



Université Ouaga 3S

Faculté des Sciences et Technologie

Année universitaire : 2024/2025

LICENCE SCIENCES, TECHNOLOGIE (LST)

MENTION MATHÉMATIQUES-INFORMATIQUE (MI)

PARCOURS MATHÉMATIQUES APPLIQUÉES AUX SCIENCES (MAS)

SEMESTRE 6 : S6

Support de Cours de PROBABILITE

VOLUMEHoraire : 72 H

CM : 48 h

TD : 24 h

Table des matières

1	Tribus et mesures	3
1.1	Tribus et mesures	3
1.1.1	Tribus	3
1.1.2	Mesures	4
1.1.3	Ensembles Négligeables	6
1.2	Premiers exemples	6
1.2.1	Les mesures discrètes	6
1.2.2	La mesure de Lebesgue	7
1.2.3	Les mesures à densité par rapport à la mesure de Lebesgue	8
1.3	Probabilité conditionnelle et événements indépendants	8
1.3.1	Probabilité conditionnelle	9
1.3.2	Événements indépendants	9
1.3.3	Lemme de Borel-Cantelli	10
2	Fonctions mesurables et variables aléatoires	
	Intégrale de Lebesgue et espérance	13
2.1	Fonctions mesurables et variables aléatoires	13
2.1.1	Fonctions mesurables	13
2.1.2	Variables aléatoires	14
2.2	Intégrale de Lebesgue	15
2.2.1	Construction de l'intégrale de Lebesgue	16
2.2.2	Propriétés générales	17
2.2.3	Calcul dans des cas particuliers	18
2.3	Espérance d'une variable aléatoire et théorème du transport	21
3	Théorèmes d'intégration	27
3.1	Théorèmes de convergence	27
3.1.1	Théorème de convergence monotone	27
3.1.2	Théorème de convergence dominée	29
3.1.3	Lemme de Fatou	31
3.2	Applications des théorèmes de convergence aux intégrales paramétrées	33
3.3	Théorème de changement de variable	36
4	Fonction de répartition et fonction caractéristique	39
4.1	Fonction de répartition	39
4.1.1	Définition pour une v.a.r. et propriétés générales	39
4.1.2	Variables aléatoires réelles discrètes	41
4.1.3	Variables aléatoires réelles absolument continues	43
4.1.4	Mélange de lois discrètes et absolument continues	45
4.1.5	Généralisation aux vecteurs aléatoires	46
4.2	Fonction caractéristique	47

5	Théorème de Fubini	51
5.1	Tribu produit	51
5.2	Mesure Produit	52
5.3	Théorèmes de Fubini	53
5.4	Applications des théorèmes de Fubini	55
5.4.1	Lois marginales d'un vecteur aléatoire	55
5.4.2	Échange des signes somme et somme	57
5.4.3	Échange des signes somme et intégrale	57
5.4.4	Échange des signes espérance et somme	58
5.4.5	Échange des signes espérance et intégrale	59
6	Indépendance de variables aléatoires	61
6.1	Indépendance de variables aléatoires	61
6.2	Caractérisations et conséquences	63
6.2.1	Indépendance et calcul d'espérances	63
6.2.2	Indépendance et fonctions de répartition	65
6.2.3	Indépendance et fonctions caractéristiques	66
6.2.4	Cadre des lois discrètes et des lois absolument continues	67
6.3	Addition de variables aléatoires indépendantes	69
7	Variance, moments et espaces L^p	73
7.1	Variance et covariance	73
7.2	Moments d'une variable aléatoire	74
7.2.1	Calcul pour les lois binomiales	75
7.2.2	Calcul pour les lois de Poisson	76
7.2.3	Calcul pour les lois géométrique	76
7.2.4	Calcul pour la loi uniforme sur $[a, b]$	77
7.2.5	Calcul pour les lois exponentielles	77
7.2.6	Calcul pour les lois gaussiennes	78
7.3	Inégalités classiques	78
7.4	Les espaces L^p , $p \geq 1$	80
7.4.1	Définitions et premières propriétés	80
7.4.2	L'espace L^2	81
8	Suites de variables aléatoires	87
8.1	Convergence presque sûre	87
8.1.1	Définition et propriétés	87
8.2	Convergence en probabilité	91
8.3	Convergence L^p	94
8.4	Convergence en loi	97
8.4.1	Définition et premières propriétés	97
8.4.2	Caractérisations	98
8.5	Somme de variables aléatoires indépendantes	99
8.5.1	Loi des grands nombres	99
8.5.2	Théorème central limite	100
8.6	Tableau récapitulant les liens entre les convergence	100
9	Vecteurs gaussiens	101
9.1	Définition et première propriétés	101
9.1.1	Définition	101
9.1.2	Fonction caractéristique d'un vecteur gaussien	102
9.2	Construction et existence des vecteurs gaussiens	103

9.3	Propriétés de la loi d'un vecteur gaussien	106
9.3.1	Vecteurs gaussiens et indépendance	106
9.3.2	Vecteur gaussien et densité	107
9.3.3	Cas $d = 2$	107
9.4	Vecteurs gaussiens et convergence en loi	110
9.4.1	Stabilité de l'ensemble des vecteurs gaussiens	110
9.4.2	Théorème central limite	111
9.5	Vecteurs gaussiens et projection orthogonale dans L^2	111
A	Lois classiques	113
B	Simulation d'une variable aléatoire	115
B.1	Simulation d'une variable aléatoire de loi uniforme en Matlab et en Python	115
B.2	Simulation d'une loi gaussienne par la méthode de Box-Muller	116
B.3	Simulation d'une variable aléatoire discrète	118
B.3.1	Un premier exemple : v.a. de Bernoulli	118
B.3.2	Variables aléatoires discrètes générales	118
B.4	Simulation d'une v.a à partir de sa fonction de répartition	122
B.4.1	Fonction de répartition réciproque généralisée	122
B.4.2	Méthode de la transformée réciproque généralisée	124
B.5	Méthode de rejet	126

Chapitre 1

Tribus et mesures

Dans tout le chapitre, Ω est un ensemble non vide. Notre premier objectif est de définir la notion de mesure sur Ω . En particulier, nous aboutissons à un cadre d'abstraction commun pour les notions de cardinal, de longueur, d'aire et de volume. Nous aborderons également la notion fondamentale d'indépendance d'événements. Notons que les concepts abordés ici sont une généralisation du programme de probabilités de classe préparatoire.

1.1 Tribus et mesures

1.1.1 Tribus

Une tribu est une famille qui contient les sous-ensembles de Ω que l'on pourra *mesurer*.

Définition 1.1.

Une **tribu** $\mathcal{F}$ sur l'ensemble Ω est une famille de sous-ensembles de Ω telle que

- (i) $\Omega \in \mathcal{F}$,
- (ii) $\mathcal{F}$ est stable par passage au complémentaire, c'est-à-dire que
si $A \in \mathcal{F}$, alors $\Omega \setminus A \in \mathcal{F}$,
- (iii) $\mathcal{F}$ est stable par réunion dénombrable, c'est-à-dire que

si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Si $\mathcal{F}$ est une tribu sur Ω , l'espace $(\Omega, \mathcal{F})$ est appelé un **espace mesurable** (ou espace probabilisable). Les éléments de $\mathcal{F}$ (qui sont des sous-ensembles de Ω) sont appelés les ensembles **$\mathcal{F}$ -mesurables**, ou simplement mesurables s'il n'y pas d'ambiguïté sur la tribu considérée.

Remarque 1.1.

1. Une tribu contient toujours l'ensemble vide (noté $\emptyset$) et est stable par réunion finie.
2. Nous aurions pu définir la notion de tribu en remplaçant
 - l'assertion (i) par l'assertion (i)' $\emptyset \in \mathcal{F}$.
 - et/ou l'assertion (iii) par l'assertion (iii)' $\mathcal{F}$ est stable par intersection dénombrable.
3. A fortiori une tribu est stable par union/intersection finie.

Exemple 1.1. Soit Ω un ensemble non vide.

1. $\mathcal{F} = \mathcal{P}(\Omega)$ est la plus grande tribu (au sens de l'inclusion) sur Ω . Elle est appelée **tribu discrète sur Ω** .
2. $\mathcal{F} = \{\emptyset, \Omega\}$ est la plus petite tribu (au sens de l'inclusion) sur Ω . Elle est appelée **tribu grossière sur Ω** .
3. Soit $A \in \mathcal{P}(\Omega)$ tel que $A \neq \emptyset$ et $A \neq \Omega$. Alors $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ est une tribu sur Ω .
4. Étant donnée une famille $\mathcal{S}$ de sous-ensembles de Ω , on définit la **tribu engendrée** par $\mathcal{S}$, notée $\sigma(\mathcal{S})$, comme l'intersection de toutes les tribus qui contiennent $\mathcal{S}$:

$$\sigma(\mathcal{S}) := \bigcap_{\mathcal{F} \text{ tribu, } \mathcal{S} \subset \mathcal{F}} \mathcal{F}.$$

Il s'agit de la plus petite tribu (au sens de l'inclusion) qui contient tous les éléments de $\mathcal{S}$. En particulier, la tribu engendrée par $\mathcal{S} = \{A\}$ est $\{\emptyset, A, A^c, \Omega\}$.

5. Considérons le cas $\Omega \subset \mathbb{R}^d$. Comme nous souhaitons étendre la notion de longueur, d'aire et de volume, il est intéressant de définir la plus petite tribu contenant toutes les boules de Ω . Il s'agit en fait de la tribu engendrée par les ouverts de Ω , appelée **tribu borélienne et notée $\mathcal{B}(\Omega)$** . Les ensembles $\mathcal{B}(\Omega)$ -mesurables sont appelés les ensembles boréliens.

Remarque 1.2. Dans tout ce cours, sauf mention du contraire, si Ω est un espace non vide fini ou dénombrable, nous le munirons de la tribu discrète $\mathcal{P}(\Omega)$.

1.1.2 Mesures

Définition 1.2.

Soit $(\Omega, \mathcal{F})$ un espace mesurable. Une **mesure positive** μ sur l'espace mesurable $(\Omega, \mathcal{F})$ est une application $\mu : \mathcal{F} \rightarrow [0, +\infty]$ telle que

- (i) $\mu(\emptyset) = 0$,
- (ii) μ est **σ -additive**, c'est-à-dire que pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{F}$ deux à deux disjoints

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{n \in \mathbb{N}} \mu(A_n). \quad (1.1)$$

Le triplet $(\Omega, \mathcal{F}, \mu)$ est alors appelé un **espace mesuré**. Si $\mu(\Omega) = 1$, alors la mesure μ est appelée **probabilité** et le triplet $(\Omega, \mathcal{F}, \mu)$ **espace de probabilités**.

Il est courant de noter les espaces probabilisés sous la forme $(\Omega, \mathcal{F}, \mathbb{P})$, où $\mathbb{P}$ est alors une mesure de probabilité. Dans ce cas, il est également d'usage d'appeler *événements* les ensembles mesurables. On interprète alors Ω comme l'ensemble des réalisations possibles et un événement regroupe différentes réalisations présentant une certaine caractéristique.

Exemple 1.2. Nous cherchons à modéliser le lancer de deux dés à 6 faces. Nous posons pour cela $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$, muni de la tribu discrète et nous définissons la mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ par

$$\mathbb{P}(A) = \text{Card}(A)/36, \quad \forall A \in \mathcal{P}(\Omega).$$

Ici, chaque élément $\omega = (\omega_1, \omega_2) \in \Omega$ représente un tirage possible (le premier dé de valeur ω_1 et le deuxième dé de valeur ω_2). Voici quelques exemples d'événements et leurs probabilités associées.

1. $A_1 = \{\omega \in \Omega \text{ tel que } \omega_1 = \omega_2\}$, avec $\mathbb{P}(A_1) = 1/6$,
2. $A_2 = \{\omega \in \Omega \text{ tel que } \omega_1 + \omega_2 = 5\}$, avec $\mathbb{P}(A_2) = 1/9$,
3. $A_3 = \{\omega \in \Omega \text{ tel que } \omega_1 + \omega_2 \neq 5\}$, avec $\mathbb{P}(A_3) = \mathbb{P}(A_2^c) = \mathbb{P}(\Omega) - \mathbb{P}(A_2) = 8/9$.

Les trois premières propriétés de la proposition suivante sont faciles à démontrer. Il est donc conseillé au lecteur d'établir ces résultats en guise d'exercice.

Proposition 1.1.

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré.

1. La mesure μ est additive, c'est-à-dire que si $A_0, \dots, A_N \in \mathcal{F}$ sont deux à deux disjoints, alors

$$\mu\left(\bigcup_{n=0}^N A_n\right) = \sum_{n=0}^N \mu(A_n).$$

2. La mesure μ est croissante sur $\mathcal{F}$ (pour l'inclusion), c'est-à-dire que pour tous $A, B \in \mathcal{F}$,

$$A \subset B \implies \mu(A) \leq \mu(B).$$

3. Pour tous $A, B \in \mathcal{F}$,

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \quad (1.2)$$

4. La mesure μ est σ -sous-additive, c'est-à-dire que si I est un ensemble fini ou dénombrable et si $(A_i)_{i \in I}$ est une famille d'éléments de $\mathcal{F}$, alors

$$\mu\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mu(A_i).$$

5. La mesure μ vérifie les deux propriétés suivantes appelées continuité monotone respectivement croissante et décroissante.

- (a) Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'éléments de $\mathcal{F}$ (c'est-à-dire si $A_n \in \mathcal{F}$ et $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$) alors la suite $(\mu(A_n))_{n \in \mathbb{N}}$ est une suite croissante et

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

- (b) Si $(B_n)_{n \in \mathbb{N}}$ est une suite décroissante d'éléments de $\mathcal{F}$ (c'est-à-dire si $B_n \in \mathcal{F}$ et $B_{n+1} \subset B_n$ pour tout $n \in \mathbb{N}$) telle que $\mu(B_0) < +\infty$, alors la suite $(\mu(B_n))_{n \in \mathbb{N}}$ est une suite décroissante et

$$\mu\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow +\infty} \mu(B_n).$$

Remarque 1.3. Supposons que $\mu(\Omega) < +\infty$. Alors par croissance de μ , pour tout $A \in \mathcal{F}$, $\mu(A) < +\infty$ (autrement dit μ est à valeurs dans $[0, +\infty[$). Dès lors, dans l'égalité (1.2), il est possible de faire passer un terme du membre de droite (resp. gauche) dans le membre de gauche (resp. droite). En particulier, cette égalité se réécrit sous la forme :

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

De plus, en prenant $B = A^c$ et en appliquant (1.2), nous obtenons que pour tout $A \in \mathcal{F}$, $\mu(A^c) = \mu(\Omega) - \mu(A)$. En particulier, si μ est une probabilité, alors pour tout $A \in \mathcal{F}$,

$$\mu(A^c) = 1 - \mu(A).$$

Remarquons enfin qu'une somme pondérée et dénombrable de mesures est une mesure.

1.1.3 Ensembles Négligeables

Jetons une pièce de monnaie équilibrée une infinité de fois. L'évènement « la pièce tombe toujours sur face » est un évènement non vide de probabilité nulle ; il s'agit d'un exemple d'évènement négligeable. La notion d'ensembles négligeables joue un rôle important en théorie de la mesure et en probabilités. La définition suivante précise cette notion.

Définition 1.3.

Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré.

1. Un sous-ensemble $N \subset \Omega$ est dit **μ -négligeable** (ou simplement négligeable) s'il est inclus dans un ensemble mesurable de mesure nulle, c'est-à-dire si

$$\exists A \in \mathcal{F} \text{ tel que } N \subset A \text{ et } \mu(A) = 0.$$

2. Une propriété Π , dépendant de $\omega \in \Omega$, est dite vraie **μ -presque partout** si l'ensemble

$$\{\omega \in \Omega / \Pi(\omega) \text{ est fausse}\}$$

est μ -négligeable.

3. Si μ est une probabilité, une propriété vraie μ -presque partout est dite vraie **μ -presque sûrement**.

On observe que tout sous-ensemble d'un ensemble négligeable est négligeable. De plus, on montre facilement à l'aide de la proposition 1.1 que la famille des ensembles négligeables est stable par réunion finie ou dénombrable.

Exemple 1.3.

1. L'ensemble vide est négligeable.
2. Soient $A \subset \mathbb{R}^d$ un ensemble fini ou dénombrable et λ_d la mesure de Lebesgue sur $\mathbb{R}^d$. L'ensemble A est alors λ_d -négligeable.
3. Une droite de $\mathbb{R}^2$ est toujours négligeable par rapport à λ_2 . De manière générale, les hyperplans affines de $\mathbb{R}^d$ sont négligeables par rapport à λ_d .
4. La mesure λ_1 est une mesure de probabilité sur $[0, 1]$. Les éléments de $[0, 1]$ sont irrationnels λ_1 -presque sûrement.

Remarque 1.4. Dans la suite du cours, nous supposerons très souvent que les ensembles μ -négligeables sont $\mathcal{F}$ -mesurables (on dit alors $(\Omega, \mathcal{F}, \mu)$ est **complet**). Cette hypothèse a priori cavalière est rendue possible en *ajoutant* à $\mathcal{F}$ les ensembles négligeables, auquel cas on dira que la tribu est **complétée** (par rapport à μ). Dans ce cas, il est possible d'étendre la définition de μ à la tribu complétée.

1.2 Premiers exemples

1.2.1 Les mesures discrètes

Cette partie présente un exemple classique et fondamental de mesures : les mesures discrètes. Un autre exemple important sera donné dans la partie suivante (la mesure de Lebesgue).

Soit $a \in \Omega$. La **mesure de Dirac en a** , notée δ_a , est la mesure sur $(\Omega, \mathcal{P}(\Omega))$ définie par

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

Soient $(a_n)_{n \in \mathbb{N}}$ une famille d'éléments de Ω et $(\alpha_i)_{i \in \mathbb{N}}$ une famille d'éléments de $[0, +\infty]$. Alors en adoptant la convention $+\infty \times 0 = 0$, pour tout ensemble $A \in \mathcal{P}(\Omega)$,

$$\mu(A) = \sum_{n \in \mathbb{N}} \alpha_n \delta_{a_n}(A) = \sum_{n \in \mathbb{N} / a_n \in A} \alpha_n \quad (1.3)$$

est bien défini et l'application $\mu : \mathcal{P}(\Omega) \rightarrow [0, +\infty]$ est une mesure positive sur $(\Omega, \mathcal{P}(\Omega))$. Une telle mesure est appelée une **mesure discrète**.¹

Remarque 1.5. La mesure discrète $\mu = \sum_{n \in \mathbb{N}} \alpha_n \delta_{a_n}$ est une probabilité si et seulement si $\sum_{n \in \mathbb{N}} \alpha_n = 1$. On parle alors de probabilité discrète.

Exemple 1.4.

1. Pour tout $a \in \Omega$, la mesure de Dirac en a , notée δ_a , est une probabilité discrète.
2. Soit Ω un ensemble non vide fini ou dénombrable. Remarquons alors que pour tout $A \in \mathcal{P}(\Omega)$,

$$\text{Card}(A) = \sum_{\omega \in \Omega} \delta_\omega(A).$$

L'ensemble Ω étant fini ou dénombrable, Card s'écrit sous la forme (1.3) et est donc une mesure discrète sur Ω , appelée **mesure de comptage sur Ω** .

3. Si l'ensemble non vide Ω est fini, la mesure définie par

$$\mu(A) = \frac{1}{\text{Card}(\Omega)} \sum_{\omega \in \Omega} \delta_\omega(A), \quad \forall A \in \mathcal{P}(\Omega),$$

est une probabilité discrète, appelée **mesure uniforme discrète** sur Ω .

4. D'autres exemples de probabilités discrètes sont données dans le tableau 2.2 à la fin du chapitre suivant.

1.2.2 La mesure de Lebesgue

Cette partie introduit la mesure de Lebesgue sur $\mathbb{R}^d$, qui étend la notion de longueur, d'aire et de volume aux ensembles boréliens de $\mathbb{R}^d$.

Soit $d \in \mathbb{N}^*$. Il existe une unique mesure λ_d sur $\mathbb{R}^d$ muni de la tribu borélienne telle que

$$\lambda_d([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d),$$

pour toutes familles $(a_i)_{1 \leq i \leq d}$ et $(b_i)_{1 \leq i \leq d}$ de $[-\infty, +\infty]$ vérifiant $a_i < b_i$ pour tout $1 \leq i \leq d$. Cette mesure est la **mesure de Lebesgue** sur $\mathbb{R}^d$.²

Étant donné que la mesure de Lebesgue étend les notions de longueur, aire ou volume, nous sommes autorisés à appliquer les formules connues pour calculer les aires et volumes classiques (triangle, disque, boule, etc...). Par exemple, la mesure de Lebesgue d'un disque $\mathcal{C}$ de $\mathbb{R}^2$ de rayon $r > 0$ est

$$\lambda_2(\mathcal{C}) = \text{aire}(\mathcal{C}) = \pi r^2.$$

1. Dans l'écriture de μ , il est toujours possible de choisir les éléments a_n deux à deux distincts (dans ce cas $\alpha_n = \mu(\{a_n\})$). Il est aussi toujours possible de supposer $\alpha_n > 0$.

2. L'existence d'une telle mesure est une conséquence du Théorème d'extension de Carathéodory, son unicité provient du lemme de classe monotone. Ces deux résultats sont hors programme.

Proposition 1.2.

La mesure de Lebesgue est invariante par les isométries euclidiennes de $\mathbb{R}^d$. En particulier, elle est invariante par translation, par symétrie et par rotation.

Remarque 1.6. La mesure de Lebesgue ne permet pas de mesurer tous les ensembles de $\mathbb{R}^d$. En effet, en supposant que tous les sous-ensembles de $\mathbb{R}^d$ sont mesurables, on arrive à des paradoxes célèbres : on peut construire un ensemble de longueur nulle contenant $[0, 1]$ (ensemble de Vitali) et on peut couper une boule de $\mathbb{R}^3$ en morceaux disjoints pour en reformer deux distinctes, chacune de même volume que la première (paradoxe de Banach-Tarski) ! Cependant tous les ensembles boréliens peuvent être mesurés par la mesure de Lebesgue. Les ensembles boréliens contiennent en particulier les réunions et intersections dénombrables d'ensembles ouverts et fermés de $\mathbb{R}^d$: cela suffira amplement à notre étude !

1.2.3 Les mesures à densité par rapport à la mesure de Lebesgue

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux et à valeurs positives. Alors on peut poser, pour tout $a < b$,

$$\nu([a, b]) = \int_a^b f(x) dx.$$

Cela permet de définir une mesure sur tous les intervalles $[a, b]$, qui peut en fait être étendue à tous les ensembles boréliens. La mesure obtenue est alors dite *absolument continue par rapport à la mesure de Lebesgue*, ou simplement *absolument continue* si il n'y a pas d'ambiguïté. La fonction f est appelée la *densité de la mesure ν* . Nous reviendrons sur ce type d'exemples dans le chapitre suivant, après avoir introduit l'intégrale de Lebesgue.

Exemple 1.5.

1. La mesure de Lebesgue est un cas particulier de mesure absolument continue, où l'on prend $f = 1$.
2. La mesure définie, pour tout $a < b$, par

$$\nu([a, b]) = \int_a^b e^{-x} dx$$

est une mesure absolument continue appelée *loi exponentielle de paramètre 1*. Il s'agit d'une mesure de probabilité.

Remarque 1.7.

1. Une mesure absolument continue est une mesure de probabilité si et seulement si l'intégrale de sa densité sur $\mathbb{R}$ vaut 1.
2. D'autres exemples de probabilités absolument continues sont données dans le tableau 2.3 à la fin du chapitre suivant.
3. On généralise facilement la notion de mesure absolument continue par rapport à la mesure de Lebesgue en dimension $d \geq 2$.

1.3 Probabilité conditionnelle et événements indépendants

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Nous appelons *événement de $\mathcal{F}$* tout ensemble $A \in \mathcal{F}$.

1.3.1 Probabilité conditionnelle

Définition 1.4 (Probabilité conditionnelle).

Soit $A \in \mathcal{F}$ tel que $\mathbb{P}(A) > 0$. On définit alors, pour tout $B \in \mathcal{F}$, la **probabilité conditionnelle de B par rapport à A** , par

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

L'application $\mathbb{P}(\cdot | A) : B \in \mathcal{F} \mapsto \mathbb{P}(B | A)$ est alors une mesure de probabilité sur Ω , appelée **la probabilité conditionnelle par rapport à A** .

Attention! Cette définition ne permet pas de construire les probabilités conditionnelles par rapport à des événements de probabilité nulle.

Terminons cette partie par la formule dite de Bayes, qui permet "*d'inverser*" les connaissances des probabilités conditionnelles.

Proposition 1.3 (Formule de Bayes).

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\mathcal{F}$ telle que $A_i \cap A_j = \emptyset$ pour $i \neq j$, et

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = 1.$$

Alors, pour tout $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$ et pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(A_k | B) = \frac{\mathbb{P}(B | A_k) \mathbb{P}(A_k)}{\sum_{n \in \mathbb{N}} \mathbb{P}(B | A_n) \mathbb{P}(A_n)}.$$

On trouve souvent cette proposition énoncée dans le cas où la famille $(A_n)_{n \in \mathbb{N}}$ ne contient qu'un événement non vide et son complémentaire : étant donnés deux événements A et B ,

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

1.3.2 Événements indépendants

Définition 1.5 (Indépendance d'événements).

Soient $A_1, A_2, \dots, A_n, \dots$ des événements de $\mathcal{F}$. Les événements $A_1, A_2, \dots, A_n, \dots$ sont dits **mutuellement indépendants** si, pour toute sous-famille finie $\{i_1, \dots, i_k\}$,

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}).$$

Attention! Si les événements $A_1, A_2, \dots, A_n, \dots$ sont indépendants deux à deux, c'est-à-dire si $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j)$ pour tout $i \neq j$, on ne peut pas en déduire que les événements $A_1, A_2, \dots, A_n, \dots$ sont mutuellement indépendants.

