{\substack{x_j \rightarrow +\infty \\ j \neq i}} F_X(x_1, \dots, x_i, \dots, x_d).$$

*Proof.* On peut construire la fonction  $g_i$  qui associe à un vecteur de  $\mathbb{R}^d$  sa  $i$ ème coordonnée. La fonction sur  $X_i : \Omega \rightarrow \mathbb{R}$  est donc telle que  $X_i(\omega) = g_i(X(\omega))$  et comme  $g$  est continue, donc mesurable,  $X_i$  est une variable aléatoire.

Soit  $(x_1^{(n)})_n, \dots, (x_{i-1}^{(n)})_n, (x_{i+1}^{(n)})_n, \dots, (x_d^{(n)})_n$   $d-1$  suites croissantes de réels telles que  $x_j^{(n)} \xrightarrow[n \rightarrow \infty]{} +\infty$  pour tout  $j \neq i$ .

Il est clair que les pavés  $A_n = ]-\infty, x_1^{(n)}] \times \dots \times ]-\infty, x_{i-1}^{(n)}] \times ]-\infty, x_i] \times ]-\infty, x_{i+1}^{(n)}] \times \dots \times ]-\infty, x_d^{(n)}]$  forment une suite croissante de pavés de  $\mathbb{R}^n$  (on a  $A_m \subset A_{m+1}$  pour tout  $m$ ). On sait alors que  $\lim_{n \rightarrow \infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\bigcup_{n \in \mathbb{N}} A_n) = \mathbb{P}_X(\mathbb{R}^{i-1} \times ]-\infty, x_i] \times \mathbb{R}^{d-i}) = \mathbb{P}(X_i \leq x_i) = F_{X_i}(x_i)$ . Comme ceci est vrai pour toute suite croissante, on a bien le résultat.  $\square$

**Remarque:** Les mesures de probabilités  $\mathbb{P}_{X_i}$  déterminées de façon univoque à partir des  $F_{X_i}$  sont appelées **lois marginales** de  $X$ . Cependant, les lois marginales ne permettent pas d'identifier la loi du vecteur, sauf à spécifier d'autres propriétés (indépendance par exemple).

**Corollaire.** Si  $F_X(x)$  est continue sur  $\mathbb{R}^d$  et si  $\frac{\partial^d}{\partial x_1 \dots \partial x_d} F_X(x_1, \dots, x_d)$  existe presque partout sur  $\mathbb{R}^d$ , alors la densité de la loi  $\mathbb{P}_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  vaut:

$$f_X(x_1, \dots, x_d) = \frac{\partial^d}{\partial x_1 \dots \partial x_d} F_X(x_1, \dots, x_d) \quad \text{pour presque tout } (x_1, \dots, x_d) \in \mathbb{R}^d.$$

**Propriété.** Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire défini sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$  dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  et de densité  $f_X$ . Alors les  $X_i$  sont des v.a. à loi absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  de densité  $f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(z_1, \dots, z_{i-1}, x_i, z_{i+1}, \dots, z_d) dz_1 \dots dz_{i-1} dz_{i+1} \dots dz_d$ .

*Proof.* Si  $F_X$  est la fonction de répartition de  $X$ , alors pour tout  $x_i \in \mathbb{R}$ ,  $F_{X_i}(x_i) = \lim_{x_j \rightarrow \infty, j \neq i} F_X(x_1, \dots, x_d)$ . Mais comme la mesure est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , il existe  $f_X$  tel que  $F_X(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_X(z_1, \dots, z_d) dz_1 \dots dz_d$ . Donc en utilisant Fubini

$$\begin{aligned} F_{X_i}(x_i) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{x_i} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(z_1, \dots, z_{i-1}, z_i, z_{i+1}, \dots, z_d) dz_1 \dots dz_d \\ &= \int_{-\infty}^{x_i} \left( \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(z_1, \dots, z_{i-1}, z_i, z_{i+1}, \dots, z_d) dz_1 \dots dz_{i-1} dz_{i+1} \dots dz_d \right) dz_i \\ &= \int_{-\infty}^{x_i} f_{X_i}(z_i) dz_i, \end{aligned}$$

avec  $f_{X_i}(z_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(z_1, \dots, z_{i-1}, z_i, z_{i+1}, \dots, z_d) dz_1 \dots dz_{i-1} dz_{i+1} \dots dz_d$ , fonction mesurable positive: la loi de  $X_i$  est bien absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  et sa densité est  $f_{X_i}$ .  $\square$

**Remarque:** Si  $X$  est à valeurs dans  $\{x_i\}_{i \in I}$ , où  $x_i \in \mathbb{R}^d$  et  $I \subset \mathbb{N}$ , alors  $X$  est discrète et sa densité par rapport à la mesure de comptage sur  $\{x_i\}_{i \in I}$  est donnée par  $\mathbb{P}(X = x_i) = \mathbb{P}(X_1 = x_{i,1} \cap \dots, X_d = x_{i,d})$ .

**Définition.** Soit  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . La fonction caractéristique de  $X$  est la fonction  $\phi_X : \mathbb{R}^d \mapsto \mathbb{C}$  telle que pour tout  $t \in \mathbb{R}^d$ ,

$$\phi_X(t) = \mathbb{E}[\exp(i \langle t, X \rangle)] = \int_{\mathbb{R}^d} e^{i \langle t, x \rangle} d\mathbb{P}_X(x),$$

où  $\langle . \rangle$  désigne le produit scalaire euclidien sur  $\mathbb{R}^d$  vérifiant  $\langle t, x \rangle = \sum_{i=1}^d t_i x_i$  pour  $t = (t_1, \dots, t_d)$

et  $x = (x_1, \dots, x_d)$ .

**Remarque:** La fonction caractéristique existe sur  $\mathbb{R}$  et  $\phi_X(0) = 1$ .  $\phi_X$  est aussi la transformée de Fourier de la mesure  $\mathbb{P}_X$ .

**Théorème.** Soit  $X$  et  $Y$  des vecteurs aléatoires sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ , de lois  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ . Alors  $\mathbb{P}_X = \mathbb{P}_Y$  si et seulement si  $\phi_X = \phi_Y$ .

*Proof.* Même type de preuve que dans  $\mathbb{R}$  sauf que dans le cas multidimensionnel on choisit des  $a_j \leq b_j$  où  $j = 1, \dots, n$  vérifiant  $\mathbb{P}_X(\{a_1, b_1\} \times \dots \times \{a_d, b_d\}) = 0$ , et alors on montre que

$$\mathbb{P}(X \in [a_1, b_1] \times \dots \times [a_d, b_d]) = \lim_{T \rightarrow \infty} \prod_{j=1}^d \left( \frac{1}{2\pi} \int_{-T}^T \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \right) \phi_X(t_1, \dots, t_d) dt_1 \dots dt_d.$$

□

**Théorème** (Théorème d'inversion). Si  $X$  est un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$  et si  $\phi_X$  est une fonction intégrable par rapport à la mesure de Lebesgue  $\lambda_d$  sur  $\mathbb{R}^d$ , alors  $X$  admet une densité  $f_X$  par rapport à  $\lambda_d$  telle que pour  $x \in \mathbb{R}^d$ ,

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \phi_X(t) dt.$$

*Proof.* Même type de preuve que dans  $\mathbb{R}$ . □

**Corollaire. (Lemme de Cramèr-Wold)** Si  $X$  et  $Y$  sont deux vecteurs aléatoires définis sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . Alors  $\mathbb{P}_X = \mathbb{P}_Y$  si et seulement si  $\langle u, X \rangle \stackrel{\mathcal{L}}{\sim} \langle u, Y \rangle$  pour tous les  $u \in \mathbb{R}^d$ .

Ceci signifie que la loi d'un vecteur aléatoire  $X = (X_1, \dots, X_d)$  peut être déterminée à partir de la loi de des combinaisons linéaires de ses coordonnées qui sont les variables aléatoires  $X_i$ . Cela peut être intéressant pour montrer notamment des théorèmes de la limite centrale multidimensionnels: on pourra se restreindre à montrer un théorème de la limite centrale unidimensionnel pour une combinaison linéaire générale.

**Définition.** Pour  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$  telle que  $\mathbb{E}[\|X\|^r] < \infty$  avec  $r > 0$ .

1. Si  $r = 1$ , alors on définit le vecteur  $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])^\top$ , qui est appelé **espérance** de  $X$ ;
2. Si  $r = 2$ , alors on définit la matrice  $\text{cov}(X) = \mathbb{E}[(X - \mathbb{E}[X]) (X - \mathbb{E}[X])^\top] = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq d}$  appelée **matrice de variance-covariance** de  $X$ .

**Propriété.** Soit  $X$  et  $Y$  deux vecteurs aléatoires définis sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ .

1. Si  $A$  et  $B$  sont des matrices de réels de taille  $(\ell, d)$  et  $C$  un vecteur de  $\mathbb{R}^d$ , alors:

$$\mathbb{E}[AX + BY + C] = A \mathbb{E}[X] + B \mathbb{E}[Y] + C.$$

2. Si  $A$  est une matrice de réels de taille  $(\ell, d)$  et  $C$  un vecteur de  $\mathbb{R}^d$ ,

$$\text{cov}(AX + C) = A \text{cov}(X) A^\top.$$

*Proof.* 1. Voir la dimension 1.

2. On a  $\mathbb{E}[AX + C] = A\mathbb{E}[X] + C$  d'où:

$$\text{cov}(AX + C) = \mathbb{E}[(AX + C) - \mathbb{E}[AX + C]]((AX + C) - \mathbb{E}[AX + C])^\top = \mathbb{E}[A(X - \mathbb{E}[X])^t(A(X - \mathbb{E}[X]))^\top] = A \text{cov}(X) A^\top.$$

□

**Propriété.** Pour  $X$  un vecteur aléatoire sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$  telle que  $\mathbb{E}[\|X\|^2] < \infty$  alors sa matrice de variance-covariance  $\Sigma$  est une matrice symétrique positive.

*Proof.* Comme  $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$  la matrice  $\Sigma$  est clairement symétrique. Elle est donc la matrice d'une forme quadratique et est diagonalisable dans  $\mathbb{R}$ . De plus, pour tout  $u \in \mathbb{R}^d$ ,  ${}^t u X$  est une v.a. et sa variance  $\text{var}({}^t u X) \geq 0$ , comme toute variance. Mais  $\text{var}({}^t u X) = \text{cov}({}^t u X) = {}^t u \text{cov}(X) u$  par la formule précédente. Donc  ${}^t u \text{cov}(X) u \geq 0$  pour tout  $u \in \mathbb{R}^d$ : la matrice est positive (les valeurs propres sont donc toutes positives ou nulles). □

## 2.2 Indépendance

**Définition.** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité et  $I \subset \mathbb{N}$ .

- Soit  $(A_i)_{i \in I}$  une famille d'événements de  $\mathcal{A}$ . On dit que les événements  $(A_i)_{i \in I}$  sont indépendants si et seulement si pour tous les sous-ensembles finis  $K \subset I$ ,

$$\mathbb{P}\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} \mathbb{P}(A_i).$$

- Soit  $(\mathcal{A}_i)_{i \in I}$  une famille de sous-tribus de  $\mathcal{A}$  (donc pour tout  $i \in I$ ,  $\mathcal{A}_i \subset \mathcal{A}$ ). On dit que les tribus  $(\mathcal{A}_i)_{i \in I}$  sont indépendantes si et seulement si pour tous les sous-ensembles finis  $K \subset I$ , et pour tous les événements  $A_k \in \mathcal{A}_k$  avec  $k \in K$ , les  $A_k$  sont indépendants.
- Soit  $(X_i)_{i \in I}$  une famille de variables aléatoires sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . On dit que les v.a.  $(X_i)_{i \in I}$  sont indépendantes si et seulement si les tribus engendrées  $(X_i^{-1}(\mathcal{B}(\mathbb{R})))_{i \in I}$  sont indépendantes, ou encore pour tous les sous-ensembles finis  $K \subset I$ :

$$\mathbb{P}\left(\bigcap_{i \in K} X_i \in B_i\right) = \prod_{i \in K} \mathbb{P}(X_i \in B_i) \quad \text{pour tous } B_i \in \mathcal{B}(\mathbb{R}).$$

**Proposition.** Si  $(X_1, \dots, X_n)$  sont des variables aléatoires sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors les  $(X_i)$  sont indépendantes si et seulement si  $\mathbb{P}_{(X_1, \dots, X_n)} = \bigotimes_{i=1}^n \mathbb{P}_{X_i}$ .

*Proof.* Par le théorème du transport et Fubini, il est clair que pour toute famille  $(B_i)_{1 \leq i \leq n}$  de boréliens de  $\mathcal{B}(\mathbb{R})$  alors:

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \int_{B_1 \times \dots \times B_n} d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n \int_{B_i} d\mathbb{P}_{X_i}(x_i) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i).$$

□

**Proposition.** Si  $(X_i)_{i \in I}$  sont des variables aléatoires indépendantes sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors les  $(X_i)$  sont indépendantes si et seulement si pour tout  $J \subset I$ ,  $J$  fini, pour toutes fonctions boréliennes  $(g_j)_{j \in J}$  telles que  $g_j(X_j)$  soit intégrable, alors

$$\mathbb{E}\left[\prod_{j \in J} g_j(X_j)\right] = \prod_{j \in J} \mathbb{E}[g_j(X_j)].$$

*Proof.*  $\implies$  Par le théorème du transport, puis par Fubini,

$$\mathbb{E}\left[\prod_{j \in J} g_j(X_j)\right] = \int_{\mathbb{R}^{|J|}} \prod_{j \in J} g_j(x_j) d\mathbb{P}_{(X_j)_{j \in J}}((dx_j)_{j \in J}) = \prod_{j \in J} \int_{\mathbb{R}} g_j(x_j) d\mathbb{P}_{X_j}(dx_j) = \prod_{j \in J} \mathbb{E}[g_j(X_j)].$$

$\Leftarrow$  On prenant le cas particulier  $g_j(x_j) = \mathbb{1}_{x_j \in B_j}$ , avec  $(B_j)$  des boréliens de  $\mathcal{B}(\mathbb{R})$  on retombe sur la définition de l'indépendance. □

**Corollaire.**  $(X_1, \dots, X_d)$  sont des variables aléatoires indépendantes si et seulement si pour tout  $(t_1, \dots, t_d) \in \mathbb{R}^d$ ,

$$\phi_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = \prod_{j=1}^d \phi_{X_j}(t_j).$$

*Proof.*  $\implies$  Comme  $\phi_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = \mathbb{E}[e^{i \sum_{j=1}^d t_j X_j}] = \mathbb{E}[\prod_{j=1}^d e^{i t_j X_j}]$ , on utilise la caractérisation précédente de l'indépendance avec les  $g_j(x) = e^{i t_j x}$ .

$\Leftarrow$  D'après la formule de caractérisation de la loi par la fonction caractéristique, avec  $\mathbb{P}(\{a_i, b_i\}_{1 \leq i \leq d}) = 0$ ,

$$\begin{aligned} \mathbb{P}(X \in [a_1, b_1] \times \dots \times [a_d, b_d]) &= \lim_{T \rightarrow \infty} \int_{-T}^T \dots \int_{-T}^T \prod_{j=1}^d \left( \frac{1}{2\pi} \int_{-T}^T \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \right) \phi_X(t_1, \dots, t_d) dt_1 \dots dt_d \\ &= \lim_{T \rightarrow \infty} \prod_{j=1}^d \left( \frac{1}{2\pi} \int_{-T}^T \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \phi_{X_j}(t_j) dt_j \right) = \prod_{j=1}^d \mathbb{P}(X_j \in [a_j, b_j]), \end{aligned}$$

soit la caractérisation de l'indépendance pour presque tous les pavés fermés.  $\square$

**Corollaire.** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$  dont la mesure de probabilité  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue  $\lambda_d$  sur  $\mathbb{R}^d$  avec pour densité  $f_X$ . Alors:

$$(X_1, \dots, X_n) \text{ v.a. indépendantes} \iff f_X(x_1, \dots, x_n) = \prod_{i=1}^d f_{X_i}(x_i) \text{ pour tout } (x_1, \dots, x_n) \in \mathbb{R}^n.$$

*Proof.* On utilise la formule d'inversion pour la fonction caractéristique:

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u_1, \dots, u_n) e^{-i(u_1 x_1 + \dots + u_n x_n)} du_1 \dots du_n = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \prod_{j=1}^n \phi_{X_j}(u_j) e^{-i u_j x_j} du_1 \dots du_n$$

d'après la caractérisation précédente. Il ne reste plus qu'à utiliser Fubini et à réécrire l'intégrale multiple comme un produit d'intégrales simples.  $\square$

**Corollaire.** Deux vecteurs aléatoires définis sur  $(\Omega, \mathcal{A}, \mathbb{P})$  sont indépendants si et seulement si toute combinaison linéaire de l'un est indépendante de toute combinaison linéaire de l'autre.

**Corollaire.** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $\mathbb{R}^d$ . On suppose que  $\mathbb{E}[\|X\|^2] < \infty$  et  $(X_1, \dots, X_n)$  mutuellement indépendantes. Alors:

$cov(X)$  est une matrice diagonale avec les  $var(X_i)$  sur sa diagonale.

*Proof.* On utilise ici le fait que  $cov(X) = (cov(X_i, X_j))_{1 \leq i, j \leq n}$  et  $cov(X_i, X_j) = 0$  si  $i \neq j$ .  $\square$

**Remarque:** Il est bien connu que  $cov(X_1, X_2) = 0$  n'entraîne pas que  $X_1$  et  $X_2$  soient indépendantes. Exemple:  $X_1 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, 1)$  et  $X_2 = X_1^2$ .