Remarque 1.8.

1. L'événement total Ω et l'événement vide est indépendant de tout autre événement.

2. Nous verrons en TD que, si $A_1, \dots, A_n, \dots$ sont indépendants, alors $A_1^{(c)}, \dots, A_n^{(c)}, \dots$ sont mutuellement indépendants, où $A_n^{(c)}$ désigne A ou son complémentaire A^c .
3. Si A et B sont indépendants, avec $\mathbb{P}(B) \neq 0$, alors on vérifie aisément que $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

1.3.3 Lemme de Borel-Cantelli

Étant donné une suite infinie d'événements $A_0, A_1, \dots, A_n, \dots$ d'une tribu $\mathcal{F}$, le lemme de Borel-Cantelli s'intéresse à la probabilité qu'une infinité d'entre eux se réalisent, c'est-à-dire à la probabilité de l'événement appelé *limite supérieure* de $(A_n)_{n \in \mathbb{N}}$ et défini par

$$\limsup_{n \in \mathbb{N}} A_n = \{\omega \in \Omega / \text{il existe une infinité d'indices } n \in \mathbb{N} \text{ tels que } \omega \in A_n\}.$$

Pour vérifier que $\limsup_{n \in \mathbb{N}} A_n$ est mesurable, nous pouvons le réécrire sous la forme d'une intersection de réunions d'événements :

$$\limsup_{n \in \mathbb{N}} A_n = \bigcap_{n=0}^{+\infty} \bigcup_{k=n}^{+\infty} A_k.$$

En effet, étant donné $\omega \in \Omega$, il existe une infinité d'indices $n \in \mathbb{N}$ tels que $\omega \in A_n$ si et seulement si pour tout $n \in \mathbb{N}$, il existe $k \geq n$ tel que $\omega \in A_k$; c'est-à-dire si et seulement si pour tout $n \in \mathbb{N}$, $\omega \in \bigcup_{k=n}^{+\infty} A_k$; c'est-à-dire si et seulement si $\omega \in \bigcap_{n=0}^{+\infty} \bigcup_{k=n}^{+\infty} A_k$.

Lemme 1.4 (Lemme de Borel-Cantelli).

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements de $\mathcal{F}$.

1. Si $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 0$.

2. Si les événements A_n , $n \in \mathbb{N}$, sont mutuellement indépendants alors

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty \quad \Leftrightarrow \quad \mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 0$$

et

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty \quad \Leftrightarrow \quad \mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 1.$$

D'après le lemme de Borel-Cantelli, si les événements $A_0, \dots, A_n, \dots$ sont indépendants, alors

$$\mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = \mathbb{P}(\omega \in \Omega / \text{il existe une infinité d'indices } n \in \mathbb{N} \text{ tels que } \omega \in A_n) \in \{0, 1\}.$$

Dans ce cas, on dit que l'événement $\limsup_{n \in \mathbb{N}} A_n$ suit une loi du 0 – 1.

Preuve du lemme de Borel-Cantelli.

1. Supposons que $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < +\infty$. Soit $\varepsilon > 0$. Alors il existe $N \in \mathbb{N}$ tel que $\sum_{n=N}^{+\infty} \mathbb{P}(A_n) < \varepsilon$.

D'après la propriété de sous-additivité (proposition 1.1),

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

Or $\limsup_{n \in \mathbb{N}} A_n = \bigcap_{n=0}^{+\infty} \bigcup_{k=n}^{+\infty} A_k \subset \bigcup_{n=N}^{+\infty} A_n$, donc

$$\mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) < \varepsilon$$

pour tout $\varepsilon > 0$. Par conséquent, $\mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 0$.

2. Supposons que les ensembles A_n , $n \in \mathbb{N}$, sont mutuellement indépendants et que

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty.$$

Soit $n \in \mathbb{N}$ et $m \geq n$. Par indépendance mutuelle des A_p , $p \in \mathbb{N}$,

$$\mathbb{P}\left(\bigcup_{k=n}^m A_k\right) = 1 - \mathbb{P}\left(\bigcap_{k=n}^m A_k^c\right) = 1 - \prod_{k=n}^m \mathbb{P}(A_k^c) = 1 - \prod_{k=n}^m (1 - \mathbb{P}(A_k)).$$

Étant donné que pour tout $x \in [0, 1]$, $0 \leq 1 - x \leq e^{-x}$,

$$\prod_{k=n}^m (1 - \mathbb{P}(A_k)) \leq \exp\left(-\sum_{k=n}^m \mathbb{P}(A_k)\right)$$

et donc

$$\mathbb{P}\left(\bigcup_{k=n}^m A_k\right) \geq 1 - \exp\left(-\sum_{k=n}^m \mathbb{P}(A_k)\right)$$

En faisant tendre $m \rightarrow +\infty$, on obtient par continuité monotone de $\mathbb{P}$:

$$\forall n \in \mathbb{N}, \mathbb{P}\left(\bigcup_{k=n}^{+\infty} A_k\right) \geq 1 \quad \text{et donc} \quad \forall n \in \mathbb{N}, \mathbb{P}\left(\bigcup_{k=n}^{+\infty} A_k\right) = 1$$

car $\mathbb{P}$ est à valeurs dans $[0, 1]$. Or la suite $(\bigcup_{k=n}^{+\infty} A_k)_{n \in \mathbb{N}}$ est une suite décroissante d'éléments de $\mathcal{A}$, donc, par continuité monotone (décroissante) de $\mathbb{P}$,

$$\mathbb{P}\left(\bigcap_{n=0}^{+\infty} \bigcup_{k=n}^{+\infty} A_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=n}^{+\infty} A_k\right) = 1.$$

Le point 1. du lemme de Borel-Cantelli permet de conclure. ■

Exemple 1.6. Considérons une suite dénombrable de dés, telle que le n -ième dé possède $2^n + 1$ faces, notées de 0 à 2^n , et qui sont équiprobables. Quelle est la probabilité qu'une infinité de dés tombent sur le nombre 1 ?

Pour répondre à cette question, nous définissons, pour chaque $n \in \mathbb{N}$, l'événement

$$A_n = \{\text{le } n\text{-ième dé tombe sur le nombre 1}\}.$$

Alors

$$\{\text{une infinité de dés tombent sur le nombre 1}\} = \limsup_{n \in \mathbb{N}} A_n.$$

Or $\mathbb{P}(A_n) = 1/(2^n + 1)$, donc

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty,$$

donc, d'après le lemme de Borel-Cantelli,

$$\mathbb{P}(\{\text{une infinité de dés tombent sur le nombre 1}\}) = \mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 0.$$

Remarquons qu'on ne suppose par que les dés sont tirés de manière indépendante.

Exemple 1.7. Considérons à nouveau une suite dénombrable de dés, telle que le n -ième dé possède $n + 1$ faces, notées de 0 à n , et qui sont équiprobables. Nous supposons de plus que les dés sont tirés de manière indépendantes, c'est-à-dire que, pour toute suite $u_1, \dots, u_n, \dots$, les événements $\{\text{le } n\text{-ième dé tombe sur } u_n\}$ sont mutuellement indépendants. Quelle est alors la probabilité qu'une infinité de dés tombent sur le nombre 1 ?

Avec les mêmes notations que dans l'exemple précédent, nous avons cette fois-ci $\mathbb{P}(A_n) = 1/(n + 1)$, donc

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty.$$

Les événements considérés étant mutuellement indépendants, on déduit de la deuxième partie du Lemme de Borel Cantelli que

$$\mathbb{P}(\{\text{une infinité de dés tombent sur le nombre 1}\}) = \mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 1.$$

Remarque 1.9. Pour définir proprement les événements considérés ci-dessus, on peut procéder comme suit. Considérons l'ensemble des suites de nombres entiers

$$\Omega = \mathbb{N} \times \mathbb{N} \times \dots = \mathbb{N}^{\mathbb{N}}.$$

On le muni de la plus petite tribu $\mathcal{F}$ contenant les ensembles $\Omega_{n,k}$, $n \geq 1, k \geq 0$, définis par

$$\Omega_{n,k} = \{\omega = (\omega_l)_{l \in \mathbb{N}} / \omega_n = k\}.$$

Les événements considérés dans les exemples ci-dessus sont alors donnés par $A_n = \Omega_{n,1}$.

De plus, on peut définir sur $(\Omega, \mathcal{F})$ une mesure de probabilité $\mathbb{P}$ telle que, pour tout $n \geq 1, k \geq 0$,

$$\mathbb{P}(\Omega_{n,k}) = \begin{cases} \frac{1}{2^{n+1}} & \text{si } k \leq 2^n, \\ 0 & \text{si } k > 2^n. \end{cases}$$

L'existence d'une telle probabilité est un résultat difficile et hors programme. Remarquons toutefois qu'il y a plusieurs façon de définir une telle probabilité et que l'on peut la choisir de telle sorte à rendre ou non les événements A_n mutuellement indépendants.

Chapitre 2

Fonctions mesurables et variables aléatoires

Intégrale de Lebesgue et espérance

Ce chapitre introduit les notions de fonctions mesurables et de variables aléatoires. Nous introduisons deux points clefs en probabilités : la notion de loi et le théorème du transport. Nous abordons également les principales étapes de la construction de l'intégrale de Lebesgue par rapport à une mesure¹, notion fondamentale en théorie moderne des probabilités.

2.1 Fonctions mesurables et variables aléatoires

2.1.1 Fonctions mesurables

Définition 2.1.

Soient $(\Omega_1, \mathcal{F}_1)$ et $(\Omega_2, \mathcal{F}_2)$ deux espaces mesurables. Une fonction $f : \Omega_1 \rightarrow \Omega_2$ est dite **mesurable** si

$$\forall B \in \mathcal{F}_2, f^{-1}(B) \in \mathcal{F}_1,$$

où on rappelle que $f^{-1}(B) = \{x \in \Omega_1 \mid f(x) \in B\}$.

Notons que la notion de fonction mesurable dépend des tribus choisies pour Ω_1 et Ω_2 . Pour bien signaler cette dépendance, nous dirons aussi que « f est mesurable par rapport à $\mathcal{F}_1$ et $\mathcal{F}_2$ » ou écrirons, pour être concis, « $f : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ ».

Remarque 2.1. Si $\Omega_1 \subset \mathbb{R}^{d_1}$ et $\Omega_2 \subset \mathbb{R}^{d_2}$ sont munis de leur tribu borélienne, une fonction mesurable sera encore appelée **fonction borélienne**.

Exemple 2.1.

1. Les fonctions constantes sont toutes mesurables (quelles que soient les tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ considérées).
2. Soient $(\Omega_1, \mathcal{F}_1)$ un espace mesurable et $A \subset \Omega_1$. On définit la fonction indicatrice $\mathbf{1}_A : \Omega_1 \rightarrow \{0, 1\}$ par

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

Avec $\{0, 1\}$ muni de la tribu discrète, la fonction indicatrice $\mathbf{1}_A$ est mesurable si et seulement si A est un ensemble $\mathcal{F}_1$ -mesurable. Il est conseillé au lecteur d'établir ce résultat, en guise d'exercice.

1. Nous ne donnons pas les preuves des propositions énoncées, car leur technicité s'écarte des objectifs du cours. Toutefois, le lecteur intéressé est invité à parcourir les nombreux ouvrages de références disponibles sur la question (par exemple *Mesure et intégration* de Daniel Revuz, 1997 aux éditions Hermann).

3. Soient $\Omega_1 \subset \mathbb{R}^{d_1}$ et $\Omega_2 \subset \mathbb{R}^{d_2}$. Nous munissons ces deux espaces de leur tribu borélienne. Alors les fonctions continues par morceaux ou encore les fonctions dérivées sont boréliennes, c'est-à-dire mesurables. Cela reste vrai si les tribus boréliennes considérées sont complétées (c'est-à-dire qu'on leur ajoute les ensembles négligeables pour la mesure de Lebesgue).

Remarque 2.2. Pour des raisons pratiques, nous souhaiterons souvent considérer des fonctions pouvant prendre les valeurs $-\infty$ et $+\infty$. Pour y parvenir, nous introduisons l'ensemble $\overline{\mathbb{R}} := [-\infty, +\infty]$ muni de sa tribu borélienne (la plus petite tribu qui contient les ouverts de $\overline{\mathbb{R}}$). Nous parlerons alors de fonctions mesurables à valeurs dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.

Terminons cette partie en donnant quelques propriétés de stabilité de l'ensemble des fonctions mesurables.

Proposition 2.1.

1. Toute composée de fonctions mesurables est une fonction mesurable.
2. Soit $(f_n)_{n \in \mathbb{N}}$ une suite d'applications mesurables de $(\Omega, \mathcal{F})$ dans $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors
 - (a) dès qu'ils sont bien définis, les combinaisons linéaires et les produits d'applications de la famille $(f_n)_{n \in \mathbb{N}}$ forment des fonctions mesurables,
 - (b) les fonctions $\sup_{n \in \mathbb{N}} f_n$ et $\inf_{n \in \mathbb{N}} f_n$ sont aussi mesurables,
 - (c) de plus si $(f_n)_{n \in \mathbb{N}}$ converge simplement, alors sa limite $f := \lim_{n \rightarrow +\infty} f_n$ est aussi mesurable.

Les fonctions que nous rencontrerons dans ce cours seront en général mesurables et nous aurons rarement à nous soucier de problèmes de non-mesurabilité. Cependant, la notion de mesurabilité est pertinente et même cruciale dès que l'on considère l'étude de processus stochastiques (quantités qui évoluent aléatoirement en fonction du temps) ou de processus partiellement observés. Dans ces situations, la tribu contient l'information disponible par un observateur ; une fonction mesurable est alors une fonction qui, au mieux, maintient l'information disponible.

2.1.2 Variables aléatoires

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $(E, \mathcal{E})$ un espace mesurable. Rappelons que, dans ce cas, nous interprétons chaque élément ω de Ω comme une réalisation possible. Une variable aléatoire est une fonction qui, à chaque réalisation possible, associe une valeur.

La définition d'une variable aléatoire X prenant ses valeurs dans un espace mesurable $(E, \mathcal{E})$ quelconque est la suivante. Il s'agit en fait d'un cas particulier de fonction mesurable.

Définition 2.2.

Une application $X : \Omega \rightarrow E$ est appelée **variable aléatoire** si elle est mesurable, c'est-à-dire si

$$\forall B \in \mathcal{E}, X^{-1}(B) \in \mathcal{F}$$

où on rappelle que $X^{-1}(B)$ est donné par l'événement $\{\omega \in \Omega / X(\omega) \in B\}$.

En probabilités, l'ensemble $X^{-1}(B)$ est un événement, noté $\{X \in B\}$, interprété comme l'ensemble des réalisations telles que $X \in B$.

Remarque 2.3. La notion de variable aléatoire coïncidant avec celle de fonction mesurable, cette notion dépend des tribus $\mathcal{F}$ et $\mathcal{E}$ dont sont munis les espaces Ω et E . De plus, les mêmes propriétés de stabilité s'appliquent pour les variables aléatoires.

La loi d'une variable aléatoire X est une probabilité définie sur l'espace dans lequel X prend ses valeurs.

Définition-Proposition 2.2.

Soit $X : \Omega \rightarrow E$ une variable aléatoire. Alors l'application

$$\begin{aligned} \mathbb{P}_X : \mathcal{E} &\longrightarrow [0, +\infty] \\ B &\longmapsto \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \end{aligned}$$

est une probabilité sur $(E, \mathcal{E})$ appelée **loi de la variable aléatoire X** (sous la probabilité $\mathbb{P}$).

Remarque 2.4.

1. La loi d'une variable aléatoire dépend des tribus $\mathcal{F}$ et $\mathcal{E}$ considérées ainsi que de la probabilité $\mathbb{P}$.
2. En modélisation, la loi de X est précisée mais le plus souvent l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ ne l'est pas. Ainsi, étant donné une mesure de probabilité μ sur $\mathbb{N}$, $\mathbb{R}$ ou $\mathbb{R}^d$, on dira souvent que l'on se donne une variable aléatoire X de loi μ .

Quelques exemples de lois sur $\mathbb{N}$ et $\mathbb{R}$, très utilisées en pratique, sont présentées en fin de chapitre.

Proposition 2.3.

Supposons que $(\Omega, \mathcal{F}, \mathbb{P})$ est complet. Si $X, Y : \Omega \rightarrow E$ sont égales $\mathbb{P}$ -presque sûrement, alors X et Y ont même loi.

Attention ! La réciproque est fausse.

Démonstration. Notons $\mathbb{P}_X$ (resp. $\mathbb{P}_Y$) la loi de X (resp. de Y). Fixons $B \in \mathcal{E}$.

Par définition,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Alors, les événements $\{X = Y\}$ et $\{X \neq Y\}$ formant une partition de Ω ,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B, X = Y) + \mathbb{P}(X \in B, X \neq Y).$$

Par hypothèse, $X = Y$ $\mathbb{P}$ -presque sûrement, ce qui signifie que $\{X \neq Y\}$ est un ensemble négligeable pour $\mathbb{P}$. A fortiori, son sous-ensemble $\{X \in B, X \neq Y\}$ est négligeable et donc $\mathbb{P}(X \in B, X \neq Y) = 0$. Dès lors, il vient :

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B, X = Y).$$

Par symétrie,

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B, X = Y) = \mathbb{P}(X \in B, X = Y) = \mathbb{P}_X(B).$$

Nous venons de montrer que sur $\mathcal{E}$, $\mathbb{P}_Y = \mathbb{P}_X$, c'est-à-dire que X et Y ont même loi. ■

2.2 Intégrale de Lebesgue

Dans toute cette section, $(\Omega, \mathcal{F}, \mu)$ est un espace mesuré. Cette partie introduit l'intégrale de Lebesgue d'une fonction mesurable par rapport à la mesure μ . Étant donnée une fonction mesurable $f : \Omega \rightarrow \overline{\mathbb{R}}$, où $\overline{\mathbb{R}}$ est muni de la tribu borélienne, on notera cette intégrale

$$\int_{\Omega} f d\mu.$$

Attention ! On indique $d\mu$ dans l'intégrale pour rappeler la dépendance en μ . On trouvera de manière courante les notations suivantes, qui sont parfaitement équivalentes :

$$\int_{\Omega} f d\mu = \int_{\Omega} f(x) d\mu(x) = \int_{\Omega} f(x) \mu(dx).$$

Dans le cadre probabiliste, c'est-à-dire quand μ est une probabilité, l'intégrale de f par rapport à μ est appelée l'espérance de f par rapport à μ , notée $\mathbb{E}^{\mu}(f)$ ou simplement $\mathbb{E}(f)$ s'il n'y a pas d'ambiguïté sur μ . Nous reviendrons sur cette notion importante dans la section et les chapitres suivants.

Nous donnons dans un premier temps les principales étapes de la construction de l'intégrale de Lebesgue puis les propriétés générales de cette intégrale. Cette construction abstraite présente l'intérêt d'être très générale. Dans un deuxième temps, nous nous concentrons sur le calcul de cette intégrale dans des cas particuliers.

2.2.1 Construction de l'intégrale de Lebesgue

La construction de l'intégrale de Lebesgue se fait en 4 étapes : dans un premier temps, nous considérons les fonctions indicatrices, puis les fonctions positives dites étagées, puis les fonctions mesurables positives et enfin les fonctions dites *intégrables*. Le lecteur constatera que chaque étape est une extension naturelle de la précédente.

Étape 1. Soit f une fonction indicatrice mesurable, c'est-à-dire soit $f = \mathbf{1}_A$ avec $A \in \mathcal{F}$. Nous rappelons que

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

On appelle intégrale de f par rapport à μ la quantité

$$\int_{\Omega} f d\mu := \mu(A).$$

Étape 2. Soit f une fonction étagée positive, c'est-à-dire une fonction s'écrivant sous la forme

$$f(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x).$$

avec $n \in \mathbb{N}^*$, $A_1, A_2, \dots, A_n \in \mathcal{F}$ et $\alpha_1, \alpha_2, \dots, \alpha_n \in [0, +\infty]$. On appelle intégrale de f par rapport à μ la quantité

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

avec la convention $0 \times +\infty = +\infty \times 0 = 0$. Cette quantité est bien définie et appartient à $[0, +\infty]$.

Étape 3. Soit $f : \Omega \rightarrow [0, +\infty]$ une fonction mesurable positive. On appelle intégrale de f par rapport à μ la quantité

$$\int_{\Omega} f d\mu := \sup \left\{ \int_{\Omega} h d\mu \mid 0 \leq h \leq f \text{ et } h \text{ est étagée} \right\}.$$

Cette valeur est toujours bien définie, éventuellement égale à $+\infty$.

Étape 4. Soit $f : \Omega \rightarrow \overline{\mathbb{R}}$ une fonction mesurable à valeurs dans $\overline{\mathbb{R}}$. On note f_+ et f_- les fonctions appelées partie positive et partie négative de f , qui sont définies par

$$f_+(x) = \max(f(x), 0) \text{ et } f_-(x) = \max(-f(x), 0), \quad \forall x \in \Omega,$$

Par définition de f_+ et f_- ,

$$\begin{cases} f &= f_+ - f_- \\ |f| &= f_+ + f_- \end{cases}$$

Remarquons que $|f|, f_+$ et f_- sont des fonctions mesurables positives dont on peut construire l'intégrale grâce à l'étape 3. On dira que f est **intégrable** si

$$\int_{\Omega} |f| d\mu < \infty.$$

Dans ce cas, $\int_{\Omega} f_+ d\mu < \infty$ et $\int_{\Omega} f_- d\mu < \infty$ et on appelle intégrale de f par rapport à μ le réel

$$\int_{\Omega} f d\mu = \int_{\Omega} f_+ d\mu - \int_{\Omega} f_- d\mu.$$

Ceci clôt la construction de l'intégrale de Lebesgue par rapport à la mesure μ .

Remarque 2.5. Si f n'est pas intégrable, c'est-à-dire si $\int_{\Omega} |f| d\mu = +\infty$, alors on ne peut pas définir l'intégrale de f , sauf si f est à valeurs positives (auquel cas l'intégrale de $f = |f|$ est $+\infty$).

Remarque 2.6. Soit f une fonction à valeurs complexes, c'est-à-dire telle que $f = f_1 + if_2$ avec f_1 et f_2 mesurables à valeurs réelles. On dit que f est intégrable si $\int_{\Omega} |f| d\mu < \infty$, ce qui est équivalent à f_1 et f_2 sont intégrables. Dans ce cas, on pose

$$\int_{\Omega} f d\mu = \int_{\Omega} f_1 d\mu + i \int_{\Omega} f_2 d\mu.$$

Les propriétés énoncées dans la section suivante, lorsqu'elles s'appliquent, restent vraies pour les fonctions à valeurs complexes.

2.2.2 Propriétés générales

Dans cette section, $\overline{\mathbb{R}}$ et ses intervalles sont munis de la tribu borélienne.

Proposition 2.4.

1. Soient $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ deux fonctions mesurables égales μ -presque partout. Alors f est intégrable si et seulement si g est intégrable. Dans ce cas, ou si les deux fonctions sont positives, leurs intégrales coïncident.

2. Linéarité de l'intégrale :

Si $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ sont deux fonctions intégrables par rapport à μ , alors, pour tous $\alpha, \beta \in \mathbb{R}$, la fonction $\alpha f + \beta g$ est intégrable et

$$\int_{\Omega} (\alpha f + \beta g) d\mu = \alpha \int_{\Omega} f d\mu + \beta \int_{\Omega} g d\mu.$$

L'égalité ci-dessus reste vraie pour $f, g : \Omega \rightarrow [0, +\infty]$ mesurables, avec $\alpha, \beta \in [0, +\infty]$.

3. Positivité/Croissance de l'intégrale

(a) Si $f, g : \Omega \rightarrow [0, +\infty]$ sont deux fonctions mesurables positives telles que $f \leq g$ μ -presque partout, alors

$$\int_{\Omega} f d\mu \leq \int_{\Omega} g d\mu.$$

(b) Si $f, g : \Omega \rightarrow \overline{\mathbb{R}}$ sont deux fonctions intégrables par rapport à μ telles que $f \leq g$ μ -presque partout, alors

$$\int_{\Omega} f d\mu \leq \int_{\Omega} g d\mu,$$

avec égalité si et seulement si $f = g$ μ -presque-sûrement.

(c) Soit $f : \Omega \rightarrow \overline{\mathbb{R}}$ une fonction intégrable par rapport à μ telle que $f \geq 0$ μ -presque partout. Alors

$$\int_{\Omega} f d\mu = 0 \iff f = 0 \text{ } \mu\text{-presque-sûrement.}$$

4. Si $f : \Omega \rightarrow \overline{\mathbb{R}}$ est une fonction intégrable par rapport à μ , alors f est finie μ -presque partout et

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu.$$

Remarque 2.7. Soit $A \in \mathcal{F}$ un ensemble mesurable. Pour toute fonction $f : \Omega \rightarrow \mathbb{R}$ telle que $\int_{\Omega} \mathbf{1}_A(x) f(x) d\mu(x)$ est bien définie (c'est-à-dire telle que $f\mathbf{1}_A$ est mesurable positive ou μ -intégrable), on pose

$$\int_A f(x) d\mu(x) := \int_{\Omega} \mathbf{1}_A(x) f(x) d\mu(x).$$

De cette façon, on retrouve immédiatement l'égalité de Chasles : pour tous $A, B \in \mathcal{F}$ disjoints tels que $\int_{A \cup B} f d\mu$ est bien définie, on a

$$\int_{A \cup B} f(x) d\mu(x) = \int_A f(x) d\mu(x) + \int_B f(x) d\mu(x).$$

2.2.3 Calcul dans des cas particuliers

L'intégrale de Lebesgue permet de construire des intégrales sur tous les espaces mesurés, ce qui en fait un des fondements de la théorie des probabilités modernes comme nous le verrons dans les chapitres suivants. Cependant, il est parfois délicat, voire impossible, de calculer une intégrale de Lebesgue sur un espace quelconque. Nous nous concentrons dans cette section sur des cas particuliers pour lesquels nous serons capables de mener des calculs explicites.