**Propriété.** Soit  $X$  et  $Y$  deux v.a. indépendantes et de loi absolument continues par rapport à la mesure de Lebesgue  $\lambda_1$ , de densités  $f_X$  et  $f_Y$ . Alors  $Z = X + Y$  est une v.a. absolument continue par rapport à la mesure de Lebesgue  $\lambda_1$  de densité  $f_Z$  et

$$f_Z(z) = \int_{\mathbb{R}} f_X(t) f_Y(z - t) dt = \int_{\mathbb{R}} f_Y(t) f_X(z - t) dt \text{ pour tout } z \in \mathbb{R}.$$

$f_Z$  est appelée **produit de convolution de  $f_X$  et  $f_Y$** .

*Proof.* On peut écrire en utilisant Fubini et des changements de variables, que pour tout  $z \in \mathbb{R}$ ,

$$f_{X+Y}(z) = \mathbb{P}(X+Y \leq z) = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{z-y} f_X(x) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^z f_Y(y) f_X(x'-y) dx' dy = \int_{-\infty}^z \left( \int_{-\infty}^{\infty} f_Y(y) f_X(x'-y) dy \right) dx',$$

d'où le résultat.  $\square$

**Remarque:** Attention l'hypothèse d'indépendance est nécessaire pour obtenir le produit de convolution.

### 3 Vecteurs gaussiens

#### 3.1 Définitions et premières propriétés

**Définition.** On dit qu'une variable aléatoire  $X$  est gaussienne si c'est une variable absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}\right) \quad \text{avec } m \in \mathbb{R} \text{ et } \sigma^2 > 0.$$

On note  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m, \sigma^2)$  et  $m = \mathbb{E}[X]$  et  $\sigma^2 = \text{var}(X)$ .

**Propriété.** Si  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m, \sigma^2)$ , alors:

1.  $X = m + \sigma Z$  avec  $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, 1)$ , loi normale centrée réduite.
2.  $\mathbb{E}[|X|^p] = f_p(m, \sigma^2) < \infty$  pour tout  $p \in \mathbb{N}^*$ .
3. La fonction caractéristique de  $X$  est  $\phi_X(u) = \exp\left(imu - \frac{1}{2}\sigma^2 u^2\right)$ .

**Définition.** On dit qu'un vecteur aléatoire  $(X_1, \dots, X_d)$  non nul défini sur  $(\Omega, \mathcal{A}, \mathbb{P})$  est un vecteur gaussien si toute combinaison linéaire  $u_1 X_1 + \dots + u_d X_d$ , avec  $(u_i)_{1 \leq i \leq d} \in \mathbb{R}^d$ , est une variable gaussienne ou une constante.

**Conséquence.** Si  $(X_1, \dots, X_d)$  est un vecteur gaussien, alors chaque  $X_i$  est une variable gaussienne. La réciproque est fautive en général (voir exercice).

**Conséquence.** Si  $X = {}^t(X_1, \dots, X_d)$  est un vecteur gaussien, alors pour toute matrice  $M$  de réels de taille  $(d', d)$  et tout vecteur colonne  $B$  de taille  $d'$ ,  $MX + B$  est aussi un vecteur gaussien.

**Propriété.** Si  $X = (X_1, \dots, X_d)$  est un vecteur gaussien, alors sa loi de probabilité ne dépend que de son espérance et de sa matrice de variance-covariance.

*Proof.* Commençons par étudier le cas où  $X$  est centré. On a  $\phi_X(u) = \mathbb{E}[e^{i \sum_{j=1}^d u_j X_j}]$ . Comme  $\sum_{j=1}^d u_j X_j$  est une variable aléatoire gaussienne centrée, sa fonction caractéristique dépend de sa variance et  $\text{var}(\sum_{j=1}^d u_j X_j) = \sum_{1 \leq j, k \leq d} u_j u_k \text{cov}(X_j, X_k)$ . On en déduit que:

$$\phi_X(u) = \exp\left(-\frac{1}{2} \left(\sum_{1 \leq j, k \leq d} u_j u_k \text{cov}(X_j, X_k)\right)\right) = \exp\left(-\frac{1}{2} {}^t u \Sigma u\right),$$

où  $\Sigma = (\text{cov}(X_j, X_k))_{1 \leq j, k \leq d}$  est la matrice de covariance de  $X$ .

Si maintenant,  $X$  n'est plus centré, si on note  $m = {}^t(m_1, \dots, m_d)$  son espérance, on peut écrire que

$$\phi_X(u) = \mathbb{E}[e^{i \langle u, m \rangle} e^{i \sum_{j=1}^d u_j (X_j - m_j)}] = e^{i \langle u, m \rangle} \exp\left(-\frac{1}{2} {}^t u \Sigma u\right).$$

La fonction caractéristique de  $X$  ne dépend donc que de son espérance et sa matrice de covariance, et il en est de même pour sa loi de probabilité.  $\square$

**Conséquence:** Si  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m, \Sigma)$  alors  $X = m + \Sigma^{1/2} Z$ , avec  $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0_d, I_d)$  vecteur gaussien centré réduit (la matrice  $\Sigma^{1/2}$  peut être obtenue par diagonalisation avec racines des valeurs propres).

**Propriété.** Si  $X_1, \dots, X_d$  sont  $d$  v.a. gaussiennes indépendantes alors  $X = (X_1, \dots, X_d)$  est un vecteur gaussien de matrice de variance-covariance diagonale.

*Proof.* On a vu que l'indépendance entre 2 v.a. entraîne la nullité de leur covariance. Or la matrice de covariance est composée en dehors de la diagonale des  $\text{cov}(X_j, X_k)$  avec  $k \neq j$ . Si on considère la fonction caractéristique de  $X$ , alors, en utilisant l'indépendance des  $X_j$ :

$$\phi_X(u) = \prod_{j=1}^d \phi_{X_j}(u_j) = e^{i \langle u, m \rangle} e^{-\frac{1}{2} \sum_{j=1}^d \sigma_j^2 u_j^2} = e^{i \langle u, m \rangle} \exp\left(-\frac{1}{2} {}^t u \Sigma u\right),$$

avec  $\Sigma$  la matrice de covariance de  $X$  qui est donc diagonale: on aboutit à la fonction caractéristique d'un vecteur gaussien.  $\square$

**Corollaire.** Soit  $X = (X_1, \dots, X_d)$  un vecteur gaussien. Alors:

$$(X_1, \dots, X_d) \text{ indépendantes} \iff \text{cov}(X_i, X_j) = 0 \text{ pour } i \neq j.$$

*Proof.* Dans le sens  $\implies$  c'est immédiat. Dans le sens  $\impliedby$  on reprend la preuve un peu plus haut.  $\square$

**Propriété.** Si  $X = (X_1, \dots, X_d)$  est un vecteur gaussien d'espérance  $\mathbb{E}[X]$  dont la matrice de variance-covariance  $\Sigma$  est définie positive, alors  $X$  est absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  et sa densité est:

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2} (x - \mathbb{E}[X])^\top \Sigma^{-1} (x - \mathbb{E}[X])\right) \text{ pour } x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d.$$

*Proof.* En premier lieu, on obtient de manière immédiate que si le vecteur  $Z$  est centré réduit, soit  $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, I_d)$  alors toutes ses composantes sont indépendantes donc sa densité est  $\prod_{j=1}^d \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2} z_j^2} = \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2} \sum_{j=1}^d z_j^2}$ .

Par ailleurs, on a vu que si  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\mathbb{E}[X], \Sigma)$  alors  $X \stackrel{\mathcal{L}}{\sim} \mathbb{E}[X] + \Sigma^{1/2} Z$ . Ainsi pour toute fonction  $g: \mathbb{R}^d \rightarrow \mathbb{R}$  bornée, alors:

$$\mathbb{E}[g(X)] = \mathbb{E}[g(\mathbb{E}[X] + \Sigma^{1/2} Z)] = \int_{\mathbb{R}^d} g(\mathbb{E}[X] + \Sigma^{1/2} z) f_Z(z) dz = \int_{\mathbb{R}^d} g(x) \det(\Sigma^{-1/2}) f_Z(\Sigma^{-1/2}(x - \mathbb{E}[X])) dx$$

avec la formule du changement de variable pour les intégrales multiples. Comme  $\det(\Sigma^{-1/2}) = (\det(\Sigma))^{-1/2}$  et  $(\Sigma^{-1/2}(x - \mathbb{E}[X]))^\top \Sigma^{-1/2}(x - \mathbb{E}[X]) = (x - \mathbb{E}[X])^\top \Sigma^{-1}(x - \mathbb{E}[X])$  on obtient le résultat.  $\square$

### 3.2 Lois découlant de la loi gaussienne

**Définition.** Soit  $X = (X_1, \dots, X_d)$  un vecteur gaussien centré réduit,  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, I_d)$ . Alors la loi de  $Z = \|X\|^2 = X_1^2 + \dots + X_d^2$  est appelée **loi du chi2 à  $d$  degrés de liberté** notée  $Z \stackrel{\mathcal{L}}{\sim} \chi^2(d)$ .

**Propriété.** Pour  $d \in \mathbb{N}^*$ ,  $\chi^2(d) \stackrel{\mathcal{L}}{\sim} \Gamma\left(\frac{d}{2}, \frac{1}{2}\right)$ .

*Proof.* On sait que si  $U \stackrel{\mathcal{L}}{\sim} \Gamma(\alpha, \beta)$  et  $v \stackrel{\mathcal{L}}{\sim} \Gamma(\alpha, \beta)$  et si  $U \perp V$ , alors  $U + V \stackrel{\mathcal{L}}{\sim} \Gamma(2\alpha, \beta)$ . Il suffit donc de démontrer que  $X_1^2 \stackrel{\mathcal{L}}{\sim} \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ .  
Or pour  $x \geq 0$ ,

$$\mathbb{P}(X_1^2 \leq x) = 2\mathbb{P}(0 \leq X_1 \leq \sqrt{x}) = 2(F(\sqrt{x}) - \frac{1}{2}) \implies f_{X_1^2}(x) = \frac{1}{\sqrt{2\pi}\sqrt{x}} e^{-x/2} = \frac{(\frac{1}{2})^{1/2}}{\Gamma(1/2)} x^{\frac{1}{2}-1} e^{-\frac{1}{2}x},$$

qui est bien la loi  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ .  $\square$

**Théorème** (Théorème de Cochran). Soit un vecteur gaussien centré réduit  $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}_d(0_d, I_d)$ . En considérant le produit scalaire euclidien classique sur  $\mathbb{R}^d$ , on considère  $A$  et  $B$  deux sous-espaces vectoriels orthogonaux de  $\mathbb{R}^d$  et on note  $P_A$  et  $P_B$  les matrices de projection orthogonale sur  $A$  et sur  $B$ . Alors:

1.  $P_A X$  et  $P_B X$  sont deux vecteurs gaussiens **indépendants** à valeurs dans  $\mathbb{R}^d$ ;
2.  $\|P_A X\|^2 \stackrel{\mathcal{L}}{\sim} \chi^2(\dim(A))$ , loi du  $\chi^2$  à  $\dim(A)$  degrés de liberté.

*Proof.* 1. Comme  $P_A$  est une matrice de réels et  $\varepsilon$  un vecteur gaussien, alors  $P_A \varepsilon$  est un vecteur gaussien, centré car  $\varepsilon$  est centré. Donc  $(P_A \varepsilon, P_B \varepsilon)$  est une famille indépendante si  $\text{cov}(P_A \varepsilon, P_B \varepsilon) = 0$  car  $(P_A \varepsilon, P_B \varepsilon)$  est un vecteur gaussien (de taille  $2d$ ). Or  $\text{cov}(P_A \varepsilon, P_B \varepsilon) = \mathbb{E}[P_A \varepsilon {}^t(P_B \varepsilon)] = \mathbb{E}[P_A \varepsilon {}^t \varepsilon {}^t P_B] = P_A \text{cov}(\varepsilon) {}^t P_B = P_A P_B$ . Mais  $A \perp B$  donc  $P_A P_B = 0$ , d'où le résultat.

2.  $\|P_A \varepsilon\|^2 = {}^t \varepsilon {}^t P_A P_A \varepsilon = {}^t \varepsilon P_A \varepsilon$ . Mais  $P_A$  est une matrice réelle symétrique donc diagonalisable et on peut écrire que  $P_A = Q D {}^t Q$  avec  $Q$  une matrice orthogonale et  $D$  une matrice diagonale avec les valeurs propres de  $P_A$ , donc par exemple d'abord  $\dim(A)$  uns sur la diagonale puis dessous  $n - \dim(A)$  zéros. D'où  $\|P_A \varepsilon\|^2 = {}^t ({}^t Q \varepsilon) D ({}^t Q \varepsilon)$ . Soit  $\varepsilon' = {}^t Q \varepsilon$ . Alors  $\varepsilon' \stackrel{\mathcal{L}}{\sim} \mathcal{N}_n(0, {}^t Q Q) = \mathcal{N}_n(0, I_d)$  car  $Q$  est une matrice orthogonale (donc  $Q^{-1} = {}^t Q$ ). Donc  $\varepsilon'$  est un vecteur gaussien centré standard. Comme  ${}^t \varepsilon' D \varepsilon' = \sum_{j=1}^{\dim(A)} (\varepsilon'_j)^2$ , les  $\varepsilon'_j$  étant des variables gaussiennes centrées indépendantes de même variance 1, on a donc  $\sum_{j=1}^{\dim(A)} (\varepsilon'_j)^2 \stackrel{\mathcal{L}}{\sim} \chi^2(\dim(A))$ . □

**Définition.** Soit  $X$  une v.a. gaussienne de loi  $\mathcal{N}(0, 1)$  et  $Y$  une v.a. **indépendante** de  $X$  et suivant une loi  $\chi^2(d)$  avec  $d \in \mathbb{N}^*$ . Alors la variable  $T = X/\sqrt{Y/d}$  suit une loi dite de Student à  $d$  degrés de liberté, et on notera  $T \stackrel{\mathcal{L}}{\sim} t(d)$ .

**Propriété.** Si  $T$  suit une loi de Student à  $d$  degrés de liberté:

- la loi de  $T$  est symétrique et de densité  $f_T(t) = \frac{1}{\sqrt{d\pi}} \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \left(1 + \frac{t^2}{d}\right)^{-\frac{d+1}{2}}$ .
- $\mathbb{E}[|T|^\alpha] < \infty$  pour  $\alpha < d$  et  $\mathbb{E}[|T|^d] = \infty$ .
- Si  $d = 1$ ,  $T$  suit une loi de Cauchy.

*Proof.* • Pour  $t \in \mathbb{R}$ , du fait que  $X$  et  $Y$  sont indépendantes et du fait que  $Y \geq 0$ , on peut écrire que:

$$\mathbb{P}(T \leq t) = \mathbb{P}(X \leq t \sqrt{Y/d}) = \int_0^\infty \frac{(\frac{1}{2})^{d/2}}{\Gamma(\frac{d}{2})} y^{\frac{d}{2}-1} e^{-y/2} \int_{-\infty}^{t\sqrt{y/d}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx dy = \frac{(\frac{1}{2})^{d/2}}{\sqrt{2\pi} \sqrt{d} \Gamma(\frac{d}{2})} \int_0^\infty \int_{-\infty}^t y^{\frac{d}{2}-\frac{1}{2}} e^{-y/2} e^{-\frac{y u^2}{d}} du dy$$

après un changement de variable. De ceci, en appliquant Fubini et un nouveau changement de variable, on en déduit que:

$$\mathbb{P}(T \leq t) = \frac{(\frac{1}{2})^{d/2}}{\sqrt{2\pi} \sqrt{d} \Gamma(\frac{d}{2})} \int_{-\infty}^t \int_0^\infty y^{\frac{(d+1)}{2}-1} e^{-\frac{1}{2}y(1+\frac{u^2}{d})} dy du = \frac{(\frac{1}{2})^{d/2}}{\sqrt{2\pi} \sqrt{d} \Gamma(\frac{d}{2})} \int_{-\infty}^t (1+\frac{u^2}{d})^{-\frac{(d+1)}{2}} \int_0^\infty y^{\frac{(d+1)}{2}-1} e^{-\frac{1}{2}y} dy du.$$

Il ne reste plus alors qu'à utiliser le fait que  $\int_0^\infty y^{\frac{(d+1)}{2}-1} e^{-\frac{1}{2}y} dy = \Gamma(\frac{d+1}{2}) (\frac{1}{2})^{-(d+1)/2}$ , et on obtient le résultat désiré.

- En utilisant un équivalent, on a  $|t|^\alpha f_T(t) \sim |t|^\alpha t^{-(d+1)}$  pour  $t \rightarrow \infty$  et il ne reste plus qu'à utiliser les résultats connus sur les intégrales de Riemann.
- Si  $d = 1$ , on tombe bien sur la densité  $\frac{1}{\pi} \frac{1}{1+x^2}$  qui est celle d'une loi de Cauchy. □

**Propriété.** Soit  $(X_1, \dots, X_n)$  une suite de v.a.i.i.d. gaussienne de loi  $\mathcal{N}(m, \sigma^2)$ . Soit la moyenne et la variance empiriques de  $(X_1, \dots, X_n)$  définies respectivement par:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \bar{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Alors:

- La loi de  $\bar{X}_n$  est  $\mathcal{N}(m, \frac{\sigma^2}{n})$  et la loi de  $\bar{\sigma}_n^2$  est  $\frac{\sigma^2}{n-1} \chi^2(n-1)$ .