Mesures discrètes

Soient Ω un ensemble non-vidé et μ une mesure discrète sur Ω . Il existe alors une famille $(a_n)_{n \in \mathbb{N}}$ d'éléments de Ω et une famille $(\alpha_n)_{n \in \mathbb{N}}$ d'éléments de $[0, +\infty]$ telles que

$$\mu(A) = \sum_{n \in \mathbb{N}} \alpha_n \delta_{a_n}(A), \quad \forall A \in \mathcal{P}(\Omega).$$

Soit $f : \Omega \rightarrow \overline{\mathbb{R}}$ une fonction mesurable. On a alors

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

Nous en déduisons que f est intégrable si et seulement si

$$\sum_{n \in \mathbb{N}} \alpha_n |f(a_n)| < +\infty,$$

c'est-à-dire si et seulement si la série de terme général $\alpha_n f(a_n)$ est absolument sommable. De plus si f est intégrable, alors

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

Insistons sur le fait que la formule ci-dessus est vraie pour f mesurable positive ou f intégrable.

Remarque 2.8. Soit $u : \mathbb{N} \rightarrow \mathbb{R}$ une fonction à valeurs réelles et μ la mesure de comptage sur $\mathbb{N}$, donnée par

$$\mu(A) = \sum_{n \in \mathbb{N}} \delta_n(A) = \text{Card}(A), \quad \forall A \subset \mathbb{N}.$$

D'après ce qui précède, l'intégrale de $|u|$ par rapport à μ est donnée par

$$\int_{\mathbb{N}} |u(n)| d\mu(n) = \sum_{n \in \mathbb{N}} |u(n)|.$$

Ainsi, la fonction u est intégrable par rapport à la mesure de comptage si et seulement si la série de terme général $|u(n)|$ est absolument sommable. Si tel est le cas, on a de plus

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

En particulier, il sera possible de considérer les séries comme des intégrales et de leur appliquer les mêmes propriétés.

Mesure de Lebesgue

Soit $d \geq 1$. On suppose ici que $\Omega \subset \mathbb{R}^d$ est muni de la tribu borélienne et de la mesure de Lebesgue λ_d . L'intégrale par rapport à la mesure de Lebesgue généralise l'intégrale de Riemann (cette dernière étant définie pour les fonctions continues par morceaux). Ainsi, lorsque l'on cherche à calculer l'intégrale $\int_{\Omega} f d\lambda_d$ où f est continue par morceaux, on peut utiliser les techniques connues pour l'intégrale de Riemann (relation de Chasles, calcul de primitive, intégration par partie, changement de variables, etc...). Par exemple,

$$\int_{]0,1/2[} \frac{1}{x(\ln x)^2} d\lambda_1(x) = \left[-\frac{1}{\ln x} \right]_0^{1/2} = \frac{1}{\ln 2}.$$

Lorsque la fonction f n'est pas continue par morceaux, on peut essayer de se ramener à une fonction continue par morceaux g égale à f λ_d -presque partout. Par exemple, comme $\mathbf{1}_{x \notin \mathbb{Q}} = 1$ λ_1 -presque partout,

$$\mathbf{1}_{x \notin \mathbb{Q}} \frac{1}{x(\ln x)^2} = \frac{1}{x(\ln x)^2} \quad \lambda_1\text{-presque partout}$$

et donc

$$\int_{]0,1/2[} \mathbf{1}_{x \notin \mathbb{Q}} \frac{1}{x(\ln x)^2} d\lambda_1(x) = \int_{]0,1/2[} \frac{1}{x(\ln x)^2} d\lambda_1(x) = \frac{1}{\ln 2}.$$

On peut également essayer de se ramener à une fonction f limite croissante (resp. uniformément bornée) de fonctions dont on sait calculer l'intégrale et utiliser le théorème de convergence monotone (resp. dominée). Nous précisons cela au fur et à mesure du cours et notamment lors du chapitre 6 sur les grands théorèmes de l'intégration.

Attention ! La notion d'intégrale impropre n'existe pas pour l'intégrale de Lebesgue.

Mesures à densité

Étant donné un espace mesuré $(\Omega, \mathcal{F}, \mu)$ et une fonction positive mesurable $h : \Omega \rightarrow [0, +\infty]$, on peut définir une fonction $\nu : \mathcal{F} \rightarrow [0, +\infty]$ par

$$\nu(A) = \int_{\Omega} \mathbf{1}_A(x) h(x) d\mu(x), \quad \forall A \in \mathcal{F}.$$

La fonction ν ainsi définie est une mesure positive, appelée **mesure de densité h par rapport à μ** . On note souvent $d\nu = h d\mu$.

Remarque 2.9. On a $\nu(\Omega) = \int_{\Omega} \mathbf{1}_{\Omega}(x) h(x) d\mu(x) = \int_{\Omega} h(x) d\mu(x)$. On en déduit donc que ν est une probabilité si et seulement si $\int_{\Omega} h d\mu = 1$.

Pour toute fonction mesurable $f : \Omega \rightarrow \overline{\mathbb{R}}$, on a

$$\int_{\Omega} |f(x)| d\nu(x) = \int_{\Omega} |f(x)| h(x) d\mu(x).$$

Par suite, une fonction mesurable $f : \Omega \rightarrow \overline{\mathbb{R}}$ est ν -intégrable si et seulement si fh est μ -intégrable. De plus, si f est ν -intégrable, alors

$$\boxed{\int_{\Omega} f(x) d\nu(x) = \int_{\Omega} f(x) h(x) d\mu(x).}$$

Insistons sur le fait que l'égalité précédente est vraie pour f mesurable positive ou f telle que fh est μ -intégrable. Cette égalité explique la notation $d\nu = h d\mu$: pour calculer $\int_{\Omega} f d\nu$, on remplace formellement le terme $d\nu$ par $h d\mu$.

Exemple 2.2 (Cas où ν est la mesure gaussienne $\mathcal{N}(0, 1)$). Vous avez déjà rencontré cet exemple dans le cours de statistique et appris que cette mesure vérifie, pour tout $-\infty \leq a < b \leq +\infty$,

$$\nu([a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

où l'intégrale a lieu au sens de Riemann. Comme la fonction intégrée est continue par morceaux, on a

$$\nu([a, b]) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbf{1}_{x \in [a, b]} e^{-\frac{x^2}{2}} d\lambda_1(x),$$

où l'intégrale a lieu au sens de Lebesgue. En fait, cette égalité s'étend à tous les ensembles boréliens et on en déduit que la mesure gaussienne est la mesure de densité $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ par rapport à la mesure de Lebesgue λ_1 .

Remarque 2.10. Nous utiliserons couramment des mesures à densité par rapport à la mesure de Lebesgue dans la suite du cours. Suivant une pratique courante en probabilités, nous omettrons abusivement de préciser que la mesure de référence est la mesure de Lebesgue. Ainsi, nous pourrions écrire « la mesure gaussienne est une mesure de densité $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ » en omettant « par rapport à la mesure de Lebesgue λ_1 ».

Mesures mixtes

On se place dans la situation $\Omega = \mathbb{R}^d$ est muni de la tribu borélienne. Une mesure mixte sur $\mathbb{R}^d$ est une mesure qui possède une partie discrète et une partie à densité par rapport à la mesure de Lebesgue. Plus précisément, une mesure μ sur $\mathbb{R}^d$ est dite **mixte** s'il existe une famille $(a_n)_{n \in \mathbb{N}}$ d'éléments de $\mathbb{R}^d$, une famille $(\alpha_n)_{n \in \mathbb{N}}$ d'éléments de $[0, +\infty]$ et une fonction mesurable positive $h : \mathbb{R}^d \rightarrow [0, +\infty]$, telle que

$$\mu(A) = \sum_{n \in \mathbb{N}} \alpha_n \delta_{a_n}(A) + \int_{\mathbb{R}^d} \mathbf{1}_A(x) h(x) d\lambda_d(x).$$

Dans ce cas, on note souvent $d\mu = \sum_{n \in \mathbb{N}} \alpha_n d\delta_{a_n} + h d\lambda_d$.

Pour toute fonction mesurable $f : \Omega \rightarrow \overline{\mathbb{R}}$, on a

$$\int_{\Omega} |f(x)| d\mu(x) = \sum_{n \in \mathbb{N}} \alpha_n |f(a_n)| + \int_{\Omega} |f(x)| h(x) d\lambda_d(x)$$

Alors une fonction mesurable f est μ -intégrable si et seulement si elle vérifie les deux conditions suivantes :

- la série de terme général $(\alpha_n f(a_n))$ est absolument sommable ;
- fh est Lebesgue-intégrable.

De plus, si ces deux conditions sont vérifiées,

$$\int_{\Omega} f(x) d\mu(x) = \sum_{n \in \mathbb{N}} \alpha_n f(a_n) + \int_{\Omega} f(x) h(x) d\lambda_d(x).$$

2.3 Espérance d'une variable aléatoire et théorème du transport

L'espérance d'une variable aléatoire X est simplement l'intégrale de Lebesgue contre la probabilité $\mathbb{P}$ dont est muni l'espace des réalisations possibles $(\Omega, \mathcal{F})$.

Définition 2.3 (Espérance d'une variable aléatoire positive ou intégrable).

Soit $X : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une variable aléatoire $\mathbb{P}$ -intégrable ou à valeurs dans $[0, +\infty]$ $\mathbb{P}$ -presque sûrement. **L'espérance** $\mathbb{E}(X)$ de X est l'intégrale de X par rapport à la probabilité $\mathbb{P}$, c'est-à-dire que

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

Cette définition théorique est utile car elle donne un fondement rigoureux et général à la notion d'espérance. En revanche, on ne détermine que très rarement l'espace Ω et la fonction $X : \Omega \rightarrow \mathbb{R}$ explicitement, ce qui rend caduc l'utilisation de cette définition pour le calcul pratique de l'espérance. A contrario, on dispose en général de la loi de X , ce qui suffit pour calculer son espérance (et même plus), comme l'explique le théorème suivant.

Théorème 2.5 (Théorème du transport).

Soient $(E, \mathcal{E})$ un espace mesurable et $X : \Omega \rightarrow E$ une variable aléatoire de loi $\mathbb{P}_X$. Considérons une fonction mesurable $\varphi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si $\varphi(X)$ est $\mathbb{P}$ -presque sûrement à valeurs dans $[0, +\infty]$, alors

$$\mathbb{E}(\varphi(X)) = \int_E \varphi(x) \mathbb{P}_X(dx).$$

Si $\mathbb{E}(|\varphi(X)|) < \infty$, alors $\varphi(X)$ est intégrable et $\mathbb{E}(\varphi(X))$ est égale à l'intégrale ci-dessus.

Le tableau 2.1 donné en fin de section ré-écrit ce théorème dans le cas des lois discrètes, absolument continues et mixtes. Il est en général le point de départ de tout calcul d'une espérance dans le cas des lois considérées.

La notion d'espérance n'étant qu'une instance particulière de l'intégrale de Lebesgue, nous pouvons reformuler les principales propriétés de celles-ci vues dans la section précédente.

Proposition 2.6.

Soient X et Y deux variables aléatoires définies sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\overline{\mathbb{R}}$.

1. L'espérance est linéaire sur le $\mathbb{R}$ -espace vectoriel des variables aléatoires $\mathbb{P}$ -intégrables à valeurs dans $\overline{\mathbb{R}}$, c'est-à-dire que si X et Y sont intégrables alors pour tous $a, b \in \mathbb{R}$,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Si $X \geq 0$ et $Y \geq 0$ $\mathbb{P}$ -presque sûrement, alors l'égalité précédente est aussi vérifiée dès lors que $a, b \in [0, +\infty]$.

2. Si $Y \geq 0$ $\mathbb{P}$ -presque sûrement ou si X et Y sont toutes les deux $\mathbb{P}$ -intégrables, alors,

$$Y \leq X \text{ } \mathbb{P}\text{-presque sûrement} \implies \mathbb{E}(Y) \leq \mathbb{E}(X).$$

En particulier, si $X \geq 0$ $\mathbb{P}$ -presque sûrement, alors $\mathbb{E}(X) \in [0, +\infty]$.

3. Si $X \geq 0$ $\mathbb{P}$ -presque sûrement, alors

$$\mathbb{E}(X) = 0 \iff X = 0 \text{ } \mathbb{P}\text{-presque sûrement}.$$

4. Si la variable aléatoire X est $\mathbb{P}$ -intégrable, alors X est finie $\mathbb{P}$ -presque sûrement et

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

En conséquence immédiate du théorème du transport, nous pouvons donner la caractérisation suivante de la loi d'une variable aléatoire. Sa démonstration est laissée en exercice. On pourra commencer par montrer l'égalité importante suivante : pour tout ensemble mesurable $A \subset E$, $\mathbb{E}(\mathbf{1}_A(X)) = \mathbb{P}(X \in A)$.

Proposition 2.7 (Caractérisation de la loi d'une variable aléatoire).

Soient $X, Y : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ deux variables aléatoires. Alors, X et Y ont même loi si et seulement si, pour toute fonction $\varphi : (E, \mathcal{E}) \rightarrow ([0, +\infty[, \mathcal{B}([0, +\infty[))$ mesurable bornée positive,

$$\mathbb{E}(\varphi(X)) = \mathbb{E}(\varphi(Y)). \tag{2.1}$$

Remarque 2.11. Nous ne le démontrerons pas, mais, si $E \subset \mathbb{R}^d$ et $\mathcal{E}$ est la tribu borélienne sur E , alors les variables aléatoires $X, Y : \Omega \rightarrow E$ ont même loi si et seulement si (2.1) est vérifiée pour toute fonction $\varphi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ continue bornée positive.

Loi de la v.a. X	Écriture de $\mathbb{E}(\varphi(X))$ pour $\varphi(X)$ positive p.s. ou intégrable
Loi discrète $\mathbb{P}_X(A) = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}(A)$	$\rightsquigarrow \mathbb{E}(\varphi(X)) = \sum_{n \in \mathbb{N}} p_n \varphi(x_n).$ $\rightsquigarrow \varphi(X) \text{ intégrable} \iff \sum_{n \in \mathbb{N}} p_n \varphi(x_n) < +\infty.$
Loi absolument continue $\forall A \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(A) = \int_A f(x) \lambda_d(dx).$	$\rightsquigarrow \mathbb{E}(\varphi(X)) = \int_{\mathbb{R}^d} \varphi(x) f(x) \lambda_d(dx).$ $\rightsquigarrow \varphi(X) \text{ intégrable} \iff \varphi \times f \text{ Lebesgue intégrable.}$
Loi ayant une partie discrète et une partie absolument continue : Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(A) = \sum_{n \in \mathbb{N}} \alpha_n \delta_{x_n}(A) + \int_A h(x) \lambda_d(dx)$ avec $0 \leq \alpha_n \leq 1$ et h borélienne positive tels que $\sum_{n \in \mathbb{N}} \alpha_n + \int_{\mathbb{R}^d} h(x) \lambda_d(dx) = 1.$	$\rightsquigarrow \mathbb{E}(\varphi(X)) = \sum_{n \in \mathbb{N}} \alpha_n \varphi(x_n) + \int_{\mathbb{R}^d} \varphi(x) h(x) \lambda_d(dx).$ $\rightsquigarrow \varphi(X) \text{ intégrable} \iff \begin{cases} \sum_{n \in \mathbb{N}} \alpha_n \varphi(x_n) < +\infty \\ \text{et } \varphi \times h \text{ Lebesgue intégrable} \end{cases}$

TABLE 2.1 – Théorème du transport dans des cas particuliers

Loi de la v.a. X	Expression de $\mathbb{P}_X$	Modélisation
Loi uniforme sur $E = \{x_1, \dots, x_N\}$ avec $N \in \mathbb{N}^*$ et $x_i \neq x_j$ pour $i \neq j$	$\mathbb{P}_X = \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$	Le résultat X obtenu lors d'un lancé d'un dé non truqué suit une loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$.
Loi de Bernoulli $\mathcal{B}(p)$ avec $p \in [0, 1]$	$\mathbb{P}_X = (1-p)\delta_0 + p\delta_1$	Réalisons une expérience dont la probabilité de succès est p . Alors la variable aléatoire X définie par $X = \begin{cases} 1 & \text{si un succès est obtenu} \\ 0 & \text{sinon} \end{cases}$ suit la loi de Bernoulli de paramètre p .
Loi Binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$	$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k$	Réalisons n fois, de manière indépendante, une expérience dont la probabilité de succès est p . Alors le nombre de succès X obtenus au cours des n expériences suit une loi $\mathcal{B}(n, p)$.
Loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	$\mathbb{P}_X = \sum_{k=0}^{+\infty} \frac{e^{-\lambda} \lambda^k}{k!} \delta_k$	Cette loi est utilisée pour modéliser le nombre de réalisations d'un évènement A dont la probabilité d'occurrence est faible mais dont le nombre de réalisations peut être grand (par exemple, nombre d'appels reçus en une minute par un standard).
Loi Géométrique $\mathcal{G}(p)$ avec $p \in]0, 1[$	$\mathbb{P}_X = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k$	Notons p la probabilité de succès d'une expérience. Réalisons cette expérience plusieurs fois de manière indépendante. Notons Y le premier instant $n \in \mathbb{N}^*$ correspondant à un succès. Alors, Y suit la loi $\mathcal{G}(p)$.

TABLE 2.2 – Exemples de modélisation par des lois discrètes

Loi de la v.a. X	Densité de $\mathbb{P}_X$ (par rapport à λ_1)	Modélisation / Utilisation
Loi Uniforme $\mathcal{U}(\mathcal{B})$ avec $\mathcal{B} \subset \mathbb{R}^d$ borné tel que $0 < \lambda_d(\mathcal{B}) < +\infty$	$f_X(x) = \frac{1}{\lambda_d(\mathcal{B})} \mathbf{1}_{\mathcal{B}}(x)$	Un marcheur se trouve en la position $(0, 0)$ et se déplace aléatoirement dans le plan sans préférence. La direction (normée) qu'il prend suit la loi uniforme sur le cercle $\mathcal{C}(0, 1)$: l'angle de la direction avec le vecteur $(0, 1)$ choisi suit une loi uniforme sur $[0, 2\pi[$.
Loi Gaussienne $\mathcal{N}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	Erreur commise par un appareil de mesure.
Loi Exponentielle $\mathcal{E}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	$f_X(x) = \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{\mathbb{R}_+^*}(x)$	Durée de vie d'un phénomène sans mémoire ou sans vieillissement.
Loi du χ^2 $\chi^2(k)$ à $k \in \mathbb{N}^*$ degrés de liberté	$f_X(x) = \frac{x^{k/2-1} e^{-x/2}}{2^{k/2} \Gamma(k/2)}$	Utilisée pour les tests statistiques (cf. cours du 1 ^{er} semestre).

TABLE 2.3 – Exemples d'utilisation de lois absolument continues

Chapitre 3

Théorèmes d'intégration

La première partie de ce chapitre donne des conditions suffisantes pour intervertir les signes limite et intégrale/espérance¹, c'est-à-dire des conditions sous lesquelles

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} \lim_{n \rightarrow +\infty} f_n d\mu \quad \text{ou} \quad \lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n \rightarrow +\infty} X_n\right).$$

Nous énoncerons également le très utile Lemme de Fatou.

Dans la seconde partie, nous donnerons des conditions suffisantes pour montrer la continuité et la dérivabilité d'intégrale/espérance paramétrées.

La troisième et dernière partie du chapitre est consacrée au théorème de changement de variables et ses applications pour l'identification d'une loi via le théorème du transport.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mu)$ est un **espace mesuré**; comme évoqué dans le chapitre 1, il est supposé complet, c'est-à-dire que la tribu $\mathcal{F}$ contient tous les ensembles négligeables pour la mesure μ .
- Tout espace $E \subset \mathbb{R}^d$ est muni de sa tribu borélienne $\mathcal{B}(E)$ et tout espace E fini ou dénombrable est muni de sa tribu discrète $\mathcal{P}(E)$.

3.1 Théorèmes de convergence

3.1.1 Théorème de convergence monotone

Nous commençons par étudier les suites de fonctions mesurables positives.

Théorème 3.1 (Théorème de convergence monotone).

Pour tout $n \in \mathbb{N}$, soit $f_n : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable. Supposons que pour tout $n \in \mathbb{N}$,

$$f_n \in [0, +\infty] \text{ } \mu\text{-presque partout et } f_n \leq f_{n+1} \text{ } \mu\text{-presque partout.}$$

Nous notons f la limite (définie μ -presque partout) de la suite $(f_n)_{n \in \mathbb{N}}$. Alors l'intégrale de f est bien définie et

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

1. Les théorèmes que nous allons présenter ont pu être énoncés en classe préparatoire dans le cadre de l'intégrale de Riemann. La pratique que vous avez alors acquise vous sera bien sûr utile ici.

Remarque 3.1. Dans le théorème de convergence monotone, la fonction f est a priori seulement définie μ -presque partout, c'est-à-dire sur $\Omega \setminus N$, où N est négligeable. Par convention, l'intégrale de f sur Ω désigne alors l'intégrale de $f\mathbf{1}_N$ (qui est mesurable car $(\Omega, \mathcal{F}, \mu)$ est supposé complet).

Exemple 3.1. Pour tout $n \in \mathbb{N}$, considérons la fonction f_n définie sur $\mathbb{R}$ par

$$f_n(x) = \begin{cases} x^n e^{-x} & \text{si } x \in [1, n+1], \\ -1/n & \text{si } x = 0, \\ -n & \text{si } x = -1, \\ 0 & \text{sinon.} \end{cases}$$

La fonction f_n est borélienne car continue par morceaux sur $\mathbb{R}$. De plus, pour tout $x \in \mathbb{R} \setminus \{-1, 0, 1\}$, $(f_n(x))_{n \in \mathbb{N}}$ est positif et converge en croissant vers

$$f(x) = \begin{cases} +\infty & \text{si } x > 1 \\ 0 & \text{sinon.} \end{cases}$$

Or $\{-1, 0, 1\}$ est négligeable pour la mesure de Lebesgue, donc la suite $(f_n)_{n \in \mathbb{N}}$ est une suite de fonctions positives λ_1 -presque partout, croissante λ_1 -presque partout, et qui converge λ_1 -presque partout vers la fonction f . Donc, d'après le théorème de convergence monotone et d'après la convention $0 \times +\infty = 0$,

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n(x) \lambda_1(dx) = \int_{\mathbb{R}} f(x) \lambda_1(dx) = 0 \times \lambda_1([-\infty, 1]) + \infty \times \lambda_1([1, +\infty[) = +\infty.$$

Le corollaire suivant est une simple application du théorème de convergence monotone et du théorème du transport (sa démonstration est donc un bon exercice).

Corollaire 3.2.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires définies sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $[0, +\infty]$. Supposons que la suite $(X_n)_{n \in \mathbb{N}}$ est croissante presque sûrement et notons X sa limite (définie presque sûrement). Alors l'espérance de X est bien définie et

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

Exemple 3.2. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Pour tout $n \geq 1$, on définit la variable aléatoire $X_n = e^{-U/n}$. La fonction exponentielle étant à valeurs positives, les variables aléatoires X_n sont toutes positives presque sûrement. De plus, pour $n \geq 1$, $X_n \leq X_{n+1}$ presque sûrement, donc la suite $(X_n)_{n \geq 1}$ est croissante presque sûrement. Par conséquent, d'après le théorème de convergence monotone (et plus précisément d'après le corollaire 3.2), on a

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n \rightarrow \infty} X_n\right).$$

Or la suite de variables aléatoires $(X_n)_{n \geq 1}$ converge presque sûrement vers 1. En définitive,

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(1) = 1.$$

Terminons par un exemple dans lequel nous ne pouvons appliquer le corollaire 3.2, mais pour lequel nous pouvons la limite d'une suite d'espérance $(\mathbb{E}(X_n))_{n \in \mathbb{N}}$ grâce au théorème du transport et en appliquant le théorème de convergence monotone pour une mesure μ qui n'est pas une probabilité.

Exemple 3.3. Considérons une suite $(Y_n)_{n \geq 1}$ de variables aléatoires uniforme sur $[1, n+1]$. Autrement dit la loi de Y_n est absolument continue de densité $f_{Y_n} : \mathbb{R} \rightarrow \mathbb{R}_+$ donnée par

$$f_{Y_n}(y) = \frac{\mathbf{1}_{[1, n+1]}(y)}{n}.$$

Pour tout $n \in \mathbb{N}^*$, nous considérons alors la variable aléatoire

$$X_n = nY_n^n e^{-Y_n}.$$

La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ est une suite de variables positives presque sûrement car, par définition de la loi uniforme sur $[1, n+1]$, Y_n est positive presque sûrement. Mais elle n'est pas a priori croissante, car nous n'avons pas fait d'hypothèse de comparaison entre Y_n et Y_{n+1} et a priori ne converge pas presque sûrement quand $n \rightarrow +\infty$. Nous ne pouvons donc pas appliquer le corollaire 3.2 pour étudier la suite $\mathbb{E}(X_n)$ lorsque $n \rightarrow +\infty$. Néanmoins, d'après le théorème du transport,

$$\mathbb{E}(X_n) = \mathbb{E}(nY_n^n e^{-Y_n}) = \int_{\mathbb{R}} n y^n e^{-y} f_{Y_n}(y) \lambda_1(dy) = \int_{[1, n+1]} y^n e^{-y} \lambda_1(dy).$$

De même que dans l'exemple 3.1, le théorème de convergence monotone implique que

$$\lim_{n \rightarrow +\infty} \int_{[1, n+1]} y^n e^{-y} \lambda_1(dy) = +\infty, \quad \text{soit} \quad \lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = +\infty.$$

3.1.2 Théorème de convergence dominée

Intéressons-nous à présent au cas de fonctions de signes quelconques mais intégrables.

Théorème 3.3 (Théorème de convergence dominée ou théorème de Lebesgue).