- $\bar{X}_n$  et  $\bar{\sigma}_n^2$  sont deux v.a. indépendantes et la loi de  $\hat{T}_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{\bar{\sigma}_n^2}}$  est une loi  $t(n-1)$ .

*Proof.* Voir exercice de la feuille de TD3. □

**Définition.** Soit  $X_1$  et  $X_2$  deux v.a. **indépendantes** suivant respectivement des lois  $\chi^2(n_1)$  et  $\chi^2(n_2)$ . Alors la variable  $Z = \frac{\frac{1}{n_1} X_1}{\frac{1}{n_2} X_2}$  suit une loi dite de Fisher à  $(n_1, n_2)$  degrés de liberté, et on notera  $Z \stackrel{\mathcal{L}}{\sim} \mathcal{F}(n_1, n_2)$ .

## 4 Convergence de suites de variables aléatoires et théorèmes limite

### 4.1 Convergence de suites de variables aléatoires

**Remarque importante!** Toutes (ou presque) les définitions et propriétés données ci-dessous le sont pour des suites de variables aléatoires. Cependant elles sont également valables pour des suites de vecteurs aléatoires, quitte à adapter les définitions et preuves avec  $\mathbb{R}^d$  et  $\|\cdot\|$  en lieu et place de  $\mathbb{R}$  et  $|\cdot|$ .

**Définition.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a. définie sur le même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ . Et soit  $X_\infty$  une v.a. définie également sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors:

- $(X_n)$  converge en loi vers  $X_\infty$ , noté  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty$ , lorsque,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_{X_\infty}(x) \quad \text{où } x \in \mathbb{R} \text{ et } F_{X_\infty} \text{ continue en } x.$$

- $(X_n)$  converge en probabilité vers  $X_\infty$ , noté  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X_\infty$ , lorsque pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X_\infty| \geq \varepsilon) = 0.$$

- $(X_n)$  converge dans  $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$  vers  $X_\infty$ , noté  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X_\infty$ , avec  $p > 0$ , lorsque

$$\mathbb{E}[|X_n|^p + |X_\infty|^p] < \infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X_\infty|^p] = 0.$$

- $(X_n)$  converge presque sûrement vers  $X_\infty$ , noté  $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X_\infty$ , lorsque

$$\lim_{n \rightarrow \infty} X_n(\omega) = X_\infty(\omega) \quad \text{pour } \mathbb{P}\text{-presque tout } \omega \in \Omega.$$

**Théorème.** Soit  $(X_n)_{n \in \mathbb{N}}$  est une suite de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $X_\infty$  une v.a. également définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors,

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty \iff \mathbb{E}[g(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[g(X_\infty)] \text{ pour toute fonction } g : \mathbb{R} \rightarrow \mathbb{R} \text{ continue bornée.}$$

*Proof.*  $\implies$  Soit  $g$  une fonction continue bornée sur  $\mathbb{R}$  et sans perte de généralité on va supposer que  $\|g\|_\infty \leq 1$ . Soit  $\varepsilon > 0$ . Comme  $\lim_{x \rightarrow +\infty} \mathbb{P}(X > x) = 0 = \lim_{x \rightarrow -\infty} \mathbb{P}(X < x)$ , il existe un intervalle  $[a, b]$  de  $\mathbb{R}$  tel que  $\mathbb{P}(X_\infty \notin [a, b]) \leq \varepsilon$ , avec  $F_{X_\infty}$  continue en  $a$  et  $b$ . Sur  $[a, b]$ , qui est un compact,  $g$  est uniformément continue, ce qui signifie qu'il existe  $\eta_\varepsilon > 0$  tel que pour tout  $x_0 \in [a, b]$  et tout  $x \in [x_0 - \eta_\varepsilon, x_0 + \eta_\varepsilon]$  alors  $|g(x_0) - g(x)| \leq \varepsilon$ . On peut donc décomposer  $[a, b]$  en  $m_\varepsilon = \lfloor (b-a)/\eta_\varepsilon \rfloor$  intervalles  $[a_i, b_i]$ , soit  $[a, b] = \bigcup_{i=1}^{m_\varepsilon} [a_i, b_i]$ , avec pour tout  $(x, y) \in [a_i, b_i]^2$ ,  $|g(x) - g(y)| \leq \varepsilon$  et  $F_{X_\infty}$  continue en  $a_i$  et  $b_i$ . Soit  $g_\varepsilon$  la fonction en escalier telle que  $g_\varepsilon(x) = \sum_{i=1}^{m_\varepsilon} g(x_i) \mathbb{1}_{x \in [a_i, b_i]}$  pour tout  $x \in \mathbb{R}$ , où  $x_i \in [a_i, b_i]$  pour tout  $i = 1, \dots, m_\varepsilon$ . Il est clair que pour tout  $x \in [a, b]$ ,  $|g(x) - g_\varepsilon(x)| \leq \varepsilon$ . Ainsi:

$$|\mathbb{E}[g(X_n) - g_\varepsilon(X_n)]| \leq \mathbb{E}[|g(X_n) - g_\varepsilon(X_n)| \mathbb{1}_{X_n \in [a, b]}] + \mathbb{E}[|g(X_n) - g_\varepsilon(X_n)| \mathbb{1}_{X_n \notin [a, b]}] \leq \varepsilon + \mathbb{P}(X_n \notin [a, b])$$

car  $g_\varepsilon = 0$  en dehors de  $[a, b]$  et  $\|g\|_\infty \leq 1$ . De la même manière,

$$|\mathbb{E}[g(X_\infty) - g_\varepsilon(X_\infty)]| \leq \varepsilon + \mathbb{P}(X_\infty \notin [a, b])$$

Il est clair que comme  $F_{X_n} \xrightarrow[n \rightarrow \infty]{} F_X$ , alors  $\mathbb{P}(X_n \notin [a, b]) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(X_\infty \notin [a, b])$ , donc il existe  $N$  tel que pour  $n \geq N$ ,  $\mathbb{P}(X_n \notin [a, b]) \leq 2\varepsilon$ .

Enfin, il existe  $N'$  tel que pour tout  $n \geq N'$ ,

$$|\mathbb{E}[g_\varepsilon(X_n) - g_\varepsilon(X_\infty)]| \leq \sum_{i=1}^{m_\varepsilon} |g(x_i)| |\mathbb{P}(X_n \in [a_i, b_i]) - \mathbb{P}(X_\infty \in [a_i, b_i])| \leq \varepsilon,$$

car il existe  $N'$  tel que pour tout  $i$ ,  $|\mathbb{P}(X_n \in [a_i, b_i]) - \mathbb{P}(X_\infty \in [a_i, b_i])| \leq \frac{1}{m_\varepsilon} \varepsilon$ . Il ne reste plus qu'à écrire que pour  $n \geq \max(N, N')$ ,

$$|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X_\infty)]| \leq |\mathbb{E}[g(X_n) - g_\varepsilon(X_n)]| + |\mathbb{E}[g_\varepsilon(X_\infty) - g_\varepsilon(X_n)]| + |\mathbb{E}[g(X_\infty) - g_\varepsilon(X_\infty)]| \leq 3\varepsilon + \varepsilon + 2\varepsilon \leq 6\varepsilon,$$

d'où la convergence demandée. □

**Théorème (Théorème de Lévy (1920)).** Soit  $(X_n)_{n \in \mathbb{N}}$  est une suite de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $X_\infty$  une v.a. également définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors, en notant  $\phi_{X_n}$  et  $\phi_{X_\infty}$  les fonctions caractéristiques de  $X_n$  et de  $X_\infty$ ,

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty \iff \phi_{X_n}(u) \xrightarrow[n \rightarrow \infty]{} \phi_{X_\infty}(u) \text{ pour tout } u \in \mathbb{R}.$$

*Proof.*  $\implies$  Il suffit de considérer le théorème précédent et choisir  $g(x) = e^{iux}$  qui est bien continue et bornée.  
 $\longleftarrow$  Trop difficile. □

**Propriété.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a. définie sur le même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $X_\infty$  une v.a. définie également sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors:

1.  $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X_\infty$  ou  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X_\infty \implies X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X_\infty$ .
2.  $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X_\infty \implies X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X_\infty$ .
3. pour  $q \geq p$ ,  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X_\infty \implies X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^q} X_\infty$ .
4. La convergence en loi n'entraîne pas la convergence en probabilité. Mais pour  $C$  une constante,  $(X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} C) \iff (X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} C)$ .

*Proof.* 1. • Si  $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X_\infty$  alors  $\exists \tilde{\Omega} \in \mathcal{A}$  vérifiant  $\mathbb{P}(\tilde{\Omega}) = 1$ , tel que pour tout  $\omega \in \tilde{\Omega}$ ,  $X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X_\infty(\omega)$ . Soit  $\varepsilon > 0$ . Alors par la formule des probabilités totales

$$\begin{aligned} \mathbb{P}(|X_n - X_\infty| \geq \varepsilon) &= \mathbb{P}(\{\omega \in \tilde{\Omega}, |X_n(\omega) - X_\infty(\omega)| \geq \varepsilon\}) + \mathbb{P}(\{\omega \notin \tilde{\Omega}, |X_n(\omega) - X_\infty(\omega)| \geq \varepsilon\}) \\ &\leq \mathbb{P}(\{\omega \in \tilde{\Omega}, |X_n(\omega) - X_\infty(\omega)| \geq \varepsilon\}) + 1 - \mathbb{P}(\tilde{\Omega}) \leq \mathbb{P}(\{\omega \in \tilde{\Omega}, |X_n(\omega) - X_\infty(\omega)| \geq \varepsilon\}). \end{aligned}$$

Mais comme  $X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X_\infty(\omega)$  pour  $\omega \in \tilde{\Omega}$ , alors il existe  $N \in \mathbb{N}$ , tel que pour tout  $n \geq N$ ,  $|X_n(\omega) - X_\infty(\omega)| < \varepsilon$ .

Donc pour  $n \geq N$ ,  $\{\omega \in \tilde{\Omega}, |X_n(\omega) - X_\infty(\omega)| \geq \varepsilon\} = \emptyset$  soit  $\mathbb{P}(|X_n - X_\infty| \geq \varepsilon) = 0$ : d'où la convergence en probabilité.

•  $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^p} X_\infty$ , alors  $\mathbb{E}[|X_n|^p + |X_\infty|^p] < \infty$  et  $\mathbb{E}[|X_n - X_\infty|^p] \xrightarrow[n \rightarrow \infty]{} 0$ . Pour  $\varepsilon > 0$ ,

$$\mathbb{P}(|X_n - X_\infty| \geq \varepsilon) = \mathbb{P}(|X_n - X_\infty|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X_\infty|^p]}{\varepsilon^p} \xrightarrow[n \rightarrow \infty]{} 0,$$

en utilisant l'Inégalité de Markov, d'où la convergence en probabilités.

2. On suppose  $X_n \xrightarrow{\mathcal{P}} X_\infty$  donc pour  $\eta > 0$  quelconque,  $\mathbb{P}(|X_n - X_\infty| \geq \eta) \xrightarrow{n \rightarrow \infty} 0$ . Par ailleurs, pour  $x \in \mathbb{R}$  tel que  $F_{X_\infty}$  est continue en  $x$ , donc pour tout  $\varepsilon > 0$  alors il existe  $\eta_\varepsilon > 0$  tel que  $|F_{X_\infty}(x - \eta_\varepsilon) - F_{X_\infty}(x)| \leq \varepsilon$  et  $|F_{X_\infty}(x + \eta_\varepsilon) - F_{X_\infty}(x)| \leq \varepsilon$ . Mais:

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \leq x) \\ &= \mathbb{P}(X_\infty \leq x - (X_n - X_\infty) \cap |X_n - X_\infty| < \eta_\varepsilon) + \mathbb{P}(X_\infty \leq x - (X_n - X_\infty) \cap |X_n - X_\infty| \geq \eta_\varepsilon) \\ &\leq \mathbb{P}(X_\infty \leq x + \eta_\varepsilon) + \mathbb{P}(|X_n - X_\infty| \geq \eta_\varepsilon). \end{aligned}$$

En minorant de la même manière, on obtient donc:

$$F_{X_\infty}(x) - \varepsilon - \mathbb{P}(|X_n - X_\infty| \geq \eta_\varepsilon) \leq F_{X_n}(x) \leq F_{X_\infty}(x) + \varepsilon + \mathbb{P}(|X_n - X_\infty| \geq \eta_\varepsilon).$$

Comme  $\mathbb{P}(|X_n - X_\infty| \geq \eta_\varepsilon) \xrightarrow{n \rightarrow \infty} 0$ , on en déduit qu'il existe  $N \in \mathbb{N}$  tel que pour tout  $n \geq N$ ,  $\mathbb{P}(|X_n - X_\infty| \geq \eta_\varepsilon) \leq \varepsilon$ , donc pour  $n \geq N$ ,  $|F_{X_n}(x) - F_{X_\infty}(x)| \leq 2\varepsilon$ : on obtient bien la convergence en loi.

3. On utilise  $\|X_n - X_\infty\|_p \leq \|X_n - X_\infty\|_q$ , donc  $\mathbb{E}[|X_n - X_\infty|^p] \leq \mathbb{E}[|X_n - X_\infty|^q]^{p/q}$ .
4. Voir exercice. □

**Exemple instructif:** Sur  $([0, 1], \mathcal{B}([0, 1]), \mathcal{U}([0, 1]))$  on définit la suite  $(X_n)$  telle que pour  $\omega \in [0, 1]$ ,  $X_n(\omega) = (n+1)\omega^n$ . On montre alors que  $X_n \xrightarrow{\mathcal{L}} 0$ ,  $X_n \xrightarrow{\mathcal{P}} 0$  et  $X_n \xrightarrow{p.s.} 0$ , mais  $(X_n)$  ne converge pas dans  $\mathbb{L}^1$  (si elle convergeait en ce sens elle convergerait également en probabilité et ce serait nécessairement vers 0).

**Propriété.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a. définie sur le même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $X_\infty$  une v.a. définie également sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Soit  $g: \mathbb{R} \rightarrow \mathbb{R}$  est une fonction borélienne continue sur  $C \subset \mathbb{R}$  telle que  $\mathbb{P}(X_\infty \in C) = 1$ . Alors  $(X_n \xrightarrow{\mathcal{P}} X_\infty) \implies (g(X_n) \xrightarrow{\mathcal{P}} g(X_\infty))$ . Même chose pour les convergences en loi et p.s. .

*Proof.* • On fixe  $\varepsilon > 0$ . Pour tout  $\delta > 0$ , on définit

$$B_{\delta, \varepsilon} = \{x \in \mathbb{R} \cap C, \exists y \in \mathbb{R}, |x - y| \leq \delta \text{ et } |g(x) - g(y)| > \varepsilon\}.$$

Comme  $g$  est continue sur  $C$ , il est clair que  $\lim_{\delta \rightarrow 0^+} B_{\delta, \varepsilon} = \emptyset$ . En utilisant la formule des probabilités totales, on peut écrire:

$$\begin{aligned} \mathbb{P}(|g(X_n) - g(X_\infty)| > \varepsilon) &= \mathbb{P}(|g(X_n) - g(X_\infty)| > \varepsilon \cap X_\infty \in B_{\delta, \varepsilon}) + \mathbb{P}(|g(X_n) - g(X_\infty)| > \varepsilon \cap X_\infty \notin B_{\delta, \varepsilon}) \\ &\leq \mathbb{P}(X_\infty \in B_{\delta, \varepsilon}) + \mathbb{P}(|X_n - X_\infty| > \delta \cap X_\infty \notin B_{\delta, \varepsilon}) + \mathbb{P}(X_\infty \notin C) \leq \mathbb{P}(X_\infty \in B_{\delta, \varepsilon}) + \mathbb{P}(|X_n - X_\infty| > \delta), \end{aligned}$$

car  $\mathbb{P}(X_\infty \notin C) = 0$ . Comme  $\lim_{\delta \rightarrow 0^+} B_{\delta, \varepsilon} = \emptyset$  on peut toujours choisir  $\delta$  suffisamment petit pour que  $\mathbb{P}(X_\infty \in B_{\delta, \varepsilon})$  soit aussi petit que l'on veut, et comme  $X_n \xrightarrow{\mathcal{P}} X_\infty$ , on peut choisir  $n$  suffisamment grand pour que  $\mathbb{P}(|X_n - X_\infty| > \delta)$  soit aussi petit que l'on veut. Donc  $g(X_n) \xrightarrow{\mathcal{P}} g(X_\infty)$ .

•  $X_n \xrightarrow{\mathcal{L}} X_\infty$  si et seulement si pour toute fonction  $h$  continue bornée  $\mathbb{E}[h(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(X_\infty)]$ . Comme  $h \circ g$  est aussi continue bornée, on a également  $\mathbb{E}[h(g(X_n))] \xrightarrow{n \rightarrow \infty} \mathbb{E}[h(g(X_\infty))]$ .

• Si  $X_n \xrightarrow{p.s.} X_\infty$ , alors  $\exists \tilde{\Omega} \in \mathcal{A}$  vérifiant  $\mathbb{P}(\tilde{\Omega}) = 1$ , tel que pour tout  $\omega \in \tilde{\Omega}$ ,  $X_n(\omega) \xrightarrow{n \rightarrow \infty} X_\infty(\omega)$ . Donc pour tout  $\omega \in \tilde{\Omega}$ ,  $g(X_n(\omega)) \xrightarrow{n \rightarrow \infty} g(X_\infty(\omega))$  d'après la caractérisation de la continuité par les suites numériques. D'où  $g(X_n) \xrightarrow{p.s.} g(X_\infty)$ . □

**Lemme (Lemme de Slutsky (Slutsky, 1915)).** Soit  $(X_n)$  une suite de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et  $X_\infty$  une v.a. définie également sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $X_n \xrightarrow{\mathcal{L}} X_\infty$ . Soit  $(Y_n)$  une suite de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $Y_n \xrightarrow{\mathcal{P}} c$ , où  $c \in \mathbb{R}$ . Alors  $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X_\infty, c)$ .