Pour tout $n \in \mathbb{N}$, soit $f_n : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction mesurable. Supposons que

- (i) $(f_n)_{n \in \mathbb{N}}$ converge μ -presque partout,*
- (ii) il existe une fonction μ -intégrable g telle que, pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ μ -presque partout.*

Notons f la limite (définie μ -presque partout) de la suite $(f_n)_{n \in \mathbb{N}}$. Alors chaque fonction f_n et f sont μ -intégrables, de plus

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

Exemple 3.4. Pour tout $n \in \mathbb{N}$ et tout $x \in [1, +\infty[$, posons

$$f_n(x) = \frac{\sin^n x}{x^2}.$$

Alors les fonctions f_n sont toutes définies de $\Omega = [1, +\infty[$ vers $\mathbb{R}$ et boréliennes car continues. De plus,

$$|f_n(x)| \leq g(x) := \frac{1}{x^2}, \quad \forall n \in \mathbb{N}, \quad \forall x \in [1, +\infty[, \quad (3.1)$$

La fonction g définie par $g(x) = 1/x^2$ est Lebesgue-intégrable sur $[1, +\infty[$ car continue et Riemann-intégrable sur $[1, +\infty[$. Par suite, les fonctions f_n sont aussi Lebesgue-intégrables sur $[1, +\infty[$ et nous pouvons définir

$$u_n = \int_{[1, +\infty[} \frac{\sin^n x}{x^2} \lambda_1(dx).$$

Nous souhaitons étudier la limite quand $n \rightarrow +\infty$ de la suite $(u_n)_{n \in \mathbb{N}}$. Posons $N = \pi/2 + \pi\mathbb{Z}$ et remarquons que, pour tout $x \notin N$, $|\sin x| < 1$. Par conséquent, pour tout $x \in [1, +\infty[\setminus N$,

$$\lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow +\infty} \frac{\sin^n x}{x^2} = 0.$$

Or, l'ensemble N étant dénombrable, il est λ_1 -négligeable et, par conséquent, nous avons montré que $(f_n)_{n \in \mathbb{N}}$ converge vers 0 λ_1 -presque partout. De plus, (3.1) est satisfaite avec g intégrable, donc, d'après le théorème de convergence dominée,

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \int_{[1, +\infty[} f_n(x) \lambda_1(dx) = \int_{[1, +\infty[} 0 \lambda_1(dx) = 0.$$

Le corollaire suivant est une simple application du théorème de convergence dominée et du théorème du transport (sa démonstration est donc un bon exercice).

Corollaire 3.4.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\overline{\mathbb{R}}$. Supposons que

- (i) la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers une variable aléatoire X à valeurs dans $\overline{\mathbb{R}}$,*
- (ii) il existe une variable aléatoire Y intégrable telle que, pour tout $n \in \mathbb{N}$, $|X_n| \leq |Y|$ presque sûrement.*

Alors X est intégrable et

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

Remarque 3.2. Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires uniformément bornées (presque sûrement) par une constante $A \in \mathbb{R}_+$, c'est-à-dire si $|X_n| \leq A$ pour tout $n \in \mathbb{N}$ presque sûrement, alors l'hypothèse (ii) dans le corollaire ci-dessus est vérifiée.

Exemple 3.5. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Pour tout $n \geq 1$, on pose

$$X_n = \frac{1 + 1/n}{\sqrt{U} + 1/n}.$$

Nous souhaitons déterminer, si elle existe, la limite quand $n \rightarrow +\infty$ de $\mathbb{E}(X_n)$. On introduit pour cela la variable aléatoire X , définie par

$$X = \begin{cases} 1/\sqrt{U}, & \text{si } U \neq 0, \\ 0 & \text{si } U = 0. \end{cases}$$

Comme $\mathbb{P}(U = 0) = 0$ (car la loi de U est absolument continue), alors la suite $(X_n)_{n \geq 1}$ converge vers X . Par conséquent, la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X . De plus, en posant $Y = 2X$, on a $|X_n| \leq |Y|$ presque sûrement, où Y est intégrable. En effet, d'après le théorème du transport,

$$\mathbb{E}(|Y|) = \mathbb{E}\left(\left|\frac{2}{\sqrt{U}}\right|\right) = \int_{[0,1]} \frac{2}{\sqrt{u}} \lambda_1(du) = \int_{[0,1]} \frac{2}{\sqrt{u}} \lambda_1(du)$$

car $\lambda_1(\{0\}) = 0$. La fonction $u \mapsto \frac{2}{\sqrt{u}}$ étant continue Riemann-intégrable sur $]0, 1]$, elle est aussi Lebesgue-intégrable sur $]0, 1]$, c'est-à-dire que

$$\mathbb{E}(|Y|) = \int_{[0,1]} \frac{2}{\sqrt{u}} \lambda_1(du) < +\infty.$$

Les hypothèses du théorème de convergence dominée (et plus précisément du corollaire 1.4) sont donc satisfaites et, en utilisant à nouveau le théorème du transport, nous en déduisons que

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X) = \int_{[0,1]} \frac{1}{\sqrt{u}} \lambda_1(du) = [2\sqrt{u}]_{u \rightarrow 0_+}^{u=1} = 2$$

3.1.3 Lemme de Fatou

Dans de nombreuses situations, les variables aléatoires ou les fonctions considérées ne forment pas une suite convergente monotone ou dominée. Dans ces cas, on ne peut pas alterner limite et intégrale (ou limite et espérance) à l'aide des outils précédents. Toutefois, en introduisant la notion de limite inférieure (notée $\liminf$), on peut énoncer le lemme de Fatou qui lui s'applique en toute généralité, y compris pour des suites non-convergentes ! Cependant, le résultat obtenu n'est pas une égalité, mais seulement une inégalité. Il permet en particulier d'établir l'intégrabilité de certaines fonctions.

Définition 3.1.

Soit $(u_n)_{n \in \mathbb{N}}$ une suite de nombre réels. La limite inférieure de $(u_n)_{n \in \mathbb{N}}$ est donnée par

$$\liminf_{n \rightarrow +\infty} u_n = \sup_{n \in \mathbb{N}} \left(\inf_{k \geq n} u_k \right) \in \overline{\mathbb{R}}.$$

Remarque 3.3.

1. La limite inférieure d'une suite $(u_n)_{n \in \mathbb{N}}$ est égale à sa plus petite valeur d'adhérence.
 - Par exemple, si $u_n = (-1)^n$, alors $\liminf_{n \rightarrow +\infty} u_n = -1$.
 - Si $u_n = (-n)^3$, alors $\liminf_{n \rightarrow +\infty} u_n = -\infty$.
 - Si $u_n = (-1)^n + n^2$, alors $\liminf_{n \rightarrow +\infty} u_n = +\infty$.
 - Enfin, si la suite $(u_n)_{n \in \mathbb{N}}$ admet une limite ℓ (éventuellement égale à $\pm\infty$), alors sa seule valeur d'adhérence est ℓ , donc $\liminf_{n \rightarrow +\infty} u_n = \ell$.
2. On définit de même la limite supérieure d'une suite $(u_n)_{n \in \mathbb{N}}$ par

$$\limsup_{n \rightarrow +\infty} u_n = \inf_{n \in \mathbb{N}} \left(\sup_{k \geq n} u_k \right) \in \overline{\mathbb{R}}.$$

Il s'agit de la plus grande valeur d'adhérence de la suite $(u_n)_{n \in \mathbb{N}}$.

- Par exemple, si $u_n = (-1)^n$, alors $\limsup_{n \rightarrow +\infty} u_n = 1$.
 - Si $u_n = (-n)^3$, alors $\limsup_{n \rightarrow +\infty} u_n = +\infty$.
 - Enfin, si la suite $(u_n)_{n \in \mathbb{N}}$ admet une limite ℓ (éventuellement égale à $\pm\infty$), alors $\limsup_{n \rightarrow +\infty} u_n = \ell$.
3. Contrairement à la limite, les limites inférieures et supérieures d'une suite sont toujours bien définies. De plus, on peut montrer qu'une suite admet une limite (éventuellement égale à $-\infty$ ou $+\infty$) si et seulement si ses limites inférieures et supérieures sont égales. Dans ce cas, les limites, limites inférieures et limites supérieures de la suite sont égales.

Remarque 3.4. Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions définies sur Ω à valeurs dans $\overline{\mathbb{R}}$. On note $\liminf_{n \rightarrow \infty} f_n$ la fonction définie sur Ω par

$$\liminf_{n \rightarrow \infty} f_n : \omega \mapsto \liminf_{n \rightarrow \infty} f_n(\omega) \in \overline{\mathbb{R}}.$$

La limite inférieure d'une suite de fonctions à valeurs dans $\overline{\mathbb{R}}$ est toujours bien définie. De plus, si chaque fonction f_n est mesurable, alors la limite inférieure $\liminf_{n \rightarrow \infty} f_n$ est également mesurable.

Proposition 3.5 (Lemme de Fatou).

Pour tout $n \in \mathbb{N}$, soit $f_n : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une fonction mesurable. Si pour tout $n \in \mathbb{N}$, f_n est à valeurs dans $[0, +\infty]$ μ -presque partout, alors

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

Remarque 3.5. On trouve facilement des cas d'inégalité stricte. C'est le cas par exemple lorsque

$$f_n : x \in \mathbb{R} \mapsto \mathbf{1}_{[n, n+1]}(x) = \begin{cases} 1 & \text{si } x \in [n, n+1], \\ 0 & \text{sinon.} \end{cases}$$

On a alors $\liminf_{n \rightarrow \infty} f_n = 0$ et $\int_{\mathbb{R}} f_n d\lambda_1 = 1$ pour tout $n \geq 0$, donc

$$0 = \int_{\mathbb{R}} \liminf_{n \rightarrow \infty} f_n d\lambda_1 < \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} f_n d\lambda_1 = 1.$$

Cet exemple est un bon moyen mnémotechnique pour retenir le sens de l'inégalité dans le lemme de Fatou.

Exemple 3.6. Pour tout $n \in \mathbb{N}$, considérons la fonction $f_n :]0, 1[\rightarrow \mathbb{R}$ définie sur $]0, 1[$ par

$$f_n(x) = n \sin^2 \left(\frac{\sqrt{x}}{n^{1/3}} \right).$$

La fonction f_n est borélienne car continue sur $]0, 1[$ et est positive. En appliquant le lemme de Fatou à la suite $(f_n)_{n \in \mathbb{N}}$, nous obtenons :

$$+\infty = \int_{]0, 1[} \liminf_{n \rightarrow +\infty} f_n d\lambda_1 \leq \liminf_{n \rightarrow +\infty} \int_{]0, 1[} f_n d\lambda_1.$$

Par conséquent, $\liminf_{n \rightarrow +\infty} \int_{]0, 1[} f_n d\lambda_1 = +\infty$, donc $\lim_{n \rightarrow +\infty} \int_{]0, 1[} f_n d\lambda_1 = +\infty$.

Le corollaire suivant est une simple application du lemme de Fatou et du théorème du transport (sa démonstration est donc un bon exercice).

Corollaire 3.6.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $[0, +\infty]$ presque sûrement. On a alors

$$\mathbb{E} \left(\liminf_{n \rightarrow \infty} X_n \right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n).$$

Exemple 3.7. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Pour tout $n \geq 1$, on définit la variable aléatoire X_n par

$$X_n = n \left| \sin \left(\frac{1}{nU} \right) \right|.$$

La variable aléatoire X_n est bien définie presque sûrement (car $U \neq 0$ presque sûrement) et positive. De plus, quand $n \rightarrow \infty$, elle converge presque sûrement vers $1/U$ (qui est donc également sa limite inférieure). Donc, d'après le lemme de Fatou,

$$+\infty = \mathbb{E} \left(\frac{1}{U} \right) = \mathbb{E} \left(\liminf_{n \rightarrow \infty} X_n \right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n).$$

On en déduit donc que $\mathbb{E}(X_n)$ converge vers $+\infty$ quand $n \rightarrow \infty$.

3.2 Applications des théorèmes de convergence aux intégrales paramétrées

Nous donnons des critères de continuité et de dérivabilité des fonctions définies par une intégrale.

Théorème 3.7 (Théorème de continuité).

Supposons que E est un espace métrique (i.e. est muni d'une distance d_E ²). Considérons une fonction $f : E \times \Omega \rightarrow \mathbb{R}$ (ou $\mathbb{C}$) et $a \in E$. Supposons que

- (i) l'application $\omega \mapsto f(x, \omega)$ est mesurable, pour tout $x \in E$,*
- (ii) l'application $x \mapsto f(x, \omega)$ est continue en $a \in E$, $\mu(\omega)$ -presque partout,*
- (iii) il existe $g : \Omega \rightarrow \mathbb{R}_+$ μ -intégrable telle que $|f(x, \omega)| \leq g(\omega)$ $\mu(\omega)$ -presque partout, pour tout $x \in E$.*

Alors la fonction $F : x \mapsto \int_{\Omega} f(x, \omega) \mu(d\omega)$ est définie sur E et est continue en a .

Remarque 3.6. La proposition peut être « localisée ». En effet, il suffit que la propriété (iii) soit réalisée pour tout x dans une boule ouverte $B(a, r)$ de centre a et de rayon $r > 0$, auquel cas F est bien définie sur $B(a, r)$.

Démonstration.

- Notons tout d'abord que les hypothèses (i) et (iii) assurent que pour tout $x \in E$, la fonction $\omega \mapsto f(x, \omega)$ est μ -intégrable sur Ω . Par suite, pour tout $x \in E$,

$$F(x) = \int_{\Omega} f(x, \omega) \mu(d\omega) \quad \text{est bien définie.}$$

- Étudions donc à présent la continuité de la fonction F en a . Considérons pour cela $(a_n)_{n \in \mathbb{N}}$ une suite de points de E convergeant vers a . Pour tout $n \in \mathbb{N}$, considérons alors la fonction f_n définie sur E par

$$\forall n \in \mathbb{N}, \forall \omega \in \Omega, f_n(\omega) = f(a_n, \omega).$$

L'hypothèse (i) assure que chaque f_n est une fonction mesurable. De plus, d'après l'hypothèse (ii), pour μ -presque tout $\omega \in \Omega$,

$$\lim_{n \rightarrow +\infty} f_n(\omega) = f(a, \omega).$$

Par ailleurs, d'après l'hypothèse (iii), pour tout $n \in \mathbb{N}$,

$$|f_n| \leq g \quad \mu\text{-presque partout}$$

où la fonction g est μ intégrable. Par suite, le théorème de convergence dominée s'applique à la suite f_n et conduit à :

$$\lim_{n \rightarrow +\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f(a, \cdot) d\mu = F(a).$$

Autrement dit, par définition de f_n

$$\lim_{n \rightarrow +\infty} F(a_n) = F(a).$$

L'espace E étant un métrique, et l'égalité ci-dessus étant vraie pour toute suite $(a_n)_{n \in \mathbb{N}}$ de E convergeant vers a , la fonction F est continue en a . ■

2. Par exemple, les ensembles (et leurs sous-ensembles) $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{R}^d$ munis de la distance euclidienne sont métrique, de même les ensembles de fonctions bornées munies de la norme infinie.

Exemple 3.8. Soit X une variable aléatoire à valeurs réelles. Pour tout $u \in \mathbb{R}$, la variable aléatoire $\arctan(uX)$ est de valeur absolue bornée par $\pi/2$ donc elle est intégrable et $\mathbb{E}(\arctan(uX))$ est donc bien définie. On définit la fonction

$$h(u) = \mathbb{E}(\arctan(uX)), \quad \forall u \in \mathbb{R}.$$

Montrons que cette fonction est continue. Afin de se ramener au théorème de continuité des intégrales paramétriques, nous écrivons h sous la forme d'une intégrale à l'aide du théorème du transport :

$$h(u) = \int_{\mathbb{R}} \arctan(ux) \mathbb{P}_X(dx), \quad \forall u \in \mathbb{R},$$

où $\mathbb{P}_X$ est la loi de X . Introduisons la fonction $g : x \in \mathbb{R} \mapsto \pi/2$ et vérifions les hypothèses du théorème de continuité des intégrales paramétrées³. Pour tout $u_0 \in \mathbb{R}$,

- (i) pour tout $u \in \mathbb{R}$, $x \mapsto \arctan(ux)$ est mesurable,
- (ii) Pour tout $x \in \mathbb{R}$, l'application $u \mapsto \arctan(ux)$ est continue en u_0 ,
- (iii) g est $\mathbb{P}_X$ intégrable, car elle est uniformément bornée par $\pi/2$ et $\mathbb{P}_X$ est une probabilité; de plus, pour tout $u \in \mathbb{R}$ et tout $x \in \mathbb{R}$, $|\arctan(ux)| \leq \pi/2 = g(x)$.

D'après le théorème de continuité, la fonction h est donc continue en u_0 , et ce pour tout $u_0 \in \mathbb{R}$.

Théorème 3.8 (Théorème de dérivabilité).

Soient U un ouvert de $\mathbb{R}$, $f : U \times \Omega \rightarrow \mathbb{R}$ et $g : \Omega \rightarrow \mathbb{R}_+$ tels que

- (i) l'application $\omega \mapsto f(x, \omega)$ est mesurable et μ -intégrable, pour tout $x \in U$,
- (ii) $x \mapsto f(x, \omega)$ est dérivable sur U , pour μ -presque tout $\omega \in \Omega$,
- (iii) g est μ -intégrable et $\left| \frac{\partial f}{\partial x}(x, \omega) \right| \leq g(\omega)$ pour tout $x \in U$, $\mu(\omega)$ -presque partout.

Alors, pour tout $x \in U$, la fonction

$$\begin{aligned} \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto \frac{\partial f}{\partial x}(x, \omega) \end{aligned}$$

est mesurable. De plus, la fonction $F : x \mapsto \int_{\Omega} f(x, \omega) \mu(d\omega)$ est bien définie et dérivable sur U . Enfin,

$$\forall x \in U, F'(x) = \int_{\Omega} \frac{\partial f}{\partial x}(x, \omega) \mu(d\omega).$$

Remarque 3.7.

1. Si U est un ouvert de $\mathbb{R}^n$ et si nous remplaçons dans la dernière proposition (ii) par

$x \mapsto f(x, \omega)$ est différentiable par rapport à x_i , pour presque tout $\omega \in \Omega$,

alors F est différentiable sur U par rapport à x_i et

$$\forall x \in U, \frac{\partial F}{\partial x_i}(x) = \int_{\Omega} \frac{\partial f}{\partial x_i}(x, \omega) \mu(d\omega).$$

2. En itérant le théorème 3.8, nous pouvons obtenir des critères pour montrer que F est de classe $\mathcal{C}^k$.

3. Attention! Comme souvent, les notations ne sont pas les mêmes que dans le théorème : ici on intègre par rapport à x et on cherche à vérifier la continuité par rapport à u .

Démonstration Notons tout d'abord que l'hypothèse (i) assure que la fonction F est bien définie.

Soient $x, a \in U$ tels que $x \neq a$. Alors,

$$\frac{F(x) - F(a)}{x - a} = \int_{\Omega} \frac{f(x, \omega) - f(a, \omega)}{x - a} \mu(d\omega).$$

En utilisant le théorème des accroissements finis et l'hypothèse (iii), nous constatons que

$$\left| \frac{f(x, \omega) - f(a, \omega)}{x - a} \right| \leq g(\omega) \text{ } \mu\text{-presque partout.}$$

De plus,

$$\lim_{x \rightarrow a} \frac{f(x, \omega) - f(a, \omega)}{x - a} = \frac{\partial f}{\partial x}(a, \omega) \text{ } \mu(\omega)\text{-presque partout.}$$

Le lecteur conclura en appliquant le théorème de convergence dominée. ■

Exemple 3.9. Soit X une variable aléatoire à valeurs réelles de loi gaussienne centrée réduite, c'est-à-dire que la loi de X , notée $\mathbb{P}_X$, est donnée par

$$\mathbb{P}_X(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \lambda_1(dx).$$

Pour tout $u \in]-1, 1[$, la variable aléatoire e^{uX} est positive donc $\mathbb{E}(e^{uX})$ est bien définie. On définit la fonction

$$h(u) = \mathbb{E}(e^{uX}), \quad \forall u \in]-1, 1[.$$

Notre but est de montrer que cette fonction est dérivable et d'écrire sa dérivée. Afin de se ramener au théorème de dérivation des intégrales paramétrées, nous écrivons h sous la forme d'une intégrale à l'aide du théorème du transport :

$$h(u) = \int_{\mathbb{R}} e^{ux} \mathbb{P}_X(dx) = \int_{\mathbb{R}} e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \lambda_1(dx), \quad \forall u \in \mathbb{R}.$$

Introduisons la fonction

$$g : x \in \mathbb{R} \mapsto \frac{|x|}{\sqrt{2\pi}} e^{|x|} e^{-x^2/2}$$

et vérifions les hypothèses du théorème de continuité des intégrales paramétrées⁴ :

- (i) pour tout $u \in]-1, 1[$, $x \mapsto e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est mesurable (car continue),
- (ii) Pour tout $x \in \mathbb{R}$, l'application $u \mapsto e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est dérivable en tout point de $] -1, 1[$,
- (iii) g est bornée et dominée par $x \mapsto 1/x^2$ en $\pm\infty$, donc elle est $\lambda_1(x)$ -intégrable ; de plus, pour tout $u \in]-1, 1[$ et tout $x \in \mathbb{R}$,

$$\left| \frac{\partial}{\partial u} e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = \left| x e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \leq g(x)$$

D'après le théorème de dérivation, la fonction h est dérivable sur $] -1, 1[$ et

$$\frac{\partial h}{\partial u}(u) = \int_{\mathbb{R}} \frac{\partial}{\partial u} e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \lambda_1(dx) = \int_{\mathbb{R}} x e^{ux} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \lambda_1(dx).$$

Ainsi, en utilisant à nouveau le théorème du transport,

$$\frac{\partial h}{\partial u}(u) = \mathbb{E}(X e^{uX}), \quad \forall u \in]-1, 1[.$$

4. Attention ! Comme dans le cas précédent, les notations ne sont pas les mêmes que dans le théorème : ici on intègre par rapport à x et on cherche à vérifier la dérivabilité par rapport à u .

3.3 Théorème de changement de variable

Supposons que la loi de X est absolument continue de densité $f_X : \mathbb{R}^d \rightarrow [0, +\infty]$. Partant de la proposition 2.7, si nous trouvons une mesure positive ν telle que pour toute fonction φ borélienne positive bornée,

$$\mathbb{E}(\varphi(Y)) = \int_{\mathbb{R}^d} \varphi(y) \nu(dy)$$

alors ν est une probabilité (prendre $\varphi = 1$) et est la loi de $Y = g(X)$. Utilisant le théorème du transport, nous pouvons tout d'abord écrire que

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi \circ g(X)) = \int_{\mathbb{R}^d} \varphi \circ g(x) f_X(x) d\lambda_d(x). \quad (3.2)$$

Il est ensuite naturel de se poser la question suivante : est-il possible d'effectuer le changement de variable $y = g(x)$ dans l'intégrale de Lebesgue ci-dessus ? Même si φ n'est pas régulière, le changement de variable dans l'intégrale de Lebesgue est valable sous les mêmes hypothèses que dans le cadre de l'intégrale au sens de Riemann. Il est ainsi possible d'utiliser par exemple le théorème suivant.

Théorème 3.9.

Soit U un ouvert de $\mathbb{R}^d$ et $g : U \rightarrow \mathbb{R}^d$ une application injective de classe C^1 dont la matrice jacobienne $\text{Jac } g$, définie par

$$\text{Jac } g(x) = \left(\frac{\partial g_i}{\partial x_j}(x) \right)_{1 \leq i, j \leq d}, \quad \forall x \in U,$$

est inversible pour tout $x \in U$. Posons $V = g(U)$.

1. Alors, pour toutes fonctions mesurables positives $\varphi : V \rightarrow [0, +\infty]$ et $f : U \rightarrow [0, +\infty]$,

$$\int_U \varphi \circ g(x) f(x) \lambda_d(dx) = \int_V \varphi(y) f \circ g^{-1}(y) |\det \text{Jac}(g^{-1})(y)| \lambda_d(dy),$$

où $\text{Jac}(g^{-1})$ est la matrice jacobienne de l'inverse de g .

2. La formule ci-dessus reste vraie pour toutes fonctions mesurables φ et f telle que $f \times (\varphi \circ g)$ est Lebesgue-intégrable.

Revenons à l'expression (3.2). La question à se poser est donc : peut-on poser $y = g(x)$ et appliquer un théorème de changement de variable ? Notons d'ores et déjà que ceci suppose que $g(X)$ et X sont toutes deux à valeurs dans le même espace $\mathbb{R}^d$. De plus, dans l'expression (3.2), il semble nécessaire d'avoir $U = \mathbb{R}^d$, or, en général, g ne satisfait pas les hypothèses du théorème de changement de variable dans ce cadre général. Voici quelques méthodes pour s'en sortir.

1er cas : Avant de vérifier des hypothèses sur g , supposons qu'il existe un ouvert $U \subset \mathbb{R}^d$ tel que f_X est nulle λ_d -presque partout en dehors de U , c'est-à-dire que $\lambda_d(\{x \notin U, f_X(x) \neq 0\}) = 0$. Ainsi, (3.2) peut être réécrite sous la forme

$$\mathbb{E}(\varphi(X)) = \int_U \varphi(g(x)) f_X(x) \lambda_d(dx).$$

Si g satisfait à présent les hypothèses du théorème 3.9 avec ce choix de U , nous pouvons poser $y = g(x)$ et réécrire l'expression (3.2) sous la forme :

$$\mathbb{E}(\varphi(Y)) = \int_{g(U)} \varphi(y) f_X(g^{-1}(y)) |\det \text{Jac}(g^{-1})(y)| \lambda_d(dy) = \int_{\mathbb{R}^d} \varphi(y) \nu(dy)$$

avec ν la mesure absolument continue de densité

$$y \mapsto f_X(g^{-1}(y)) |\det \text{Jac}(g^{-1})(y)| \mathbf{1}_{y \in g(U)}.$$

L'écriture ci-dessus étant vraie pour toute fonction mesurable positive φ , on déduit du théorème du transport que la loi de $Y = g(X)$ est la mesure de probabilité ν .

Exemple 3.10. Supposons que X suit la loi uniforme sur $[0, 1]$. Et considérons la variable $Y = X^2$. Rappelons que la densité f_X de X est définie par

$$f_X(x) = \mathbf{1}_{[0,1]}(x), \quad x \in \mathbb{R}.$$

Fixons $\varphi : \mathbb{R} \rightarrow [0, +\infty]$ une fonction borélienne positive. Alors, par le théorème du transport,

$$\mathbb{E}(\varphi(Y)) = \mathbb{E}(\varphi(X^2)) = \int_{\mathbb{R}} \varphi(x^2) f_X(x) \lambda_1(dx) = \int_{[0,1]} \varphi(x^2) \lambda_1(dx).$$

Notons que comme $\lambda_1(\{0, 1\}) = 0$,

$$\mathbb{E}(\varphi(Y)) = \int_{]0,1[} \varphi(x^2) \lambda_1(dx).$$

Le lecteur vérifiera que la fonction $g : x \mapsto x^2$ satisfait les hypothèses du théorème 3.9 en prenant $U =]0, 1[$ et $V = g(U) =]0, 1[$. En appliquant ce théorème de changement de variable, le lecteur vérifiera aussi que

$$\mathbb{E}(\varphi(Y)) = \int_{]0,1[} \frac{\varphi(y)}{2\sqrt{y}} \lambda_1(dy).$$

Ceci étant vrai pour toute fonction borélienne positive φ , la loi de Y est absolument continue de densité

$$\begin{aligned} f_Y : \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ y &\longmapsto \frac{\mathbf{1}_{]0,1[}(y)}{2\sqrt{y}} \end{aligned}$$

En appliquant la démarche proposée dans ce premier cas, il est aussi possible d'établir des propriétés de stabilité pour certains ensembles de lois classiques.

Proposition 3.10.

Soient $a, b \in \mathbb{R}$ et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle.

- 1. Si X est de loi uniforme sur $[c, d]$ (avec $c, d \in \mathbb{R}$ et $c < d$) et si $a \neq 0$, alors $aX + b$ suit la loi uniforme sur l'intervalle $[ac + b, ad + b]$.*
- 2. Si X est de loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$ et si $a > 0$, alors aX suit la loi exponentielle de paramètre $a\lambda$.*
- 3. Si X suit la loi gaussienne de paramètre $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, alors $aX + b$ suit la loi gaussienne de paramètre $(am + b, a^2\sigma^2)$.*

2nd cas : Dans le cas où g ne satisfait pas directement le théorème du changement de variable après avoir choisi U , il faut être plus astucieux. L'idée est de découper l'intégrale sur $\mathbb{R}^d$ en K intégrales, chacune sur un ensemble U_i choisi de sorte à pouvoir appliquer le théorème du changement de variables. Le plus simple est d'illustrer ce cas par un exemple.

Exemple 3.11. Soit X une variable aléatoire de loi uniforme sur $[-1, 2]$. Déterminons la loi de la variable aléatoire $Y = X^2$. Fixons $\varphi : \mathbb{R} \rightarrow [0, +\infty]$ une fonction borélienne positive. Alors, par le théorème du transport,

$$\mathbb{E}(\varphi(Y)) = \frac{1}{3} \int_{[-1, 2]} \varphi(x^2) \lambda_1(dx) = \frac{1}{3} \int_{]-1, 2[} \varphi(x^2) \lambda_1(dx)$$

car $\lambda_1(\{-1, 2\}) = 0$. Nous souhaitons poser $y = x^2$ mais ceci ne définit pas un changement de variable sur $U =]-1, 2[$ (car $x \mapsto x^2$ n'est pas bijective). Par contre $x \mapsto x^2$ est bijective de $] -1, 0[$ (respectivement $]0, 2[$) sur $]0, 1[$ (respectivement $]0, 4[$). L'idée est donc de décomposer l'intégrale en deux :

$$\mathbb{E}(\varphi(Y)) = \frac{1}{3} \int_{]-1, 0[} \varphi(x^2) \lambda_1(dx) + \frac{1}{3} \int_{]0, 2[} \varphi(x^2) \lambda_1(dx)$$

d'après la relation de Chasles et car $\lambda_1(\{0\}) = 0$. Il devient alors possible d'effectuer le changement de variable $y = x^2$ sur l'ouvert $U =]-1, 0[$ puis sur l'ouvert $U =]0, 2[$. Le lecteur vérifiera que l'on obtient ainsi :

$$\mathbb{E}(\varphi(Y)) = \frac{1}{3} \int_{]0, 1[} \frac{\varphi(y)}{2\sqrt{y}} \lambda_1(dy) + \frac{1}{3} \int_{]0, 4[} \frac{\varphi(y)}{2\sqrt{y}} \lambda_1(dy) = \int_{\mathbb{R}} \frac{\varphi(y)}{3\sqrt{y}} \left(\mathbf{1}_{]0, 1[}(y) + \frac{1}{2} \mathbf{1}_{]1, 4[}(y) \right) \lambda_1(dy).$$

Ceci étant vrai pour toute fonction borélienne positive φ , la loi de Y est absolument continue de densité

$$\begin{aligned} f_Y : \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ y &\longmapsto \frac{1}{3\sqrt{y}} \left(\mathbf{1}_{]0, 1[}(y) + \frac{1}{2} \mathbf{1}_{]1, 4[}(y) \right). \end{aligned}$$

Chapitre 4

Fonction de répartition et fonction caractéristique

Dans le chapitre précédent, nous avons défini les notions de loi et d'espérance d'une variable aléatoire. Dans ce chapitre, nous allons proposer des méthodes pour calculer et identifier une loi à travers sa fonction de répartition et sa fonction caractéristique.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités (qui ne sera, sauf exception, pas explicité) ;
- $\mathbb{R}^d$ est muni de sa tribu borélienne.

4.1 Fonction de répartition

4.1.1 Définition pour une v.a.r. et propriétés générales

Définition 4.1 (Fonction de répartition d'une variable aléatoire réelle).

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle de loi $\mathbb{P}_X$. **La fonction de répartition de X** est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$\forall t \in \mathbb{R}, F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X^{-1}([-\infty, t])) = \mathbb{P}_X([-\infty, t]).$$

Exemple 4.1. Considérons une variable aléatoire réelle X . Notons $\mathbb{P}_X$ sa loi et F_X sa fonction de répartition.

1. Soit $a \in \mathbb{R}$. Supposons que $X = a$ presque sûrement, c'est-à-dire que $\mathbb{P}_X = \delta_a$. Alors,

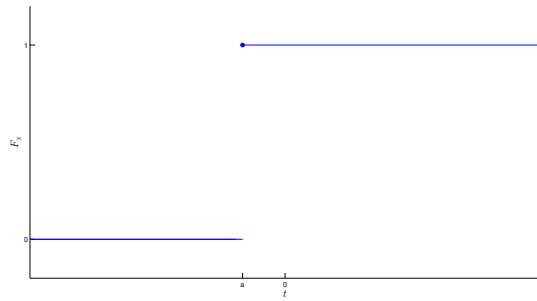
$$F_X(t) = \mathbb{P}_X([-\infty, t]) = \begin{cases} 0 & \text{si } t \in]-\infty, a[\\ 1 & \text{si } t \in [a, +\infty[. \end{cases}$$

Remarquons que la fonction de répartition de la loi discrète $\mathbb{P}_X = \delta_a$ est constante par morceaux (voir figure 4.1a). Nous verrons que ceci est caractéristique des lois discrètes.

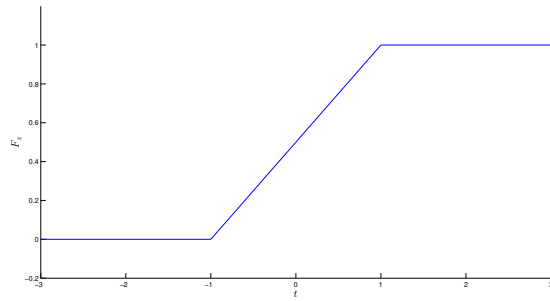
2. Supposons que X est de loi uniforme sur $[a, b]$ ($a < b$). Alors,

$$F_X(t) = \frac{1}{b-a} \int_{-\infty}^t \mathbf{1}_{[a,b]}(x) \lambda_1(dx) = \frac{\lambda_1([-\infty, t] \cap [a, b])}{b-a} = \begin{cases} 0 & \text{si } t < a, \\ \frac{t-a}{b-a} & \text{si } a \leq t < b, \\ 1 & \text{si } t \geq b. \end{cases}$$

Remarquons que la fonction de répartition de la loi uniforme sur $[a, b]$ est continue et même $\mathcal{C}^1$ sauf en quelques points (voir figure 4.1b). Nous verrons que ceci est caractéristique des lois absolument continues.



(a) Fonction de répartition de la loi δ_a



(b) Fonction de répartition de la loi uniforme sur $[-1, 1]$.

FIGURE 4.1 – Exemples de fonctions de répartition

La loi d'une variable aléatoire X à valeurs dans $\mathbb{R}$ est caractérisée par sa fonction de répartition

Proposition 4.1 (Caractérisation de la loi d'une variable aléatoire réelle).

Soient X et Y deux variables aléatoires réelles. Les variables aléatoires X et Y ont même fonction de répartition si et seulement si elles ont même loi.

Terminons par quelques propriétés générales sur les fonctions de répartitions.

Proposition 4.2.

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle.

1. *La fonction de répartition F_X de X est une fonction croissante, continue à droite telle que*

$$\lim_{t \rightarrow -\infty} F_X(t) = 0 \quad \text{et} \quad \lim_{t \rightarrow +\infty} F_X(t) = 1. \quad (4.1)$$

2. *Par ailleurs, F_X admet une limite à gauche en tout point et*

$$\forall t \in \mathbb{R}, F_X(t_-) = \lim_{x \rightarrow t_-} F_X(x) = \mathbb{P}_X(]-\infty, t]). \quad (4.2)$$

3. *De plus, F_X admet au plus un nombre fini ou dénombrable de points de discontinuité.*

4. *Enfin, pour tous réels a, b tels que $a < b$,*

$$\left\{ \begin{array}{l} \mathbb{P}_X(]a, b]) = \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a), \\ \mathbb{P}_X([a, b]) = \mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a_-), \\ \mathbb{P}_X([a, b[) = \mathbb{P}(a \leq X < b) = F_X(b_-) - F_X(a_-), \\ \mathbb{P}_X([a, +\infty[) = \mathbb{P}(X \geq a) = 1 - F_X(a_-), \\ \mathbb{P}_X(]a, +\infty[) = \mathbb{P}(X > a) = 1 - F_X(a). \end{array} \right.$$

Preuve de la proposition 4.2 :

1. • Soient $t, s \in \mathbb{R}$ tels que $s \leq t$. Alors, $]-\infty, s] \subset]-\infty, t]$ et donc par croissance de $\mathbb{P}_X$ (cf. proposition 1.1 du chapitre 1 page 5),

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

L'inégalité précédente étant vraie pour tous $t, s \in \mathbb{R}$ tels que $s \leq t$, la fonction F_X est croissante sur $\mathbb{R}$.

- La continuité à droite de F_X , la propriété (4.1) s'obtiennent en utilisant la continuité monotone de $\mathbb{P}_X$ (cf. proposition 1.1 du chapitre 1 page 5).

2. La propriété (4.2) s'obtient aussi en utilisant la continuité monotone de $\mathbb{P}_X$.

3. La fonction F_X étant croissante continue à droite, l'ensemble de ses points de discontinuité est

$$S = \bigcup_{n \in \mathbb{N}^*} D_n \quad \text{avec} \quad D_n = \left\{ x \in \mathbb{R} / F_X(x) - F_X(x_-) \geq \frac{1}{n} \right\}.$$

Étant donné que $0 \leq F_X \leq 1$ et que F_X est croissante, pour tout $n \in \mathbb{N}^*$, D_n est de cardinal fini car

$$\frac{\text{Card} D_n}{n} \leq \sum_{x \in D_n} (F_X(x) - F_X(x_-)) \leq 1.$$

Par conséquent, S est un ensemble fini ou dénombrable.

4. Laissé en exercice. ■

Remarque 4.1. En fait, l'assertion 1. de la proposition 4.2 caractérise la notion de fonction de répartition d'une variable aléatoire réelle. En effet, on peut montrer que, pour toute fonction F satisfaisant cette assertion, il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire à valeurs réelles X tels que $F_X = F$.

4.1.2 Variables aléatoires réelles discrètes

Cette partie s'intéresse aux fonctions de répartition de variables aléatoires discrètes à valeurs dans $\mathbb{R}$. La preuve de la proposition suivante est laissée en exercice.

Proposition 4.3.

Soit $I = \{1, \dots, n\}$ (avec $n \in \mathbb{N}^*$) ou $I = \mathbb{N}^*$ et considérons

$$S = \{x_i / i \in I\} \subset \mathbb{R}$$

une famille strictement croissante de réels (i.e. telle que $x_i < x_{i+1}$ pour tout $i \in I$ tel que $i+1 \in I$). Supposons que X une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs $\mathbb{P}$ -presque sûrement dans S . Rappelons que la loi de X s'écrit alors

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i} \quad \text{avec} \quad p_i = \mathbb{P}(X = x_i) \in [0, 1].$$

1. Nous avons

$$F_X(t) = \begin{cases} 0 & \text{si } t \in]-\infty, x_1[\\ p_1 + \dots + p_i & \text{si } x_i \leq t < x_{i+1} \text{ avec } 1 \leq i < n \\ 1 & \text{si } t \in [\sup S, +\infty[. \end{cases}$$

2. De plus, la fonction F_X est constante par morceaux et continue sauf éventuellement aux points x_i , $i \in I$. Enfin, pour tout $i \in I$,

$$p_i = \mathbb{P}(X = x_i) = F_X(x_i) - F_X(x_{i-})$$

est la valeur du saut de la fonction F_X au point x_i .

Exemple 4.2. Soit X une variable aléatoire de loi binomiale de paramètre $(2, 1/3)$, c'est-à-dire de loi

$$\mathbb{P}_X = \sum_{k=0}^2 C_2^k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{2-k} \delta_k = \frac{4}{9}\delta_0 + \frac{4}{9}\delta_1 + \frac{1}{9}\delta_2.$$

Alors, d'après la formule donnée dans la proposition précédente, la fonction de répartition de X est définie par

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{4}{9} & \text{si } 0 \leq t < 1 \\ \frac{4}{9} + \frac{4}{9} = \frac{8}{9} & \text{si } 1 \leq t < 2 \\ 1 & \text{si } t \geq 2. \end{cases}$$

Le tracé de cette fonction est représentée dans la figure 4.2a.

Terminons par le problème inverse. Se donnant une fonction $F : \mathbb{R} \rightarrow \mathbb{R}_+$, nous nous demandons si il existe une loi discrète dont F est la fonction de répartition et, si oui, quelle est la loi associée? Les conditions nécessaires de la proposition 4.2 et l'assertion 2. de la proposition 4.3 s'avèrent être aussi suffisantes.

Proposition 4.4.

Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante et continue à droite telle que

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

Supposons que F est constante par morceaux, au sens où il existe une famille strictement croissante de réels

$$S = \{x_i / i \in I\} \subset \mathbb{R}$$

finie ou dénombrable telle que F est constante sur chaque intervalle $[x_i, x_{i+1}[$.

1. Alors, il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire réelle discrète X définie sur cet espace dont F est la fonction de répartition.
2. De plus, pour tout $i \in I$, $\mathbb{P}(X = x_i) = F(x_i) - F(x_{i-})$ la loi de X est

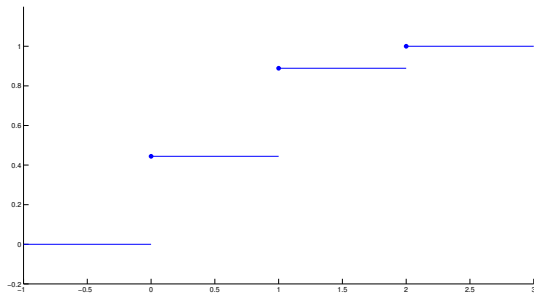
$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$$

avec pour tout $i \in I$, $p_i = \mathbb{P}(X = x_i) = F(x_i) - F(x_{i-})$ la valeur du saut de F au point x_i .

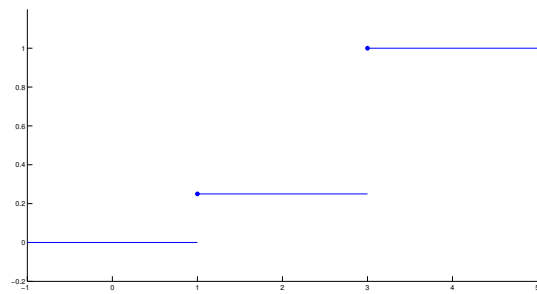
Exemple 4.3. Considérons la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ représentée sur la figure 4.2b et définie par

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.25 & \text{si } 1 \leq x < 3 \\ 1 & \text{si } x \geq 3. \end{cases}$$

Le lecteur vérifiera que la fonction F satisfait bien les hypothèses de la proposition 4.4. En particulier, elle est constante, sauf en $x = 1$ et en $x = 3$. De plus, son saut en $x = 1$ vaut $1/4$ et son saut en $x = 3$ vaut $3/4$. Par suite, F est la fonction de répartition d'une variable aléatoire X de loi $\mathbb{P}_X = \frac{1}{4}\delta_1 + \frac{3}{4}\delta_3$.



(a) Fonction de répartition d'une variable de loi $\mathcal{B}(2, 1/3)$.



(b) Fonction $F = \frac{1}{4}\mathbf{1}_{[1,3[} + \frac{3}{4}\mathbf{1}_{[3,+\infty[}$.

FIGURE 4.2 – Exemples de fonctions de répartition de lois discrètes

4.1.3 Variables aléatoires réelles absolument continues

Étudions à présent les fonctions de répartition des variables aléatoires réelles de loi absolument continue.

Proposition 4.5.

Soit X une variable aléatoire réelle de loi absolument continue de densité f_X . Notons F_X sa fonction de répartition.

1. Alors, d'après le théorème du transport,

$$\forall t \in \mathbb{R}, F_X(t) = \mathbb{P}(X \leq t) = \mathbb{E}(\mathbf{1}_{X \leq t}) = \int_{]-\infty, t]} f_X(x) \lambda_1(dx).$$

2. De plus, la fonction F_X est continue. Elle est aussi dérivable λ_1 -presque partout et $F'_X = f_X$ λ_1 -presque partout.

3. Enfin, si la densité f_X est continue sur $\mathbb{R}$, alors la fonction F_X est $\mathcal{C}^1$ sur $\mathbb{R}$ et

$$\forall t \in \mathbb{R}, F'_X(t) = f_X(t).$$

Exemple 4.4. Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$, i.e. de densité f_X avec

$$f_X(x) = \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

Notons F_X la fonction de répartition de X . D'après le théorème du transport,

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{]-\infty, t]} f_X d\lambda_1 = \int_{]-\infty, t]} \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{\mathbb{R}_+^*}(x) d\lambda_1(x).$$

Pour $t \leq 0$, $F_X(t) = \int_{-\infty}^t 0 \, dx = 0$. Et, pour $t > 0$,

$$F_X(t) = \int_{-\infty}^t f_X \, d\lambda_1 = \int_{-\infty}^t \frac{e^{-x/\lambda}}{\lambda} dx = \left[e^{-x/\lambda} \right]_{x \rightarrow -\infty}^{x=t} = 1 - e^{-t/\lambda}.$$

Ainsi, la fonction de répartition de X est donnée par

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq 0, \\ 1 - e^{-t/\lambda} & \text{si } t > 0. \end{cases}$$

La fonction F_X est représentée sur la figure 4.3a.

Terminons par le problème inverse. Se donnant une fonction $F : \mathbb{R} \mapsto \mathbb{R}_+$, existe-t-il des conditions suffisantes pour que F soit la fonction de répartition d'une loi absolument continue. Les propositions 4.2 et 4.5 donnent des conditions nécessaires, qui ne sont en général pas suffisantes. La proposition suivante donne une condition suffisante proche mais qui impose une propriété de régularité à F .

Proposition 4.6.

Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante, continue et $\mathcal{C}^1$ par morceaux telle que

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

Alors, il existe une variable aléatoire X définie sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$ dont F est la fonction de répartition. De plus, la loi de X est la loi absolument continue de densité $f = F'$ (fonction bien définie sauf en les points où F n'est pas $\mathcal{C}^1$).

Remarque 4.2. Dans la proposition précédente, la fonction f n'est pas a priori bien définie en x_i , $i \in I$. Ceci n'est pas un problème car $\{x_i, i \in I\}$ est λ_1 -négligeable car fini ou dénombrable, donc la loi de densité f ne dépend pas des valeurs de f en les x_i , qui peuvent donc être choisies arbitrairement.

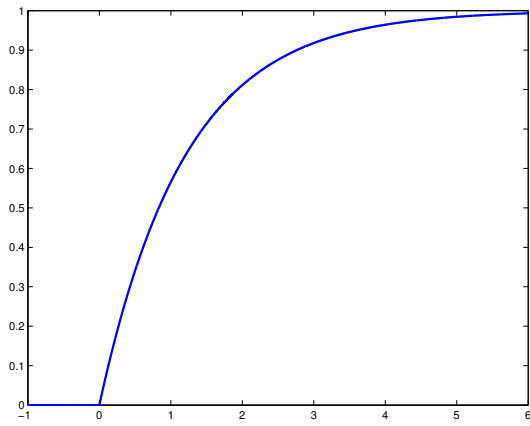
Exemple 4.5. Considérons la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ représentée sur la figure 4.3b et définie par

$$F(t) = \begin{cases} 0 & \text{si } t < 0 \\ t^2 & \text{si } t \in [0, 1[\\ 1 & \text{si } t \geq 1. \end{cases}$$

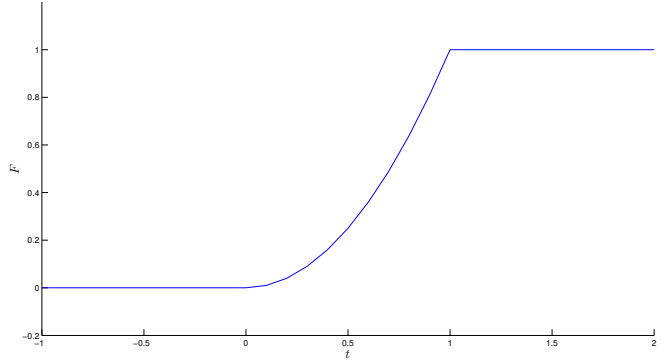
La fonction F est continue croissante sur $\mathbb{R}$ telle que

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

Par ailleurs, F est $\mathcal{C}^1$ sauf en 1. Par conséquent, F est la fonction de répartition d'une variable aléatoire de loi absolument continue de densité $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = 2x\mathbf{1}_{]0,1[}(x)$.



(a) Fonction de répartition de la loi $\mathcal{E}(1)$



(b) Fonction F de l'exemple 4.5

FIGURE 4.3 – Exemples de fonctions de répartition de lois discrètes

4.1.4 Mélange de lois discrètes et absolument continues

Nous considérons à présent le cas d'une variable aléatoire dont la loi est la somme d'une mesure discrète et d'une mesure absolument continue.

Proposition 4.7.

Soit $I = \{1, \dots, n\}$ avec $n \in \mathbb{N}^*$ ou $I = \mathbb{N}$. Considérons X une variable aléatoire de loi

$$d\mathbb{P}_X = f_X d\lambda_d + \sum_{i \in I} p_i d\delta_{x_i} \quad (4.3)$$

avec

- (i) f_X une fonction borélienne positive,
- (ii) $(x_i)_{i \in I}$ une famille finie ou dénombrable de réels strictement croissante (i.e. telle que $x_i < x_{i+1}$)
- (iii) pour tout $i \in I$, $p_i \in [0, 1]$ (dans ce cas $p_i = \mathbb{P}(X = x_i)$).

Alors,

1. la fonction de répartition F_X de X est continue sauf peut-être en les points x_i , $i \in I$. Le saut de F_X en x_i est

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

2. de plus F est dérivable λ_1 -presque partout et $F'_X = f_X$ λ_1 -presque partout ;
3. si la fonction f_X est continue sur $]x_i, x_{i+1}[$, alors F_X est $\mathcal{C}^1$ sur $]x_i, x_{i+1}[$ et

$$\forall t \in]x_i, x_{i+1}[, F'_X(t) = f_X(t).$$

Notons que la proposition précédente généralise les propositions 4.3 et 4.5. Comme dans les paragraphes précédents, nous terminons en donnant des conditions suffisantes pour qu'une fonction soit la fonction de répartition d'une loi $\mathbb{P}_X$ donnée par (4.3).

Proposition 4.8.

Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante continue à droite, $\mathcal{C}^1$ par morceaux et telle que

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

Supposons qu'il existe $(x_i)_{i \in I}$, avec $I = \{1, \dots, n\}$ ou $I = \mathbb{N}^*$, une famille strictement croissante de réels telle que F est $\mathcal{C}^1$ sur chaque intervalle $]x_i, x_{i+1}[$. Alors, il existe une variable aléatoire X définie sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$ dont F est la fonction de répartition. De plus, la loi de X est

$$d\mathbb{P}_X = f_X d\lambda_d + \sum_{i \in I} p_i d\delta_{x_i}$$

avec

- (i) pour tout $i \in I$, $p_i = \mathbb{P}(X = x_i) = F_X(x_i) - F_X(x_{i-})$ le saut de la fonction F_X en x_i
- (ii) $f_X = F'_X$ (fonction bien définie sauf peut-être en x_i).