*Proof.* Soit  $u = (u_1, u_2) \in \mathbb{R}^2$ . En notant  $\phi_{(X_n, Y_n)}$  la fonction caractéristique de  $(X_n, Y_n)$ , on a:

$$\begin{aligned} |\phi_{(X_n, Y_n)}(u) - \phi_{(X_\infty, c)}(u)| &\leq |\phi_{(X_n, Y_n)}(u) - \phi_{(X_n, c)}(u)| + |\phi_{(X_n, c)}(u) - \phi_{(X_\infty, c)}(u)| \\ &\leq |\mathbb{E}[e^{iu_1 X_n} (e^{iu_2 Y_n} - e^{iu_2 c})]| + |\mathbb{E}[e^{iu_2 c} (e^{iu_1 X_n} - e^{iu_1 X_\infty})]| \\ &\leq |\mathbb{E}[e^{iu_2 Y_n} - e^{iu_2 c}]| + |\phi_{X_n}(u_1) - \phi_{X_\infty}(u_1)|. \end{aligned}$$

Mais  $\mathbb{E}[|e^{iu_2 Y_n} - e^{iu_2 c}|] \leq \sqrt{\mathbb{E}[|e^{iu_2 Y_n} - e^{iu_2 c}|^2]} \leq \sqrt{\mathbb{E}[2 - e^{iu_2(Y_n - c)} - e^{iu_2(c - Y_n)}]} \leq \sqrt{2 - \frac{\phi_{Y_n}(u_2)}{\phi_c(u_2)} - \frac{\phi_{Y_n}(-u_2)}{\phi_c(-u_2)}} \xrightarrow{n \rightarrow \infty} 0$

car  $\phi_{Y_n}(u_2) \xrightarrow{n \rightarrow \infty} \phi_c(u_2)$  pour tout  $u_2$ , puisque  $Y_n \xrightarrow{\mathcal{P}} c$  donc  $Y_n \xrightarrow{\mathcal{L}} c$ . Par ailleurs,  $|\phi_{X_n}(u_1) - \phi_{X_\infty}(u_1)| \xrightarrow{n \rightarrow \infty} 0$  pour tout  $u_1$  puisque  $X_n \xrightarrow{\mathcal{L}} X_\infty$ . En conséquence,  $|\phi_{(X_n, Y_n)}(u) - \phi_{(X_\infty, c)}(u)| \xrightarrow{n \rightarrow \infty} 0$ , dou le résultat. □

## 4.2 Théorèmes limites

Dans la suite, on va s'intéresser à la convergence de la moyenne empirique. On commence par obtenir les propriétés suivantes:

**Propriété.** Pour  $(X_1, \dots, X_n)$  une famille de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ .

- Si  $\mathbb{E}[|X_k|] < \infty$  pour tout  $k$ , alors  $\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k)$ ;
- Si  $\mathbb{E}[X_k^2] < \infty$  pour tout  $k$ ,

$$\text{var}(\bar{X}_n) = \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \text{cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

*Proof.* • On obtient l'existence de  $\mathbb{E}[|\bar{X}_n|]$  par l'inégalité triangulaire et sa valeur par linéarité de l'espérance.

- Comme  $\text{var}(\bar{X}_n) = \mathbb{E}[(\bar{X}_n - \mathbb{E}[\bar{X}_n])^2] = \frac{1}{n^2} \sum_{k=1}^n (X_k - \mathbb{E}[\bar{X}_n])^2$ , on obtient l'existence de  $\text{var}(\bar{X}_n)$  par l'inégalité triangulaire pour  $\|Z\|_2 = \mathbb{E}[Z^2]^{1/2}$ .  
De plus, par bilinéarité de la covariance, on a

$$\text{var}(\bar{X}_n) = \frac{1}{n^2} \text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \text{cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

□

**Conséquence.** Si  $(X_1, \dots, X_n)$  sont des v.a. **indépendantes** et **identiquement distribuées** (notées v.a.i.i.d.) alors

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}[X_1] \quad \text{si } \mathbb{E}[|X_1|] < \infty \quad \text{et} \quad \text{var}(\bar{X}_n) = \frac{1}{n} \text{var}(X_1) \quad \text{si } \mathbb{E}[X_1^2] < \infty.$$

*Proof.* La preuve est immédiate en utilisant le fait que  $\mathbb{E}[X_k] = \mathbb{E}[X_1]$  et  $\text{var}(X_k) = \text{var}(X_1)$  pour tout  $k$ , et le fait que l'indépendance entre les  $X_i$  implique que  $\text{cov}(X_i, X_j) = 0$  pour  $i \neq j$ . □

**Propriété (Inégalité de Markov (Tchebychev, 1867, Markov, 1884)).** Soit  $X$  une v.a. positive définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon} \quad (\text{valable si } \mathbb{E}[X] = \infty).$$

*Proof.* Soit  $\varepsilon > 0$ . On considère  $Y = X \mathbb{I}_{X \geq \varepsilon}$ . Alors  $Y$  est une v.a. positive et  $\mathbb{E}[Y] = \mathbb{E}[X \mathbb{I}_{X \geq \varepsilon}] \geq \mathbb{E}[\varepsilon \mathbb{I}_{X \geq \varepsilon}] = \varepsilon \mathbb{P}(X \geq \varepsilon)$ . Mais on a également  $Y \leq X$  par définition, donc  $\mathbb{E}[Y] \leq \mathbb{E}[X]$ . D'où le résultat. □

Voici une conséquence directe de cette inégalité:

**Propriété (Inégalité de Bienaymé-Tchebitchev (1853-1867)).** Soit  $X$  une v.a. définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[X^2] < \infty$ . Alors  $m = \mathbb{E}[X]$  et  $\sigma^2 = \text{var}(X)$ , pour tout  $\varepsilon > 0$ ,

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

*Proof.* On reprend l'Inégalité de Markov appliquée à  $(X - \mathbb{E}[X])^2$  qui est bien une v.a. positive et à  $\varepsilon^2$  plutôt que  $\varepsilon$ . Le résultat est obtenu puisque  $\mathbb{E}[(X - \mathbb{E}[X])^2] = \text{var}(X)$  et en remarquant que  $\mathbb{P}((X - \mathbb{E}[X])^2 \geq \varepsilon^2) = \mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon)$ . □

On va maintenant appliquer tout ce qui précède à la moyenne empirique d'une famille de variables aléatoires indépendantes ayant toute la même loi:

**Théorème (Loi Faible des Grands Nombres avec moment d'ordre 2 (Tchebychev, 1867)).** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.i.i.d. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[X_1^2] < \infty$ . Alors:

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \mathbb{E}[X_1].$$

*Proof.* Application directe de l'inégalité de Bienaymé-Tchebychev: pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}[X_1]| \geq \varepsilon) \leq \frac{\text{var}(\bar{X}_n)}{\varepsilon^2} = \frac{\text{var}(X_1)}{n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0,$$

d'où la convergence en probabilité de  $\bar{X}_n$ . □

**Théorème (Loi Faible des Grands Nombres avec moment d'ordre 1 (Khinchine, 1929)).**

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.i.i.d. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[|X_1|] < \infty$ . Alors:

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \mathbb{E}[X_1].$$

*Proof.* Sans perte de généralité on peut supposer que  $(X_i)$  est centré (sinon on considère  $X'_k = X_k - \mathbb{E}[X_1]$ ). Alors pour  $u \in \mathbb{R}$ , en passant par la fonction caractéristique on:

$$\phi_{\bar{X}_n}(u) = \mathbb{E}[e^{i \frac{u}{n} \sum_{k=1}^n X_k}] = \phi_{X_1}^n(u/n),$$

grâce à l'indépendance et le fait d'être identiquement distribué. Mais  $|\phi_{X_1}(u/n)| \leq 1$  comme toute fonction caractéristique et on peut effectuer un développement limité de Lagrange d'ordre 1 en 0 de  $\phi_{X_1}(u/n)$ , sachant que puisque  $\mathbb{E}[|X_1|] < \infty$  alors  $\phi_{X_1}$  est de classe  $\mathcal{C}^1$  sur  $\mathbb{R}$ :

$$\phi_{X_1}(u/n) = \phi_{X_1}(0) + \frac{u}{n} \phi' \left( \frac{v}{n} \right) = 1 + \frac{u}{n} \phi' \left( \frac{v}{n} \right),$$

avec  $|v| \leq |u|$ . Mais  $\phi'(0) = i \mathbb{E}[X_1] = 0$ , donc par continuité de  $\phi'$  en 0 alors  $\phi' \left( \frac{v}{n} \right) \xrightarrow[n \rightarrow \infty]{} 0$ .