Exemple 4.6. Considérons la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$F(t) = \begin{cases} 0 & \text{si } t < 0 \\ \sqrt{t} & \text{si } t \in [0, 1/4[\\ 1 & \text{si } t \in [1/4, +\infty[. \end{cases}$$

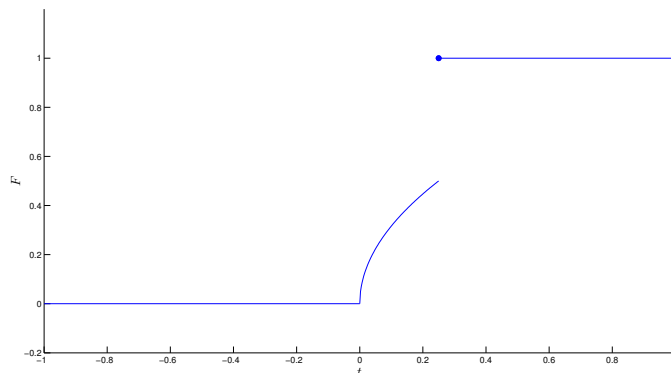


FIGURE 4.4 – Fonction F

La fonction F étant croissante continue à droite telle que $\lim_{x \rightarrow +\infty} F(x) = 1$ et $\lim_{x \rightarrow -\infty} F(x) = 0$, il s'agit de la fonction de répartition d'une variable aléatoire réelle X . En appliquant la proposition précédente,

$$d\mathbb{P}_X = f d\lambda_1 + \frac{1}{2} d\delta_{1/4}$$

avec $f : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par $f(t) = \frac{1}{2\sqrt{t}} \mathbf{1}_{]0, 1/4[}(t)$.

4.1.5 Généralisation aux vecteurs aléatoires

Nous terminons ce paragraphe en généralisant la notion de fonction de répartition aux variables aléatoires à valeurs dans $\mathbb{R}^d$.

Définition 4.2.

Pour tout entier $1 \leq i \leq d$, considérons $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire. Notons $\mathbb{P}_X$ la loi de $X = (X_1, \dots, X_d)$. La **fonction de répartition de X** est la fonction $F_X : \mathbb{R}^d \rightarrow [0, 1]$ définie par

$$\forall (t_1, \dots, t_d) \in \mathbb{R}^d, F_X(t_1, \dots, t_d) = \mathbb{P}\left(\bigcap_{i=1}^d \{X_i \leq t_i\}\right) = \mathbb{P}_X\left(\prod_{i=1}^d]-\infty, t_i]\right).$$

Remarque 4.3. La proposition 4.1 reste vraie pour des variables à valeurs dans $\mathbb{R}^d$.

4.2 Fonction caractéristique

Nous terminons ce chapitre en introduisant une nouvelle fonction caractérisant la loi d'une variable aléatoire.

Définition 4.3 (Fonction caractéristique).

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$. La **fonction caractéristique de X** est la fonction $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par

$$\forall t = (t_1, \dots, t_d) \in \mathbb{R}^d, \varphi_X(t) = \mathbb{E}\left(e^{i\langle t, X \rangle}\right) = \mathbb{E}\left(e^{i\sum_{k=1}^d t_k X_k}\right) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mathbb{P}_X(x)$$

où $\mathbb{P}_X$ désigne la loi de X .

Remarque 4.4.

1. La fonction caractéristique de la variable aléatoire X ne dépend que de la loi de X . Alors, si les variables aléatoires X et Y ont même loi, elles ont même fonction caractéristique. Nous verrons que la réciproque est vraie (cf. théorème 4.9).
2. (a) Si X admet comme densité la fonction f_X , alors

$$\varphi_X(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} f_X(x) \lambda_d(dx).$$

Dans ce cas, φ_X est la transformée de Fourier de la fonction f_X .

- (b) Si la loi de X est la loi discrète $\mathbb{P}_X = \sum_{n \in \mathbb{N}} a_n \delta_{x_n}$, alors

$$\varphi_X(t) = \sum_{n \in \mathbb{N}} a_n e^{i\langle t, x_n \rangle}.$$

En particulier, si $x_n = \frac{2\pi n}{T}$ pour tout n , φ_X est une série de Fourier, associée à une fonction T -périodique (dont les a_n sont les coefficients dit de Fourier).

3. Le tableau 4.1 donne quelques exemples de fonctions caractéristiques de lois classiques. Le lecteur est encouragé à essayer de calculer ses fonctions (sauf celle de la loi gaussienne).

Théorème 4.9 (Caractérisation de la loi d'une variable aléatoire).

Si deux variables aléatoires X et Y ont même fonction caractéristique alors X et Y ont même loi.

Loi de la v.a. X	Fonction caractéristique de X
Loi de Bernoulli $\mathcal{B}(p)$ avec $p \in [0, 1]$	$\varphi_X(t) = 1 - p + pe^{it}$
Loi binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$	$\varphi_X(t) = (pe^{it} + 1 - p)^n$
Loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	$\varphi_X(t) = (pe^{it} + 1 - p)^n$
Loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0, 1[$	$\varphi_X(t) = \frac{pe^{it}}{1 - (1 - p)e^{it}}$
Loi Uniforme $\mathcal{U}([0, 1])$	$\varphi_X(t) = \begin{cases} \frac{e^{it}-1}{it} & \text{si } t \neq 0 \\ 1 & \text{si } t = 0 \end{cases}$
Loi Gaussienne $\mathcal{N}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$	$\varphi_X(t) = \exp\left(itm - \frac{t^2\sigma^2}{2}\right)$

TABLE 4.1 – Fonction caractéristique de quelques lois classiques

Nous énonçons à présent quelques propriétés de la fonction caractéristique.

Proposition 4.10.

Si X une variable aléatoire réelle de fonction caractéristique φ_X , alors la fonction φ_X est uniformément continue sur $\mathbb{R}^d$, c'est-à-dire que

$$\forall \varepsilon > 0, \exists \delta > 0, \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d, \|x - y\| \leq \delta \implies |\varphi_X(x) - \varphi_X(y)| \leq \varepsilon.$$

A fortiori φ_X est continue sur $\mathbb{R}^d$. De plus, elle est bornée par 1.

Proposition 4.11 (Dérivation d'une fonction caractéristique).

Considérons une variable aléatoire réelle X et notons φ_X sa fonction caractéristique.

1. Soit $k \in \mathbb{N}$ un entier tel que $\mathbb{E}(|X|^k) < +\infty$. Alors φ_X est k fois dérivable et

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k).$$

De plus, au voisinage de 0,

$$\varphi_X(t) = \sum_{n=0}^k \frac{(it)^n}{n!} \mathbb{E}(X^n) + o(|t|^k).$$

En particulier, si X est intégrable, alors φ_X est dérivable et $\mathbb{E}(X) = -i\varphi_X'(0)$. De plus, si X est de carré intégrable, alors φ_X est deux fois dérivable et $\mathbb{E}(X^2) = -\varphi_X''(0)$.

2. Si φ_X est dérivable $(2p)$ fois, avec $p \in \mathbb{N}^*$, alors $\mathbb{E}(X^{2p}) < +\infty$ et

$$\mathbb{E}(X^k) = (-i)^k \varphi_X^{(k)}(0)$$

pour tout entier $0 \leq k \leq 2p$.

Remarque 4.5.

1. Lorsque $\mathbb{E}(|X|^k) < +\infty$, le réel $\mathbb{E}(X^k)$ est appelé moment d'ordre k de X . Nous reviendrons sur cette notion dans le chapitre 7.
2. La démonstration de l'assertion 1. de la proposition précédente repose sur un des théorèmes clefs de l'intégration : le théorème de convergence dominée que nous énoncerons dans le chapitre 6. Plus précisément, ce théorème permet en particulier d'étudier la régularité d'une intégrale paramétrée et d'échanger les signes dérivations et intégration.
3. Il est possible que φ_X soit $(2p+1)$ fois dérivable avec $p \in \mathbb{N}^*$ et que $\mathbb{E}(|X|^{2p+1}) = +\infty$.

Appliquons la proposition précédente pour la loi gaussienne centrée réduite $\mathcal{N}(0,1)$, c'est-à-dire la loi de densité $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

Corollaire 4.12.

Soit Z une variable aléatoire de loi gaussienne réduite et centrée. Alors pour tout $k \in \mathbb{N}$, $\mathbb{E}(|Z|^k) < +\infty$ et

$$\mathbb{E}(Z^k) = \begin{cases} \frac{(2p)!}{2^p p!} & \text{si } k = 2p \text{ avec } p \in \mathbb{N} \\ 0 & \text{si } k \text{ est impair} \end{cases}$$

Preuve du corollaire 4.12. Laissée en exercice. ■

Chapitre 5

Théorème de Fubini

Ce chapitre est dédiée au théorème de Fubini, théorème qui permet de ramener le calcul d'une intégrale « double » sur un espace produit $\Omega_1 \times \Omega_2$ aux calculs d'intégrales « simples ». Ce résultat a aussi été vu en classe préparatoire pour des fonctions continues sur $\mathbb{R}^p$. Comme pour les théorèmes permettant l'échange des signes limite/espérance, il est énoncé dans le cadre général de la théorie de la mesure car se restreindre à l'intégration sur des espaces $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ de probabilités n'est pas en pratique suffisant, même dans une étude probabiliste.

Nous donnons également des applications des théorèmes de Fubini : obtention des lois des coordonnées d'un vecteur et échange des signes somme, intégrale et espérance. Attention ! Il n'y aura pas de TD associé à ce chapitre. C'est pour vous aider à travailler en autonomie que de nombreux exemples vous sont fournis.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mu)$ est un **espace mesuré** ; comme évoqué dans le chapitre 1, il est supposé complet, c'est-à-dire que la tribu $\mathcal{F}$ contient tous les ensembles négligeables pour la mesure μ .
- Si μ est une probabilité sur Ω , on préférera la noter encore $\mathbb{P}$ et on rappelle la notation

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P}$$

pour l'espérance d'une variable aléatoire X , lorsque celle-ci est bien définie.

- Tout espace $E \subset \mathbb{R}^d$ est muni de sa tribu borélienne $\mathcal{B}(E)$ et tout espace E fini ou dénombrable est muni de sa tribu discrète $\mathcal{P}(E)$.

5.1 Tribu produit

Dans cette partie, nous considérons d espaces mesurables $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_d, \mathcal{F}_d)$, avec $d \in \mathbb{N}^*$, ainsi que l'espace produit $\Omega = \Omega_1 \times \dots \times \Omega_d$. Nous définissons la tribu produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_d$ des tribus $\mathcal{F}_1, \dots, \mathcal{F}_d$.

Définition 5.1 (Tribu produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_d$).

La tribu produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_d$ sur $\Omega_1 \times \dots \times \Omega_d$ est la tribu engendrée sur $\Omega_1 \times \dots \times \Omega_d$ par l'ensemble des rectangles de $\mathcal{F}_1 \times \dots \times \mathcal{F}_d$, c'est-à-dire par

$$\{A_1 \times A_2 \times \dots \times A_d / A_i \in \mathcal{F}_i, 1 \leq i \leq d\}.$$

Lorsque $\Omega_1 = \dots = \Omega_n$ et $\mathcal{F}_1 = \dots = \mathcal{F}_d = \mathcal{F}$, nous notons : $\mathcal{F}^{\otimes d} = \underbrace{\mathcal{F} \otimes \dots \otimes \mathcal{F}}_{d \text{ fois}}.$

Exemple 5.1.

1. Si $\mathcal{F}_1 = \mathcal{P}(\Omega_1)$ et si $\mathcal{F}_2 = \{\emptyset, \Omega_2\}$, alors $\mathcal{F}_1 \otimes \mathcal{F}_2 = \{A_1 \times \Omega_2 / A_1 \in \mathcal{F}_1\}$.
2. Soit $d \in \mathbb{N}^*$. Alors la tribu borélienne sur $\mathbb{R}^d$ est la tribu produit des tribus boréliennes $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$. Autrement dit, $\mathcal{B}(\mathbb{R})^{\otimes d} = \mathcal{B}(\mathbb{R}^d)$. De même $\mathcal{B}(\overline{\mathbb{R}}^d) = \mathcal{B}(\overline{\mathbb{R}})^{\otimes d}$.

Dans le théorème de Fubini, nous considérerons une fonction définie sur un espace produit, qui sera muni d'une tribu produit, à valeurs dans $\overline{\mathbb{R}}$. Mais avant d'énoncer ce résultat, nous nous permettons une digression portant sur la mesurabilité d'une fonction à valeurs dans un espace produit. Ainsi, la proposition suivante établit en particulier que $(X_1, \dots, X_d)$ est un vecteur aléatoire (i.e. une fonction mesurable) si et seulement si chacune de ses coordonnées X_i est une variable aléatoire.

Proposition 5.1 (Mesurabilité d'une fonction à valeurs dans $\Omega_1 \times \dots \times \Omega_d$).

Soient $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_d, \mathcal{F}_d)$ des espaces mesurables. Nous munissons l'espace $\Omega_1 \times \dots \times \Omega_d$ de la tribu produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_d$. Nous rappelons que l'espace Ω est lui muni de la tribu $\mathcal{F}$.

Considérons des applications

$$X^{(i)} : \Omega \rightarrow \Omega_i, \quad 1 \leq i \leq d$$

ainsi que l'application $X = (X^{(1)}, \dots, X^{(d)})$. Alors, l'application

$$X : (\Omega, \mathcal{F}) \rightarrow (\Omega_1 \times \dots \times \Omega_d, \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_d)$$

est mesurable si et seulement si pour tout $1 \leq i \leq d$, $X^{(i)} : (\Omega, \mathcal{F}) \rightarrow (\Omega_i, \mathcal{F}_i)$ est mesurable.

Exemple 5.2. Considérons U une variable aléatoire réelle et l'application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \times \mathbb{R} \\ x &\longmapsto (\cos(U), [U]) \end{aligned}$$

où $[y]$ désigne la partie entière du réel y . Comme $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, montrer que l'application X à valeurs dans $\mathbb{R}^2$ est un vecteur aléatoire (i.e. une fonction mesurable) revient à établir que chacune de ses coordonnées est une variable aléatoire réelle.

- La fonction $\cos$ étant continue sur $\mathbb{R}$ et U étant une variable aléatoire réelle, $\cos(U)$ est aussi une variable aléatoire.
- La fonction partie entière étant constante par morceaux sur $\mathbb{R}$, elle est borélienne. Dès lors, comme U est une variable aléatoire réelle, $[U]$ est aussi une variable aléatoire.

Par suite, d'après la proposition 5.1, l'application X est bien une variable aléatoire.

5.2 Mesure Produit

La notion de mesure produit $\mu_1 \otimes \mu_2$ sur un espace produit $\Omega_1 \times \Omega_2$ est bien définie lorsque l'on considère deux mesures μ_1 et μ_2 σ -finie. Donnons tout d'abord la définition de la notion de mesure σ -finie.

Définition 5.2 (Mesure σ -finie).

Une mesure μ sur un espace mesurable $(\Omega, \mathcal{F})$ est **σ -finie** s'il existe une suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{F}$ telle que pour tout $n \in \mathbb{N}$, $\mu(A_n) < +\infty$ et telle que

$$\Omega = \bigcup_{n \in \mathbb{N}} A_n.$$

Dans ce cadre de mesures σ -finies, la proposition suivante définit la mesure produit $\mu_1 \otimes \mu_2$ sur $\Omega_1 \times \Omega_2$. Elle se généralise aisément à un produit de d espaces avec $d \geq 2$.

Proposition 5.2 (Mesure produit).

Si μ_1 est une mesure σ -finie sur l'espace mesurable $(\Omega_1, \mathcal{F}_1)$ et si μ_2 est une mesure σ -finie sur l'espace mesurable $(\Omega_2, \mathcal{F}_2)$, alors il existe une unique mesure μ sur $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ telle que

$$\forall A_1 \in \mathcal{F}_1, \forall A_2 \in \mathcal{F}_2, \mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

avec pour convention $0 \times +\infty = +\infty \times 0 = 0$ et $+\infty \times +\infty = +\infty$. Cette mesure μ , que nous notons sous la forme

$$\mu = \mu_1 \otimes \mu_2,$$

*est appelée **mesure produit** sur $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ et est σ -finie.*

Remarque 5.1. Si μ_1 et μ_2 sont deux mesures finies, alors la mesure produit $\mu_1 \otimes \mu_2$ est bien définie et est finie. Si de plus μ_1 et μ_2 sont deux probabilités, alors $\mu_1 \otimes \mu_2$ est une probabilité.

Remarque 5.2. La proposition 5.2 se généralise à l'espace produit $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_d$.

Exemple 5.3.

1. Pour tout $a \in \Omega_1$ et tout $b \in \Omega_2$, $\delta_a \otimes \delta_b = \delta_{(a,b)}$.
2. La mesure de Lebesgue λ_1 étant σ -finie, $\lambda_d = \underbrace{\lambda_1 \otimes \cdots \otimes \lambda_1}_{d \text{ fois}} = \lambda_1^{\otimes d}$.

5.3 Théorèmes de Fubini

Sous de bonnes hypothèses sur l'espace produit, les théorèmes de Fubini établissent le résultat suivant : l'intégrale d'une fonction f par rapport à une mesure produit $\mu_1 \otimes \mu_2$, lorsqu'elle a un sens, se ramène « en général » au calcul d'intégrales par rapport à μ_1 et μ_2 .

Comme souvent, nous commençons par étudier le cas de fonctions positives. Dans le cas général, nous sommes souvent amenés tout d'abord à étudier l'intégrable du module de la fonction, étude que l'on peut mener en appliquant le théorème de Fubini pour les fonctions positives.

Théorème 5.3 (Théorème de Fubini-Tonelli).

Soient $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_2, \mu_2)$ deux espaces mesurés. Pour tout $i \in \{1, 2\}$, supposons que μ_i est une mesure σ -finie sur $(\Omega_i, \mathcal{F}_i)$. Soit

$$f : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$$

une fonction mesurable à valeurs dans $[0, +\infty]$. Nous considérons

$$F_1(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \quad \text{et} \quad F_2(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1).$$

Alors pour $i \in \{1, 2\}$, la fonction $F_i : (\Omega_i, \mathcal{F}_i) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$ est mesurable. De plus

$$\iint_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) (\mu_1 \otimes \mu_2)(d\omega_1, d\omega_2) = \int_{\Omega_1} F_1(\omega_1) \mu_1(d\omega_1) = \int_{\Omega_2} F_2(\omega_2) \mu_2(d\omega_2). \quad (5.1)$$

Remarque 5.3.

1. Autrement dit, sous les hypothèses du théorème précédent,

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

Le calcul de l'intégrale double sur $\Omega_1 \times \Omega_2$ se ramène donc au calcul de deux intégrales simples et il est possible d'intégrer dans l'ordre que l'on souhaite.

2. Le théorème de Fubini-Tonelli se généralise à un espace produit $\Omega_1 \times \cdots \times \Omega_d$.

Intéressons-nous au cas des fonctions a priori non positives.

Théorème 5.4 (Théorème de Fubini).

Soient $(\Omega_1, \mathcal{F}_1, \mu_1)$ et $(\Omega_2, \mathcal{F}_2, \mu_2)$ deux espaces mesurés complets avec μ_i , $i \in \{1, 2\}$, une mesure σ -finie sur $(\Omega_i, \mathcal{F}_i)$. Considérons $f : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une fonction $\mu_1 \otimes \mu_2$ -intégrable.

1. Alors la fonction $F_1 : \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2)$ est bien définie μ_1 -presque partout et est μ_1 -intégrable.
2. De même la fonction $F_2 : \omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1)$ est bien définie μ_2 -presque partout et est μ_2 -intégrable.
3. Enfin l'équation (5.1) est vérifiée, c'est-à-dire que

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

Remarque 5.4. Montrer qu'une fonction mesurable f est $\mu_1 \otimes \mu_2$ -intégrable revient à établir que

$$\iint_{\Omega_1 \times \Omega_2} |f| d(\mu_1 \otimes \mu_2) < +\infty.$$

Pour étudier la dernière intégrale, on peut en particulier lui appliquer le théorème de Fubini-Tonelli car $|f|$ est une fonction mesurable positive. On peut par exemple écrire :

$$\iint_{\Omega_1 \times \Omega_2} |f| d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} |f(\omega_1, \omega_2)| \mu_2(d\omega_2) \right) \mu_1(d\omega_1)$$

pour tenter de montrer qu'elle est finie.

Remarque 5.5. Pour $i \in \{1, 2\}$, considérons $(\Omega_i, \mathcal{F}_i, \mu_i)$ un espace mesuré complet avec μ_i une mesure σ -finie. Si $f_1 : (\Omega_1, \mathcal{F}_1) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$ et $f_2 : (\Omega_2, \mathcal{F}_2) \rightarrow ([0, +\infty], \mathcal{B}([0, +\infty]))$ sont des fonctions boréliennes,

$$\boxed{\int_{\Omega_1 \times \Omega_2} f_1(\omega_1) f_2(\omega_2) (\mu_1 \otimes \mu_2)(d\omega_1, d\omega_2) = \left(\int_{\Omega_1} f_1(\omega_1) \mu_1(d\omega_1) \right) \left(\int_{\Omega_2} f_2(\omega_2) \mu_2(d\omega_2) \right).}$$

La formule précédente reste vraie si f_1 est μ_1 -intégrable et si f_2 est μ_2 -intégrable.

5.4 Applications des théorèmes de Fubini

5.4.1 Lois marginales d'un vecteur aléatoire

Lorsque X est une variable aléatoire à valeurs dans un espace produit, nous pouvons nous intéresser à la loi de ses coordonnées.

Définition 5.3 (Lois marginales).

Soit $X = (X_1, \dots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$. La loi de la variable aléatoire X_i est appelée **loi marginale** de la $i^{\text{ème}}$ composante.

Nous supposons dans la suite pour simplifier $n = 2$. Mais il est facile de transposer les résultats au cas général. L'énoncé suivant est complètement général et est réécrit dans le tableau 5.1 dans le cas des lois discrètes, absolument continues ou mixtes.

Proposition 5.5.

Considérons deux mesures σ -finies μ_1 et μ_2 sur $\mathbb{R}$. Supposons que la loi du vecteur aléatoire (X_1, X_2) est absolument continue par rapport à la mesure produit $\mu = \mu_1 \otimes \mu_2$ de densité $f_{(X_1, X_2)}$. Autrement dit pour tout ensemble $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}_X(A) = \int_A f_{(X_1, X_2)}(x_1, x_2) \nu(dx_1, dx_2).$$

Alors la loi de la variable aléatoire X_1 (respectivement X_2) est absolument continue par rapport à la mesure μ_1 (respectivement μ_2) et admet pour densité la fonction f_{X_1} (respectivement f_{X_2}) définie par

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) \mu_2(dx_2) \quad \text{et} \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) \mu_1(dx_1)$$

Démonstration. La fonction $f_{(X_1, X_2)}$ étant mesurable positive, d'après le théorème de Fubini-Tonelli, l'application

$$f_{X_1} : x_1 \mapsto \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) \mu_2(dx_2)$$

est bien définie sur $\mathbb{R}$ et est mesurable positive. De plus, pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{X_1}(A) = \mathbb{P}(X_1 \in A) = \mathbb{P}((X_1, X_2) \in A \times \mathbb{R}) = \int_{\mathbb{R}} \mathbf{1}_A(x_1) f_{(X_1, X_2)}(x_1, x_2) (\mu_1 \otimes \mu_2)(dx_1, dx_2)$$

car la loi de (X_1, X_2) est absolument continue de densité $f_{(X_1, X_2)}$ par rapport à la mesure produit $\mu_1 \otimes \mu_2$. Alors, pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{X_1}(A) = \int_{\mathbb{R}} \mathbf{1}_A(x_1) \left(\int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) = \int_{\mathbb{R}} \mathbf{1}_A(x_1) f_{X_1}(x_1) \mu_1(dx_1)$$

d'après le théorème de de Fubini-Tonelli appliqué à la fonction mesurable positive $f = \mathbf{1}_A f_{(X_1, X_2)}$. L'égalité ci-dessus étant vraie pour tout $A \in \mathcal{B}(\mathbb{R})$, la loi de X_1 est absolument continue par rapport à la mesure μ_1 de densité f_{X_1} . De même on montre que la loi de X_2 est absolument continue par rapport à la mesure μ_2 de densité f_{X_2} définie dans l'énoncé. ■

Remarque 5.6.

1. La proposition précédente se généralise au cas d'un vecteur aléatoire à valeurs dans tout espace produit $\Omega_1 \times \Omega_2$, muni d'une mesure produit de mesures σ -finies.
2. Soit Y une variable aléatoire discrète à valeurs presque sûrement dans un espace E fini ou dénombrable. Alors, pour A mesurable,

$$\mathbb{P}_Y(A) = \mathbb{P}(Y \in A) = \sum_{y \in A} \mathbb{P}(Y = y) = \int_A \mathbb{P}(Y = y) \nu(dy)$$

où $\nu = \sum_{y \in E} \delta_y$ est la mesure de comptage sur l'ensemble E . Par suite, la loi de Y est absolument continue par rapport à la mesure ν de densité f_Y donnée par

$$f_Y(y) = \mathbb{P}(Y = y).$$

Il est alors possible de vérifier que le cas d'un vecteur aléatoire discret (X_1, X_2) à valeurs dans $E_1 \times E_2$ est couvert par les hypothèses de la proposition précédente en prenant

$$\mu_1 = \sum_{x \in E_1} \delta_x \quad \text{et} \quad \mu_2 = \sum_{x \in E_2} \delta_x$$

les mesures de comptage respectivement sur E_1 et E_2 (dans ce cas $\nu = \mu_1 \otimes \mu_2$ est la mesure de comptage sur $E = E_1 \times E_2$). Nous renvoyons au tableau 5.1 pour l'application de la proposition dans ce cadre.

3. Le cas d'un vecteur de loi absolument continue (sous-entendu par rapport à la mesure de Lebesgue) se retrouve en prenant $\mu_1 = \mu_2 = \lambda_1$.

Hypothèses	Loi des marginales sous ces hypothèses
La loi de (X_1, X_2) est discrète, donnée par $\mathbb{P}_{(X_1, X_2)} = \sum_{(i, j) \in I \times J} p_{ij} \delta_{(x_i, y_j)}$ avec I fini ou dénombrable, J fini ou dénombrable, $x_i \neq x_{i'}$ si $i \neq i'$ et $y_j \neq y_{j'}$ si $j \neq j'$	• La loi de X_1 est $\mathbb{P}_{X_1} = \sum_{i \in I} p_{i \cdot} \delta_{x_i} \quad \text{avec} \quad p_{i \cdot} = \sum_{j \in J} p_{ij}.$ • La loi de X_2 est $\mathbb{P}_{X_2} = \sum_{j \in J} p_{\cdot j} \delta_{y_j} \quad \text{avec} \quad p_{\cdot j} = \sum_{i \in I} p_{ij}.$
La loi de $X = (X_1, X_2)$ est absolument continue de densité $f_X : \mathbb{R}^2 \rightarrow \mathbb{R}_+$, i.e. $\forall A \in \mathcal{B}(\mathbb{R}^2), \mathbb{P}_X(A) = \int_A f_X(x) \lambda_2(dx).$	• La loi de X_1 est absolument continue de densité f_{X_1} définie par $\forall x_1 \in \mathbb{R}, f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2) \lambda_1(dx_2).$ • La loi de X_2 est absolument continue de densité f_{X_2} définie par $\forall x_2 \in \mathbb{R}, f_{X_2}(x_2) = \int_{\mathbb{R}} f_X(x_1, x_2) \lambda_1(dx_1).$

TABLE 5.1 – Lois des marginales d'un vecteur discret ou absolument continue

5.4.2 Échange des signes somme et somme

Appliquant les théorèmes de Fubini pour des mesures μ_1 et μ_2 de comptages permet d'en déduire des critères pour échanger deux signes somme. La proposition suivante peut aussi se déduire des théorèmes de convergence monotone et dominée.

Proposition 5.6.

Soit $(u_{m,n})_{m,n \in \mathbb{N}}$ une famille dénombrable de nombres réels. Supposons que l'une au moins des deux propriétés suivantes est vérifiée :

- i) les termes de la famille $(u_{m,n})_{m,n \in \mathbb{N}}$ sont tous positifs,*
- ii) la somme double $\sum_{m \in \mathbb{N}, n \in \mathbb{N}} |u_{m,n}|$ est finie.*

Alors la somme $\sum_{m \in \mathbb{N}, n \in \mathbb{N}} u_{m,n}$ est bien définie et

$$\sum_{m \in \mathbb{N}, n \in \mathbb{N}} u_{m,n} = \sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} u_{m,n} = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} u_{m,n}.$$

Démonstration. Appliquer le théorème de Fubini-Tonelli et le théorème de Fubini avec $\Omega_1 = \Omega_2 = \mathbb{N}$,

$$\mu_1 = \mu_2 = \sum_{n \in \mathbb{N}} \delta_n \quad \text{et} \quad f(m, n) = u_{m,n}. \quad \blacksquare$$

5.4.3 Échange des signes somme et intégrale

Appliquant les théorèmes de Fubini pour μ_1 une mesure de comptage et μ_2 une mesure de Lebesgue permet d'en déduire des critères pour échanger une somme et une intégrale. La proposition suivante peut aussi se déduire des théorèmes de convergence monotone et dominée.

Proposition 5.7.

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions boréliennes de $\mathbb{R}^d$ dans $\mathbb{R}$. Supposons que l'une au moins des trois propriétés suivantes est vérifiée :

- i) les fonctions f_n sont toutes à valeurs positives,*
- ii) la somme $\sum_{n \in \mathbb{N}} \int_{\mathbb{R}^d} |f_n(x)| \lambda_1(dx)$ est finie,*
- iii) l'intégrale $\int_{\mathbb{R}^d} \sum_{n \in \mathbb{N}} |f_n(x)| \lambda_1(dx)$ est finie.*

Alors les sommes et intégrales suivantes sont bien définies et on a

$$\sum_{n \in \mathbb{N}} \int_{\mathbb{R}^d} f_n(x) \lambda_1(dx) = \int_{\mathbb{R}^d} \sum_{n \in \mathbb{N}} f_n(x) \lambda_1(dx).$$

Remarque 5.7. Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions boréliennes de $\mathbb{R}^d$ dans $\mathbb{R}$. Les fonctions $x \mapsto |f_n(x)|$ sont toutes boréliennes à valeurs positives, en particulier elles vérifient le point *i*). Par conséquent, on a toujours

$$\sum_{n \in \mathbb{N}} \int_{\mathbb{R}^d} |f_n(x)| \lambda_1(dx) = \int_{\mathbb{R}^d} \sum_{n \in \mathbb{N}} |f_n(x)| \lambda_1(dx)$$

et les hypothèses *ii*) et *iii*) sont donc équivalentes. De plus, si *ii*) ou *iii*) sont vérifiées, alors $x \mapsto \sum_{n \in \mathbb{N}} f_n(x)$ est intégrable et $(\int_{\mathbb{R}^d} f_n(x) \lambda_1(dx))_{n \in \mathbb{N}}$ est absolument sommable.

Démonstration. Appliquer le théorème de Fubini-Tonelli et le théorème de Fubini avec $\Omega_1 = \mathbb{N}$, $\Omega_2 = \mathbb{R}$,

$$\mu_1 = \sum_{n \in \mathbb{N}} \delta_n, \quad \mu_2 = \lambda_1 \quad \text{et} \quad f(n, x) = f_n(x). \quad \blacksquare$$

Exemple 5.4. Considérons la fonction $f : x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \mapsto \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2}$. Nous souhaitons montrer que f est intégrable par rapport à λ_1 et que son intégrale vaut $\sum_{n=1}^{\infty} \frac{2(-1)^n}{(2n+1)^3}$. Pour cela nous utilisons le théorème de Fubini à deux reprises : dans un premier temps, nous appliquons le point *i*) de la proposition 5.7 pour montrer que l'intégrale $\int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \sum_{n=1}^{\infty} \left| \frac{\cos(nx)}{n^2} \right| \lambda_1(dx)$ est finie ; ceci nous autorisera, dans un deuxième temps, à appliquer le point *iii*) de la proposition 5.7 à l'intégrale $\int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2} \lambda_1(dx)$ pour conclure le calcul de l'intégrale de f par rapport à λ_1 .

Pour tout $n \geq 1$, $u_n : x \in [-\frac{\pi}{2}, \frac{\pi}{2}] \mapsto \left| \frac{\cos(nx)}{n^2} \right|$ est borélienne (car continue) et positive. D'après le théorème de Fubini-Tonelli (et plus précisément d'après le point *i*) de la proposition 1.10),

$$\begin{aligned} \int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \sum_{n=1}^{\infty} \left| \frac{\cos(nx)}{n^2} \right| \lambda_1(dx) &= \sum_{n=1}^{\infty} \int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \left| \frac{\cos(nx)}{n^2} \right| \lambda_1(dx) \\ &\leq \sum_{n=1}^{\infty} \frac{\pi}{n^2} < \infty. \end{aligned}$$

Donc, en appliquant à nouveau le théorème de Fubini-Tonelli (et plus précisément d'après le point *iii*) de la proposition 1.10), l'intégrale de f est bien définie (f est même intégrable) et

$$\begin{aligned} \int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} f(x) \lambda_1(dx) &= \int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^2} \lambda_1(dx) = \sum_{n=1}^{\infty} \int_{[-\frac{\pi}{2}, \frac{\pi}{2}]} \frac{\cos(nx)}{n^2} \lambda_1(dx) \\ &= \sum_{n=1}^{\infty} \left[\frac{\sin(nx)}{n^3} \right]_{-\pi/2}^{\pi/2} = \sum_{n=1}^{\infty} \frac{2(-1)^n}{(2n+1)^3}. \end{aligned}$$

5.4.4 Échange des signes espérance et somme

Proposition 5.8.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs réelles. Supposons que l'une au moins des trois propriétés suivantes est vérifiée :

- i)* les variables aléatoires X_n sont toutes positives presque sûrement,
- ii)* la somme $\sum_{n \in \mathbb{N}} \mathbb{E}(|X_n|)$ est finie,
- iii)* l'espérance $\mathbb{E}(\sum_{n \in \mathbb{N}} |X_n|)$ est finie.

Alors les sommes et espérances suivantes sont bien définies et on a

$$\sum_{n \in \mathbb{N}} \mathbb{E}(X_n) = \mathbb{E} \left(\sum_{n \in \mathbb{N}} X_n \right).$$

Remarque 5.8. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs réelles. Les variables aléatoires $|X_n|$ sont toutes positives, donc elles vérifient le point *i*). Par conséquent, on a toujours

$$\sum_{n \in \mathbb{N}} \mathbb{E}(|X_n|) = \mathbb{E} \left(\sum_{n \in \mathbb{N}} |X_n| \right).$$

En particulier les hypothèses *ii*) et *iii*) sont donc équivalentes. De plus, si *ii*) ou *iii*) sont vérifiées, alors $x \mapsto \sum_{n \in \mathbb{N}} X_n$ est intégrable et $(\mathbb{E}(X_n))_{n \in \mathbb{N}}$ est une famille absolument sommable.

Démonstration. Appliquer le théorème de Fubini-Tonelli et le théorème de Fubini avec $\Omega_1 = \mathbb{N}$, $\Omega_2 = \Omega$,

$$\mu_1 = \sum_{n \in \mathbb{N}} \delta_n, \quad \mu_2 = \mathbb{P} \quad \text{et} \quad f(n, \omega) = X_n(\omega). \quad \blacksquare$$

Exemple 5.5. Soit U une variable aléatoire à valeurs dans $]0, 1[$ presque sûrement. Nous souhaitons montrer que

$$\mathbb{E} \left(\frac{1}{1-U} \right) = \sum_{n \in \mathbb{N}} \mathbb{E}(U^n).$$

Remarquons que $\frac{1}{1-U}$ est positif presque sûrement, donc l'espérance est toujours bien définie (éventuellement égale à l'infini). De plus, le développement en série entière de $\frac{1}{1-u}$ nous donne l'égalité

$$\frac{1}{1-U} = \sum_{n \in \mathbb{N}} U^n, \text{ presque sûrement.}$$

Or les variables aléatoires U^n sont toutes positives presque sûrement, donc, d'après le théorème de Fubini-Tonelli (et plus précisément d'après le point *i*) de la proposition 5.8), nous avons

$$\mathbb{E} \left(\frac{1}{1-U} \right) = \mathbb{E} \left(\sum_{n \in \mathbb{N}} U^n \right) = \sum_{n \in \mathbb{N}} \mathbb{E}(U^n).$$

5.4.5 Échange des signes espérance et intégrale

Proposition 5.9.

Soit X une variable aléatoire à valeurs dans un espace mesurable E et $f : \mathbb{R} \times E \mapsto \mathbb{R}$ une application borélienne à valeurs dans $\mathbb{R}$. Supposons que l'une au moins des trois propriétés suivantes est vérifiée :

- (i) la fonction f est à valeurs positives,
- (ii) l'intégrale $\int_{\mathbb{R}} \mathbb{E}(|f(x, X)|) \lambda_1(dx)$ est finie,
- (iii) l'espérance $\mathbb{E} \left(\int_{\mathbb{R}} |f(x, X)| \lambda_1(dx) \right)$ est finie.

Alors les intégrales et espérances suivantes sont bien définies et on a

$$\int_{\mathbb{R}} \mathbb{E}(f(x, X)) \lambda_1(dx) = \mathbb{E} \left(\int_{\mathbb{R}} f(x, X) \lambda_1(dx) \right).$$

Remarque 5.9. Soit X une variable aléatoire à valeurs dans un espace mesurable E et $f : \mathbb{R} \times E \mapsto \mathbb{R}$ une application borélienne à valeurs dans $\mathbb{R}$. La fonction $|f|$ est positive, donc elle vérifie le point (i) de la proposition précédente. Par conséquent, on a toujours

$$\int_{\mathbb{R}} \mathbb{E}(|f(x, X)|) \lambda_1(dx) = \mathbb{E} \left(\int_{\mathbb{R}} |f(x, X)| \lambda_1(dx) \right).$$

En particulier les hypothèses (ii) et (iii) sont donc équivalentes. De plus, si (ii) ou (iii) sont vérifiées, alors la fonction $x \mapsto \mathbb{E}(f(x, X))$ est Lebesgue-intégrable et la variable aléatoire

$$Y = \int_{\mathbb{R}^d} f(x, X) \lambda_1(dx)$$

est absolument sommable.

Exemple 5.6. Soit X une variable aléatoire à valeurs réelles et intégrable. Nous cherchons à démontrer que l'intégrale $\int_{[0,1]} \mathbb{E}(X \cos(xX)) \lambda_1(dx)$ est bien définie et que l'on a l'égalité

$$\int_{[0,1]} \mathbb{E}(X \cos(xX)) \lambda_1(dx) = \mathbb{E}(\sin X).$$

Montrons dans un premier temps que l'intégrale est bien définie. On a

$$\int_{[0,1]} \mathbb{E}(|X \cos(xX)|) \lambda_1(dx) \leq \int_{[0,1]} \mathbb{E}(|X|) \lambda_1(dx) = \mathbb{E}(|X|) < \infty,$$

car X est supposée intégrable. L'assertion (ii) de la proposition 5.9 étant donc vérifiée, l'intégrale étudiée est bien définie et finie. De plus,

$$\int_{[0,1]} \mathbb{E}(X \cos(xX)) \lambda_1(dx) = \mathbb{E} \left(\int_{[0,1]} X \cos(xX) \lambda_1(dx) \right) = \mathbb{E}(\sin(X)).$$

Chapitre 6

Indépendance de variables aléatoires

Ce chapitre introduit une notion clef en probabilités : la notion d'indépendance de variables aléatoires. Cette notion permet de modéliser des expériences dont les résultats ne dépendent pas les uns des autres (dans un sens intuitif). Par exemple, lorsque l'on jette un dé bleu et un dé rouge, le résultat obtenu avec le dé bleu ne dépend pas de celui obtenu avec le dé rouge.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités (qui ne sera, sauf exception, pas explicité) sur lequel sont définies les variables aléatoires considérées dans ce chapitre.
- Les ensembles E_i , $1 \leq i \leq d$, sont munis de la tribu $\mathcal{E}_i$ et l'ensemble produit $E_1 \times \cdots \times E_d$ de la tribu produit $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_d$.
- Tout espace $E \subset \mathbb{R}^d$ est muni de sa tribu borélienne $\mathcal{B}(E)$ et tout espace E fini ou dénombrable est muni de sa tribu discrète $\mathcal{P}(E)$.
- $X = (X_1, \dots, X_d)$ est un vecteur aléatoire défini sur Ω i.e. chacune de ses coordonnées X_i est une variable aléatoire définie sur Ω .

6.1 Indépendance de variables aléatoires

Définition 6.1 (Indépendance mutuelle).

Soient $d \in \mathbb{N}^*$ et, pour tout $1 \leq i \leq d$, supposons que X_i est à valeurs dans E_i . Les variables $X_1, \dots, X_d$ sont dites **mutuellement indépendantes**¹ si

$$\mathbb{P}(X_1 \in B_1, \dots, X_d \in B_d) = \prod_{i=1}^d \mathbb{P}(X_i \in B_i)$$

pour tous $B_i \in \mathcal{E}_i, \dots, B_d \in \mathcal{E}_d$, c'est-à-dire si

$$\mathbb{P}_{(X_1, \dots, X_d)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_d}$$

avec $\mathbb{P}_Z$ la loi de Z .

Remarque 6.1. Les variables aléatoires $X_1, \dots, X_d$ seront dites **deux à deux indépendantes** si pour tout $i \neq j$, les variables aléatoires X_i et X_j sont indépendantes, c'est-à-dire si pour tout $i \neq j$ et tous $B_i \in \mathcal{E}_i$ et

1. On omet le plus souvent « mutuellement ».

$B_j \in \mathcal{E}_j$,

$$\mathbb{P}(X_i \in B_i, X_j \in B_j) = \mathbb{P}(X_i \in B_i) \mathbb{P}(X_j \in B_j).$$

Attention ! Comme dans le cas des événements, l'indépendance deux à deux de variables aléatoires n'implique pas leur indépendance mutuelle. En revanche, l'indépendance mutuelle implique l'indépendance deux à deux. En effet, il suffit de remarquer que

$$\mathbb{P}(X_i \in B_i, X_j \in B_j) = \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n)$$

avec pour $k \notin \{i, j\}$, $B_k = E_k$. Ainsi, par indépendance mutuelle,

$$\mathbb{P}(X_i \in B_i, X_j \in B_j) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n) = \mathbb{P}(X_i \in B_i) \mathbb{P}(X_j \in B_j),$$

car, pour tout $k \notin \{i, j\}$, $\mathbb{P}(X_k \in B_k) = \mathbb{P}(X_k \in E_k) = 1$.

Voici quelques exemples simples de variables aléatoires indépendantes.

Exemple 6.1.

1. Soit $a \in \mathbb{R}^d$. Alors, la variable constante $X = a$ (c'est-à-dire que X est de loi δ_a) est indépendante de toute variable aléatoire Y . En effet, pour tous ensembles mesurables B_1, B_2 ,

$$\begin{aligned} \mathbb{P}(X \in B_1, Y \in B_2) &= \begin{cases} \mathbb{P}(\emptyset) & \text{si } a \notin B_1 \\ \mathbb{P}(Y \in B_2) & \text{si } a \in B_1 \end{cases} \\ &= \delta_a(B_1) \mathbb{P}(Y \in B_2) = \mathbb{P}(X \in B_1) \mathbb{P}(Y \in B_2). \end{aligned}$$

2. Si $A_1, \dots, A_n$ sont des événements de Ω , alors les variables aléatoires $X_1 = \mathbf{1}_{A_1}, \dots, X_n = \mathbf{1}_{A_n}$ sont mutuellement indépendantes si et seulement si les événements $A_1, \dots, A_n$ sont indépendants.
3. Soit (X, Y) un couple de variables aléatoires de loi absolument continue de densité

$$f_{(X,Y)}(x, y) = \mathbf{1}_{[0,1] \times [0,1]}(x, y) = \mathbf{1}_{[0,1]}(x) \mathbf{1}_{[0,1]}(y).$$

Alors, pour tous $A, B \in \mathcal{B}([0, 1])$, d'après le théorème du transport puis le théorème de Fubini (les fonctions intégrées sont mesurables positives),

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{E}(\mathbf{1}_A(X) \mathbf{1}_B(Y)) = \int_{[0,1] \times [0,1]} \mathbf{1}_A(x) \mathbf{1}_B(y) f_{(X,Y)}(x, y) d\lambda_2(x, y) \\ &= \int_{[0,1]} \int_{[0,1]} \mathbf{1}_A(x) \mathbf{1}_B(y) \mathbf{1}_{[0,1]}(x) \mathbf{1}_{[0,1]}(y) \lambda_1(dx) \lambda_1(dy) \\ &= \int_{[0,1]} \mathbf{1}_A(x) \mathbf{1}_{[0,1]}(x) \lambda_1(dx) \times \int_{[0,1]} \mathbf{1}_B(y) \mathbf{1}_{[0,1]}(y) \lambda_1(dy) \\ &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B), \end{aligned}$$

car X et Y sont tous deux de loi uniforme sur $[0, 1]$ (c'est une conséquence immédiate du calcul ci-dessus en prenant tour à tour $A = [0, 1]$ puis $B = [0, 1]$). En conséquence, X et Y sont indépendantes. Nous verrons un peu plus loin un résultat qui généralise cet exemple.

La définition permet aisément d'établir la première propriété suivante, établissant que les images de variables aléatoires indépendantes sont indépendantes.

Proposition 6.1 (Indépendance et images de variables aléatoires).

Pour tout $1 \leq i \leq d$, considérons X_i une variable aléatoire à valeurs dans E_i (muni de la tribu $\mathcal{E}_i$) et un espace E'_i muni de la tribu $\mathcal{E}'_i$. Si les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes et si pour tout $1 \leq i \leq d$, la fonction $f_i : E_i \rightarrow E'_i$ est mesurable, alors les variables aléatoires $f_1(X_1), \dots, f_d(X_d)$ sont mutuellement indépendantes.

Démonstration. Soient $B_1 \in \mathcal{E}'_1, \dots$, et $B_d \in \mathcal{E}'_d$. Alors,

$$\mathbb{P}(f_1(X_1) \in B_1, \dots, f_d(X_d) \in B_d) = \mathbb{P}(X_1 \in f_1^{-1}(B_1), \dots, X_d \in f_d^{-1}(B_d)).$$

Étant donné que f_i est mesurable et que $B_i \in \mathcal{E}'_i$, $f_i^{-1}(B_i) \in \mathcal{E}_i$. Alors, par indépendance mutuelle des variables aléatoires $X_1, \dots, X_d$,

$$\mathbb{P}(f_1(X_1) \in B_1, \dots, f_d(X_d) \in B_d) = \prod_{i=1}^d \mathbb{P}(X_i \in f_i^{-1}(B_i)) = \prod_{i=1}^d \mathbb{P}(f_i(X_i) \in B_i).$$

Par conséquent, les variables aléatoires $f_1(X_1), \dots, f_d(X_d)$ sont mutuellement indépendantes. ■

Pour terminer cette partie, nous étendons la définition d'indépendance aux familles infinies de variables.

Définition 6.2 (Indépendance d'une famille quelconque).

Soit I un ensemble quelconque. Les variables aléatoires $(X_i)_{i \in I}$ sont alors dites **mutuellement indépendantes** (ou simplement **indépendantes**) si pour tout $n \in \mathbb{N}^*$ et pour tout $(i_1, \dots, i_n) \in I^n$, les variables aléatoires $X_{i_1}, \dots, X_{i_n}$ sont mutuellement indépendantes.

6.2 Caractérisations et conséquences

6.2.1 Indépendance et calcul d'espérances

La propriété d'indépendance des variables aléatoires $X_1, \dots, X_d$ permet de simplifier le calcul de l'espérance d'une variable aléatoire

$$Y = \prod_{i=1}^d h_i(X_i)$$

lorsque son espérance est bien définie. De plus, il existe une forme de réciproque. La proposition ci-après précise ces deux points.

Proposition 6.2.

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire avec X_i à valeurs dans E_i .

1. Supposons que les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes.

(a) Alors, si pour tout $1 \leq i \leq d$, la fonction $h_i : E_i \rightarrow [0, +\infty]$ est mesurable positive,

$$\mathbb{E}\left(\prod_{i=1}^d h_i(X_i)\right) = \prod_{i=1}^d \mathbb{E}(h_i(X_i)). \quad (6.1)$$

(b) Si pour tout $1 \leq i \leq d$, la fonction $h_i : E_i \rightarrow \overline{\mathbb{R}}$ est mesurable telle que $h_i(X_i)$ est intégrable alors, $\prod_{i=1}^d h_i(X_i)$ est intégrable et l'égalité (6.1) est vérifiée.

En particulier, si les variables aléatoires $X_1, \dots, X_d$ sont intégrables, alors, $\prod_{i=1}^d X_i$ est intégrable et

$$\mathbb{E}\left(\prod_{i=1}^d X_i\right) = \prod_{i=1}^d \mathbb{E}(X_i).$$

2. Si l'égalité (6.1) est vraie pour toutes fonctions $h_i : E_i \rightarrow [0, +\infty]$ mesurables positives, alors les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes.

Remarque 6.2.

1. Si les variables $X_1, \dots, X_d$ sont mutuellement indépendantes, l'égalité (6.1) a lieu dès que toutes les espérances écrites sont bien définies.
2. L'assertion 2. reste vraie en remplaçant « pour toutes fonctions $h_i : E_i \rightarrow \overline{\mathbb{R}}$ mesurables positives » par « pour toutes fonctions $h_i : E_i \rightarrow \overline{\mathbb{R}}$ mesurables bornées » ou encore par « pour toutes fonctions $h_i : E_i \rightarrow \overline{\mathbb{R}}$ mesurables positives bornées ».
3. Dans le cas où $E_i = \mathbb{R}^{p_i}$, l'assertion 2. reste encore vraie en remplaçant « pour toutes fonctions $h_i : E_i \rightarrow \overline{\mathbb{R}}$ mesurables positives » par « pour toutes fonctions $h_i : \mathbb{R}^{d_i} \rightarrow \mathbb{R}$ continues positives bornées ».

Démonstration.