On peut alors utiliser le résultat suivant (voir Exercice 9 TD4): si  $(z_j)$  et  $(z'_j)$  sont deux familles de nombres complexes tels que  $|z_j| \leq 1$  et  $|z'_j| \leq 1$  pour tout  $j$ , alors

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n z'_j \right| \leq \sum_{j=1}^n |z_j - z'_j|.$$

En prenant ici  $z_j = \phi_{X_1}(u/n)$  et  $z'_j = 1$  pour tout  $j$ , on aboutit à:

$$|\phi_{X_1}^n(u/n) - 1| \leq \sum_{j=1}^n \left| 1 + \frac{u}{n} \phi' \left( \frac{v}{n} \right) - 1 \right| = n \phi' \left( \frac{v}{n} \right) \xrightarrow[n \rightarrow \infty]{} 0.$$

De ceci on en déduit que pour tout  $u \in \mathbb{R}$ ,  $\phi_{\bar{X}_n}(u) \xrightarrow[n \rightarrow \infty]{} 1$ . Or 1 est la fonction caractéristique de la variable aléatoire qui vaut 0. D'où  $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} 0$  et comme c'est une convergence vers une constante  $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 0$ . □

**Théorème (Loi forte des Grands Nombres (Kolmogorov, 1940)).** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.i.i.d. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ . Alors

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{p.s.} m \iff \mathbb{E}[|X_1|] < \infty \text{ et } m = \mathbb{E}[X_0].$$

*Proof.* Trop difficile, voir cours de M1 MAEF. □

Il est possible d'être plus précis quant au comportement asymptotique de la moyenne empirique autour de son espérance: c'est ce que précise le théorème suivant, à savoir que ce comportement est gaussien et se resserre à vitesse  $1/\sqrt{n}$  autour de l'espérance.

**Théorème (Théorème de la limite centrale (Lindeberg et Lévy, ~ 1920)).** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.i.i.d. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[X_0^2] < \infty$ . Alors:

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{\sigma} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{où } m = \mathbb{E}[X_0] \text{ et } \sigma^2 = \text{var}(X_0).$$

*Proof.* En premier lieu, quitte à considérer  $X'_k = (X_k - m)\sigma$ , on se place sans perte de généralité dans le cas où les  $(X_k)$  sont centrées et de variance 1. On considère alors  $Z_n = \sqrt{n}\bar{X}_n$ . Comme les  $(X_k)$  sont des v.a.i.i.d., on a pour tout  $u \in \mathbb{R}$  (voir preuve précédente),

$$\phi_{Z_n}(u) = \mathbb{E}\left[e^{i \frac{u}{\sqrt{n}} \sum_{k=1}^n X_k}\right] = \phi_{X_1}^n(u/\sqrt{n}).$$

Comme  $\mathbb{E}[X_1^2] < \infty$ , alors  $\phi_{X_1}$  est de classe  $\mathcal{C}^2$  sur  $\mathbb{R}$ , et un développement limité d'ordre 2 en 0 de  $\phi_{X_1}(u/\sqrt{n})$  donne:

$$\phi_{X_1}(u/\sqrt{n}) = \phi_{X_1}(0) + \phi'_{X_1}(0) \frac{u}{\sqrt{n}} + \phi''_{X_1}(0) \frac{u^2}{2n} (1 + \varepsilon_n) = 1 - \frac{u^2}{2n} (1 + \varepsilon_n),$$

car  $\phi'_{X_1}(0) = i \mathbb{E}[X_1] = 0$  et  $\phi''_{X_1}(0) = -\mathbb{E}[X_1^2] = -1$ . Comme dans la preuve précédente on utilise le fait  $\left| \prod_{j=1}^n z_j - \prod_{j=1}^n z'_j \right| \leq \sum_{j=1}^n |z_j - z'_j|$  avec cette fois-ci  $z_j = \phi_{X_1}(u/\sqrt{n})$  et  $z'_j = 1 - \frac{u^2}{2n}$ , et on obtient que:

$$|\phi_{X_1}^n(u/\sqrt{n}) - (1 - \frac{u^2}{2n})^n| \leq \sum_{j=1}^n \frac{u^2}{2n} |\varepsilon_n| \leq \frac{u^2}{2} |\varepsilon_n| \xrightarrow{n \rightarrow \infty} 0.$$

Or  $(1 - \frac{u^2}{2n})^n = e^{n \ln(1 - \frac{u^2}{2n})} \xrightarrow{n \rightarrow \infty} e^{-u^2/2}$ . On en déduit donc que  $\phi_{Z_n}(u) \xrightarrow{n \rightarrow \infty} e^{-u^2/2}$  pour tout  $u \in \mathbb{R}$ , et  $e^{-u^2/2}$  est la fonction caractéristique de la loi  $\mathcal{N}(0, 1)$ .  $\square$

**Théorème (Second théorème de la limite centrale).** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.i.i.d. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  telle que  $\mathbb{E}[X_0^2] < \infty$ . Alors :

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{\hat{\sigma}_n} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{avec } m = \mathbb{E}[X_0] \text{ et } \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

*Proof.* On montre d'abord que  $\hat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \sigma^2$  (on utilise la LGN pour  $\frac{1}{n} \sum_{k=1}^n X_k^2$ , et comme  $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} m$ , donc  $\bar{X}_n^2 \xrightarrow[n \rightarrow \infty]{\mathcal{P}} m^2$  on utilise le Lemme de Slutsky pour montrer que  $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n X_k^2 - \bar{X}_n^2 \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \sigma^2$ ). Donc  $\frac{\sigma}{\hat{\sigma}_n} \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 1$  (fonction continue). Or

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{\hat{\sigma}_n} \right) = \frac{\sigma}{\hat{\sigma}_n} \sqrt{n} \left( \frac{\bar{X}_n - m}{\sigma} \right).$$

En utilisant le Lemme de Slutsky, comme  $\frac{\sigma}{\hat{\sigma}_n} \xrightarrow[n \rightarrow \infty]{\mathcal{P}} 1$  et  $\sqrt{n} \left( \frac{\bar{X}_n - m}{\sigma} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ , on obtient le résultat demandé.  $\square$

**Exemple:** Application de ce TLC pour les v.a. de Bernoulli de paramètre  $p$ .

**Remarque:** Dans le cas où les v.a. sont gaussiennes, alors on a pour tout  $n \geq 2$ ,

$$\sqrt{n} \left( \frac{\bar{X}_n - m}{\hat{\sigma}_n} \right) \stackrel{\mathcal{L}}{\sim} t(n-1).$$

On en déduit donc également que  $t(n-1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ .

Un dernier théorème limite, souvent utile en statistique, peut être énoncé. Il s'apparente à une formule de Taylor:

**Théorème (Delta-méthode (Kelley, 1928)).** Soit  $(Z_n)_{n \in \mathbb{N}}$  une suite de v.a. définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$  et telle que  $a_n(Z_n - m) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ , avec  $a_n \xrightarrow[n \rightarrow \infty]{} \infty$  et  $m \in \mathbb{R}$ . Alors, pour toute fonction  $g$  de classe  $\mathcal{C}^1$  dans un voisinage de  $m$  telle que  $g'(m) \neq 0$ ,

$$a_n (g(Z_n) - g(m)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, (g'(m))^2).$$

*Proof.* D'après le développement de Taylor-Lagrange, il existe une variable aléatoire  $\lambda$  à valeurs dans  $[0, 1]$  telle que

$$a_n (g(Z_n) - g(m)) = g'(\lambda Z_n + (1 - \lambda)m) a_n (Z_n - m).$$

Le fait que l'on ait  $a_n(Z_n - m) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$  implique que  $Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} m$  et donc  $\lambda Z_n + (1 - \lambda)m \xrightarrow[n \rightarrow \infty]{\mathcal{P}} \lambda m + (1 - \lambda)m = m$ .

Comme  $g'$  est supposée être une fonction continue implique que  $g'(\lambda Z_n + (1 - \lambda)m) \xrightarrow[n \rightarrow \infty]{\mathcal{P}} g'(m)$ . Enfin, grâce au Lemme de Slutsky,  $a_n (g(Z_n) - g(m)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} g'(m) \mathcal{N}(0, 1)$ .  $\square$

**Exemple d'application:** Si on a le TLC pour  $\bar{X}_n$  et si  $g$  de classe  $\mathcal{C}^1$  dans un voisinage  $\mathbb{E}[X_0]$ , alors:

$$\sqrt{n}(g(\bar{X}_n) - g(\mathbb{E}[X_0])) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, (g'(\mathbb{E}[X_0]))^2 \text{var}(X_0)).$$

A utiliser notamment pour obtenir un TLC pour  $1/\bar{X}_n$  dans le cas de v.a.i.i.d. de loi exponentielle de paramètre  $\lambda$ , ce qui fournit un TLC pour l'estimateur par maximum de vraisemblance de  $\lambda$ .