1. Supposons $X_1, \dots, X_d$ mutuellement indépendantes.

(a) Pour tout $1 \leq i \leq d$, soit $h_i : E_i \rightarrow [0, +\infty]$ une fonction mesurable positive. Considérons la fonction $h : E_1 \times \dots \times E_d \rightarrow [0, +\infty]$ définie par

$$h(x_1, \dots, x_d) = \prod_{i=1}^d h_i(x_i), \quad \forall (x_1, \dots, x_d) \in E_1 \times \dots \times E_d.$$

La fonction h est alors mesurable positive et d'après le théorème du transport,

$$\mathbb{E}\left(\prod_{i=1}^d h_i(X_i)\right) = \mathbb{E}(h(X_1, \dots, X_d)) = \int_{\prod_{i=1}^d E_i} h(x_1, \dots, x_d) d\mathbb{P}_X(x_1, \dots, x_d).$$

où $\mathbb{P}_X$ est la loi du vecteur $X = (X_1, \dots, X_d)$. Les variables aléatoires $X_1, \dots, X_d$ étant mutuellement indépendantes,

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$$

où $\mathbb{P}_{X_i}$ est la loi de X_i . Dès lors, d'après le théorème de Fubini-Tonelli,

$$\begin{aligned} \mathbb{E}\left(\prod_{i=1}^d h_i(X_i)\right) &= \int_{E_d} \left(\dots \int_{E_2} \left(\int_{E_1} \prod_{i=1}^d h_i(x_i) d\mathbb{P}_{X_1}(x_1) \right) d\mathbb{P}_{X_2}(x_2) \dots \right) d\mathbb{P}_{X_d}(x_d) \\ &= \prod_{i=1}^d \int h_i(x_i) d\mathbb{P}_{X_i}(x_i) \\ &= \prod_{i=1}^d \mathbb{E}(h_i(X_i)). \end{aligned}$$

- (b) Pour tout $1 \leq i \leq d$, soit $h_i : E_i \rightarrow \overline{\mathbb{R}}$ une fonction mesurable telle que $h_i(X_i)$ est intégrable. Alors d'après l'assertion 1.(a) appliquée aux fonctions mesurables positives $|h_i|$,

$$\mathbb{E} \left(\left| \prod_{i=1}^d h_i(X_i) \right| \right) = \mathbb{E} \left(\prod_{i=1}^d |h_i(X_i)| \right) = \prod_{i=1}^d \mathbb{E}(|h_i(X_i)|) < +\infty$$

car chaque $h_i(X_i)$ étant intégrable, $\mathbb{E}(|h_i(X_i)|) \in \mathbb{R}$ pour tout $1 \leq i \leq d$. Par conséquent, $\prod_{i=1}^d h_i(X_i)$ est une variable aléatoire intégrable. En reprenant la preuve de l'assertion 1.(a) et en appliquant le théorème de Fubini (au lieu du théorème de Fubini-Tonelli), nous obtenons :

$$\mathbb{E} \left(\prod_{i=1}^d h_i(X_i) \right) = \mathbb{E} \left(\prod_{i=1}^n h_i(X_i) \right).$$

2. Supposons que pour toutes fonctions $h_i : E_i \rightarrow [0, +\infty]$, $1 \leq i \leq d$, mesurables positives, l'égalité (6.1) est vérifiée. Pour tout $1 \leq i \leq d$, considérons $A_i \in \mathcal{E}_i$. Les fonctions

$$h_i = \mathbf{1}_{A_i}, \quad 1 \leq i \leq d$$

sont alors mesurables positives. De plus, par définition,

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{E} \left(\prod_{i=1}^d \mathbf{1}_{A_i}(X_i) \right).$$

Alors, d'après (6.1),

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{i=1}^d \mathbb{E}(\mathbf{1}_{A_i}(X_i)) = \prod_{i=1}^d \mathbb{P}(X_i \in A_i).$$

L'égalité précédente étant vraie pour tout $A_i \in \mathcal{E}_i$, $1 \leq i \leq d$, les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes. ■

6.2.2 Indépendance et fonctions de répartition

Dans cette section, nous nous intéressons aux cas de variables aléatoires X_i , $1 \leq i \leq d$ à valeurs dans $\mathbb{R}$. L'indépendance mutuelle de ces variables se lit sur la fonction de répartition du vecteur $(X_1, \dots, X_d)$.

Proposition 6.3 (Indépendance de variables aléatoires réelles).

1. Si les variables aléatoires réelles $X_1, \dots, X_d$ sont mutuellement indépendantes alors

$$\forall (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) = \prod_{i=1}^d \mathbb{P}(X_i \leq t_i) \stackrel{\text{déf}}{=} \prod_{i=1}^d F_{X_i}(t_i)$$

où F_{X_i} est la fonction de répartition de X_i .

2. Réciproquement, si il existe des fonctions mesurables $h_i : \mathbb{R} \rightarrow \mathbb{R}$, $i \in \{1, \dots, d\}$, telles que

$$\forall (t_1, \dots, t_d) \in \mathbb{R}^d, \quad \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) = \prod_{i=1}^d h_i(t_i),$$

alors les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes.

Démonstration.

1. Supposons que les variables aléatoires X_i , $i \in \{1, \dots, d\}$, sont mutuellement indépendantes. Alors, par mesurabilité des intervalles $] - \infty, t_i]$ et par définition de l'indépendance mutuelle,

$$\mathbb{P}(X_1 \in] - \infty, t_1], \dots, X_d \in] - \infty, t_d]) = \prod_{i=1}^d \mathbb{P}(X_i \in] - \infty, t_i]) = \prod_{i=1}^d \mathbb{P}(X_i \leq t_i),$$

pour tous $t_1, \dots, t_d \in \mathbb{R}$. Nous avons donc établi l'assertion 1.

2. Nous ne donnons qu'une esquisse de la preuve de la réciproque. On peut commencer par montrer que, pour tout $i \in \{1, \dots, d\}$ et tout $t \in \mathbb{R}$, on a

$$F_{X_i}(t) = \frac{h_i(t)}{\lim_{u \rightarrow +\infty} h_i(u)} \quad \text{et} \quad \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) = \prod_{i=1}^d F_{X_i}(t_i).$$

Puis on remarque qu'un vecteur $(Y_1, \dots, Y_d)$ de loi $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$ a pour fonction de répartition

$$\mathbb{P}(Y_1 \leq t_1, \dots, Y_d \leq t_d) = \prod_{i=1}^d F_{X_i}(t_i)$$

Par conséquent, $(Y_1, \dots, Y_d)$ et $(X_1, \dots, X_d)$ ont même loi (cf. remarque 4.3 du chapitre 4). Or les coordonnées de $(Y_1, \dots, Y_d)$ sont mutuellement indépendantes, donc les coordonnées de $(X_1, \dots, X_d)$ sont mutuellement indépendantes. ■

6.2.3 Indépendance et fonctions caractéristiques

Nous donnons à présent une caractérisation de l'indépendance via la notion de fonction caractéristique.

Proposition 6.4.

Pour tout $i \in \{1, \dots, d\}$, X_i est une variable aléatoire à valeurs dans $\mathbb{R}^{p_i}$ dont φ_{X_i} est la fonction caractéristique. Notons $\varphi_{(X_1, \dots, X_d)}$ la fonction caractéristique de $(X_1, \dots, X_d)$.

1. Si les variables aléatoires $X_1, \dots, X_d$ sont mutuellement indépendantes, alors

$$\forall (u_1, \dots, u_d) \in \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_d}, \quad \varphi_{(X_1, \dots, X_d)}(u_1, \dots, u_d) = \prod_{k=1}^d \varphi_{X_k}(u_k). \quad (6.2)$$

2. Réciproquement, si il existe des fonctions mesurables $h_i : \mathbb{R}^{p_i} \rightarrow \mathbb{C}$, $i \in \{1, \dots, d\}$, telles que

$$\forall (u_1, \dots, u_d) \in \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_d}, \quad \varphi_{(X_1, \dots, X_d)}(u_1, \dots, u_d) = \prod_{i=1}^d h_i(u_i), \quad (6.3)$$

alors les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes.

Démonstration. Dans la suite, comme dans tout le polycopié, $\mathbb{P}_Z$ (respectivement φ_Z) désigne la loi (respectivement la fonction caractéristique) de la variable aléatoire Z .

- Le terme de gauche dans l'égalité (6.3) étant non nul en $(u_1, \dots, u_d) = 0$, on en déduit que $h_i(0) \neq 0$ pour tout $i \in \{1, \dots, d\}$. De plus, on a, pour tout $u_i \in \mathbb{R}^{p_i}$,

$$\varphi_{X_i}(u_i) = \frac{h_i(u_i)}{h_i(0)} \text{ et } \prod_{i=1}^d h_i(0) = 1.$$

Par conséquent, l'égalité (6.3) est équivalente à l'égalité (6.2).

- Calculons à présent la fonction caractéristique d'un vecteur Y de loi

$$\mathbb{P}_Y = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$$

D'après le théorème du transport, pour tout $u = (u_1, \dots, u_d) \in \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_d}$,

$$\varphi_Y(u_1, \dots, u_d) = \int_{\mathbb{R}^{\sum_{j=1}^d p_j}} \prod_{j=1}^d e^{i\langle u_j, y_j \rangle} \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}(dy_1, \dots, dy_d).$$

Rappelons que la fonction intégrée est borélienne bornée par 1, et donc intégrable. En appliquant le théorème de Fubini, nous avons alors :

$$\varphi_Y(u_1, \dots, u_d) = \prod_{i=1}^d \int_{\mathbb{R}^{p_j}} e^{i\langle u_j, y_j \rangle} \mathbb{P}_{X_j}(dy_j) = \prod_{j=1}^d \varphi_{X_j}(u_j)$$

pour tout $u = (u_1, \dots, u_d) \in \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_d}$.

- Par définition, les variables aléatoires $X_1, \dots, X_d$ sont indépendantes si et seulement si

$$\mathbb{P}_{(X_1, \dots, X_d)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d},$$

c'est-à-dire si et seulement si $\mathbb{P}_{(X_1, \dots, X_d)} = \mathbb{P}_Y$ avec Y introduit précédemment. Par suite, la fonction caractéristique caractérisant la loi, les variables aléatoires $X_1, \dots, X_d$ sont indépendantes si et seulement si

$$\varphi_{(X_1, \dots, X_d)} = \varphi_Y$$

c'est-à-dire vu le point précédent, si et seulement si

$$\varphi_{(X_1, \dots, X_d)}(u_1, \dots, u_d) = \prod_{i=1}^d \int_{\mathbb{R}^{p_j}} e^{i\langle u_j, y_j \rangle} \mathbb{P}_{X_j}(dy_j) = \prod_{j=1}^d \varphi_{X_j}(u_j)$$

pour tout $u = (u_1, \dots, u_d) \in \mathbb{R}^{p_1} \times \dots \times \mathbb{R}^{p_d}$. ■

6.2.4 Cadre des lois discrètes et des lois absolument continues

Nous examinons le cas où toutes les variables sont discrètes.

Proposition 6.5 (Indépendance et lois discrètes).

1. Soit (X, Y) un vecteur aléatoire discret à valeurs dans $E_1 \times E_2$. Alors les variables aléatoires X et Y sont mutuellement indépendantes si et seulement si

$$\forall i \in E_1, \forall j \in E_2, \mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i) \mathbb{P}(Y = j).$$

2. Plus généralement, soient $X_1, \dots, X_n$ des variables aléatoires discrètes. Supposons que X_k est à valeurs dans E_k fini ou dénombrable. Alors les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes si et seulement si, pour tout $(i_1, \dots, i_n) \in E_1 \times \dots \times E_n$,

$$\mathbb{P}(X_1 = i_1, \dots, X_n = i_n) = \prod_{k=1}^n \mathbb{P}(X_k = i_k).$$

Nous nous intéressons à présent au cas d'un vecteur aléatoire dont la loi est absolument continue.

Proposition 6.6 (Indépendance des marginales d'un vecteur de loi absolument continue).

Pour tout $1 \leq i \leq d$, considérons X_i une variable aléatoire à valeurs dans $\mathbb{R}^{p_i}$. Supposons que le vecteur aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans $\mathbb{R}^p$, avec $p = p_1 + \dots + p_n$, est de loi absolument continue de densité $f_X : \mathbb{R}^p \rightarrow [0, +\infty]$. Alors, les variables aléatoires $X_1, \dots, X_d$ sont indépendantes si et seulement si il existe d fonctions boréliennes positives $g_1, \dots, g_d$ telles que

$$f_X(x_1, \dots, x_d) = \prod_{i=1}^d g_i(x_i), \quad \lambda_p\text{-presque partout.} \quad (6.4)$$

Remarque 6.3. La décomposition (6.4) n'est pas unique, il suffit de changer g_1 en g_1/λ et g_2 en λg_2 , où $\lambda > 0$. Cependant, il y a unicité si l'on impose que l'intégrale de chaque fonction g_i est égale à 1. Lorsqu'il est ainsi g_i est une densité de X_i .

Exemple 6.2. Soit $X = (X_1, X_2)$ un vecteur à valeurs dans $\mathbb{R}^2$ de loi absolument continue de densité

$$f_X(x_1, x_2) = \mathbf{1}_{x_1 > 0, x_2 > 0} e^{-x_1 - x_2}.$$

On a alors $f_X(x_1, x_2) = g_1(x_1)g_2(x_2)$ pour tout $(x_1, x_2) \in \mathbb{R}^2$, avec

$$g_1(x_1) = \mathbf{1}_{x_1 > 0} e^{-x_1} \text{ et } g_2(x_2) = \mathbf{1}_{x_2 > 0} e^{-x_2}.$$

Par conséquent, les variables aléatoires X_1 et X_2 sont indépendantes. De plus, suivant la remarque 6.3, la loi de X_i est absolument continue de densité

$$f_{X_i} = \frac{g_i}{\int_{\mathbb{R}} g_i d\lambda_1} = g_i$$

car ici g_i est déjà la densité d'une loi exponentielle de paramètre 1 (donc son intégrale vaut 1). Ainsi, X_1 et X_2 sont deux variables aléatoires indépendantes de loi exponentielle de paramètre 1.

Exemple 6.3. Soit $Y = (Y_1, Y_2)$ un vecteur à valeurs dans $\mathbb{R}^2$ de loi absolument continue de densité

$$f_Y(y_1, y_2) = \mathbf{1}_{y_1 \neq y_2, y_1 > 0, y_2 > 0} e^{-y_1 - y_2}.$$

Cette fois-ci, on ne peut pas directement écrire $f_Y(y_1, y_2)$ sous la forme d'un produit $g'_1(y_1)g'_2(y_2)$ pour tout $(y_1, y_2) \in \mathbb{R}^2$. Cependant, on remarque que pour $y_1 \neq y_2$

$$f_Y(y_1, y_2) = f_X(y_1, y_2)$$

avec f_X la densité considérée dans l'exemple précédent. Mais l'ensemble $\{(x_1, x_2) \in \mathbb{R}^2, x_1 = x_2\}$ (c'est-à-dire la diagonale de $\mathbb{R}^2$) est de mesure nulle pour la mesure de Lebesgue λ_2 . Par conséquent,

$$f_Y(y_1, y_2) = f_X(y_1, y_2) = g_1(y_1)g_2(y_2) \quad \lambda_2\text{-presque partout},$$

où g_1 et g_2 sont définies dans l'exemple précédent. Par conséquent, les variables aléatoires Y_1 et Y_2 sont indépendantes.

Nous savons que si la loi d'un vecteur X est absolument continue, alors la loi de chacune de ses marginales l'est aussi. La réciproque est en général fausse. Cependant, si les marginales d'un vecteur X sont indépendantes et de loi absolument continue, alors le vecteur X est aussi de loi absolument continu.

Proposition 6.7 (Indépendance et marginales de loi absolument continue).

Pour tout $1 \leq i \leq d$, considérons X_i une variable aléatoire à valeurs dans $\mathbb{R}^{p_i}$ de loi absolument continue de densité $f_{X_i} : \mathbb{R}^{p_i} \rightarrow [0, +\infty]$. Alors, les variables $X_1, \dots, X_d$ sont mutuellement indépendantes si et seulement si la loi du vecteur $X = (X_1, \dots, X_d)$ est absolument continue de densité

$$f_X(x_1, \dots, x_d) = \prod_{i=1}^d f_{X_i}(x_i), \quad \lambda_p\text{-presque partout}$$

où $p = \sum_{i=1}^d p_i$.

Exemple 6.4. Soient X et Y deux variables aléatoires réelles indépendantes. On suppose que X est de loi uniforme sur $[0, 2]$ et que Y est de loi gaussienne $\mathcal{N}(0, 3)$. Alors, d'après la proposition précédente, la loi du vecteur (X, Y) admet un densité f par rapport à λ_2 , donnée par

$$f(x, y) = f_X(x)f_Y(y) = \mathbf{1}_{x \in [0, 2]} \frac{1}{6\sqrt{2\pi}} e^{-y^2/18}, \quad \forall (x, y) \in \mathbb{R}^2.$$

6.3 Addition de variables aléatoires indépendantes

La propriété suivante permet de caractériser la loi de la somme de deux variables aléatoires indépendantes. Elle se déduit aisément des théorèmes de Fubini.

Proposition 6.8 (Loi d'une somme de variables indépendantes).

Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $\mathbb{R}^d$ de lois respectives $\mathbb{P}_X$ et $\mathbb{P}_Y$, alors, pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable positive (ou mesurable telle que $f(X+Y)$ est intégrable), on a

$$\mathbb{E}(f(X+Y)) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x+y) d\mathbb{P}_X \otimes \mathbb{P}_Y(x, y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x+y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y).$$

Étudions à présent le cas des variables aléatoires discrètes.

Proposition 6.9 (Loi d'une somme de variables discrètes).

Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{R}^d$. Supposons que X est une variable aléatoire discrète de loi

$$\mathbb{P}_X = \sum_{i \in I} \mathbb{P}(X = i) \delta_i$$

avec $I \subset \mathbb{R}^d$ fini ou dénombrable. De même, supposons que Y est une variable aléatoire discrète de loi

$$\mathbb{P}_Y = \sum_{j \in J} \mathbb{P}(Y = j) \delta_j$$

avec $J \subset \mathbb{R}^d$ fini ou dénombrable. Si X et Y sont indépendantes, alors la variable aléatoire $X + Y$ est discrète de loi

$$\mathbb{P}_{X+Y} = \sum_{(i,j) \in I \times J} \mathbb{P}(X = i) \mathbb{P}(Y = j) \delta_{i+j}.$$

De manière équivalente, pour tout $k \in \mathbb{R}^d$,

$$\mathbb{P}(X + Y = k) = \mathbb{P}_{X+Y}(k) = \sum_{(i,j) \in I \times J, i+j=k} \mathbb{P}(X = i) \mathbb{P}(Y = j).$$

Démonstration. Pour tout $k \in \mathbb{R}^d$,

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \sum_{(i,j) \in I \times J} \mathbb{P}(X = i, Y = j \text{ et } X + Y = k) = \sum_{(i,j) \in I \times J} \mathbb{P}(X = i, Y = j \text{ et } i + j = k) \\ &= \sum_{(i,j) \in I \times J, i+j=k} \mathbb{P}(X = i, Y = j) \\ &= \sum_{(i,j) \in I \times J, i+j=k} \mathbb{P}(X = i) \mathbb{P}(Y = j), \end{aligned}$$

par indépendance de i et j . ■

Exemple 6.5. Soient X et Y deux variables aléatoires discrètes à valeurs dans $\mathbb{N}$. Si X et Y sont indépendantes, alors $X + Y$ est à valeurs dans $\mathbb{N}$ et

$$\forall n \in \mathbb{N}, \mathbb{P}(X + Y = n) = \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i).$$

Cette fonction de n est appelée le *produit de convolution* des fonctions $i \mapsto \mathbb{P}(X = i)$ et $j \mapsto \mathbb{P}(Y = j)$.

Nous donnons à présent des exemples de convolutions pour des lois classiques discrètes.

Proposition 6.10.

Soient X et Y deux variables aléatoires réelles indépendantes.

1. Si X suit la loi binomiale $\mathcal{B}(n, p)$ et si Y suit la loi $\mathcal{B}(m, p)$, alors $X + Y$ suit la loi $\mathcal{B}(n + m, p)$.
2. Si X suit la loi de Poisson $\mathcal{P}(\lambda)$ et si Y suit la loi de Poisson $\mathcal{P}(\mu)$, alors la loi de $X + Y$ est la loi de Poisson $\mathcal{P}(\lambda + \mu)$.

Démonstration.

1. Supposons que X et Y sont indépendantes, que X suit la loi $\mathcal{B}(n, p)$ et que Y suit la loi $\mathcal{B}(m, p)$. Alors, presque sûrement

$$X + Y \in \{k \in \mathbb{N} / 0 \leq k \leq n + m\}.$$

De plus, pour tout entier k tel que $0 \leq k \leq n + m$,

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i) \\ &= \sum_{i=0}^k C_n^i C_m^{k-i} p^k (1-p)^{n+m-k}. \end{aligned}$$

Or $\sum_{i=0}^k C_n^i C_m^{k-i}$ est le coefficient de t^k du polynôme $(1+t)^n (1+t)^m = (1+t)^{n+m}$ donc est égal à C_{n+m}^k . Alors, pour tout entier k tel que $0 \leq k \leq n + m$,

$$\mathbb{P}(X + Y = k) = C_{n+m}^k p^k (1-p)^{n+m-k}.$$

Par conséquent, $X + Y$ suit une loi binomiale $\mathcal{B}(n + m, p)$.

1. Supposons que X et Y sont indépendantes, que X suit la loi $\mathcal{P}(\lambda)$ et que Y suit la loi $\mathcal{P}(\mu)$. Alors, $X + Y \in \mathbb{N}$ presque sûrement. De plus, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i) \\ &= \sum_{i=0}^n \frac{\lambda^i}{i!} e^{-\lambda} \frac{\mu^{n-i}}{(n-i)!} e^{-\mu} \\ &= \frac{1}{n!} e^{-(\lambda+\mu)} \left(\sum_{i=0}^n C_n^i \lambda^i \mu^{n-i} \right) = \frac{1}{n!} e^{-(\lambda+\mu)} (\lambda + \mu)^n. \end{aligned}$$

Par conséquent, $X + Y$ a pour loi $\mathcal{P}(\lambda + \mu)$. ■

Remarque 6.4. Cette proposition se généralise aisément à la somme de n variables aléatoires mutuellement indépendantes. En particulier, si $X_1, \dots, X_n$ sont n variables aléatoires mutuellement indépendantes de loi de Bernoulli de paramètre p , alors

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

est une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ (car $\mathcal{B}(p) = \mathcal{B}(1, p)$). Ceci explique pourquoi le nombre de piles obtenus lors de n lancers indépendants d'une pièce truquée est modélisé par une loi binomiale $\mathcal{B}(n, p)$ où p est la probabilité d'obtenir pile lors d'un lancer.

Donnons à présent la loi de la somme de deux variables aléatoires réelles de loi absolument continue.

Proposition 6.11 (Loi d'une somme de variables indépendantes de loi absolument continue).

Soient X et Y des variables aléatoires réelles indépendantes. Si les lois de X et Y sont absolument continues de densités respectives f_X et f_Y , alors la loi de $X + Y$ est absolument continue de densité

$$f_{X+Y}(u) = \int_{\mathbb{R}^d} f_X(u-t) f_Y(t) \lambda_1(dt) = \int_{\mathbb{R}^d} f_X(t) f_Y(u-t) \lambda_1(dt). \quad (6.5)$$

En d'autres termes, nous observons que f_{X+Y} est le produit de convolution de f_X par f_Y , noté $f_X * f_Y$.

Démonstration. Soit $\varphi : \mathbb{R}^2 \longrightarrow [0, +\infty]$ une fonction borélienne positive.

Les variables X et Y étant indépendantes, la loi de (X, Y) est absolument continue de densité $f_{(X,Y)}$ définie par

$$\forall (x, y) \in \mathbb{R}^2, f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

Par conséquent,

$$\mathbb{E}(\varphi(X + Y)) = \iint_{\mathbb{R} \times \mathbb{R}} \varphi(x + y) f_X(x) f_Y(y) \lambda_1(dx) \lambda_1(dy).$$

D'après le théorème de Fubini-Tonelli,

$$\mathbb{E}(\varphi(X + Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(x + y) f_X(x) \lambda_1(dx) \right) f_Y(y) \lambda_1(dy).$$

En effectuant le changement de variable $z = x + y$ (y étant fixé), on constate que

$$\mathbb{E}(\varphi(X + Y)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \varphi(z) f_X(z - y) \lambda_1(dz) \right) f_Y(y) \lambda_1(dy).$$

Ainsi, d'après le théorème de Fubini-Tonelli,

$$\mathbb{E}(\varphi(X + Y)) = \int_{\mathbb{R}} \varphi(z) \left(\int_{\mathbb{R}} f_X(z - y) f_Y(y) \lambda_1(dy) \right) \lambda_1(dz).$$

L'égalité précédente étant vraie pour toute fonction $\varphi : \mathbb{R}^2 \longrightarrow [0, +\infty]$ borélienne positive, la loi de $X + Y$ est la loi absolument continue de densité donnée par (6.5). ■

Exemple 6.6. Soient U_1 et U_2 deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Alors, d'après la proposition 6.11, la variable aléatoire $U_1 + U_2$ est une variable aléatoire absolument continue de densité

$$\begin{aligned} f_{U_1+U_2}(u) &= \int_{\mathbb{R}} \mathbf{1}_{[0,1]}(u - t) \mathbf{1}_{[0,1]}(t) \lambda_1(dt) \\ &= \int_{\mathbb{R}} \mathbf{1}_{-1+u \leq t \leq u \text{ et } 0 \leq t \leq 1} \lambda_1(dt) \\ &= \begin{cases} 0 & \text{si } u < 0 \text{ ou } u > 2 \\ u & \text{si } u \in [0, 1] \\ 2 - u & \text{si } u \in [1, 2]. \end{cases} \end{aligned}$$

Nous avons ainsi obtenu la densité de la loi de $U_1 + U_2$. Cette loi est appelée la *loi triangulaire* (n'hésitez pas à tracer la fonction $f_{U_1+U_2}$ afin de comprendre cette dénomination).

Chapitre 7

Variance, moments et espaces L^p

Dans ce chapitre, nous étudions les variables aléatoires dont la puissance $p^{\text{ème}}$ est intégrable. Nous introduisons notamment les concepts de variance, de covariance et de moments d'ordre p . Quelques inégalités classiques sont également présentées.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités (qui ne sera, sauf exception, pas explicité) ;
- $\mathbb{R}^d$ est muni de sa tribu borélienne.
- Sauf mention du contraire, les variables aléatoires considérées sont définies sur $(\Omega, \mathcal{F}, \mathbb{P})$.

7.1 Variance et covariance

Définition-Proposition 7.1 (Covariance/Variance).

Soient X, Y deux variables aléatoires à valeurs réelles et de carrés intégrables, c'est-à-dire telles que $\mathbb{E}(X^2) < \infty$ et $\mathbb{E}(Y^2) < \infty$. Alors

1. la quantité

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

est bien définie et est appelée **covariance entre X et Y** ;

2. les variables aléatoires X et Y sont dites **décorrélées** si $\text{Cov}(X, Y) = 0$;
3. la covariance $\text{Cov}(X, X)$ est encore appelée **variance de X** et notée $\text{Var}X$. Enfin, le réel positif $\sqrt{\text{Var}X}$ est appelé **écart-type de X** .

Proposition 7.2.

Soient X, Y, Z des variables aléatoires réelles de carré intégrable.

1. X est constante presque sûrement si et seulement si $\text{Var}X = 0$.
2. La covariance est bilinéaire symétrique, c'est-à-dire que, pour tout $a \in \mathbb{R}$,

$$\text{Cov}(X, Y) = \text{Cov}(Y, X) \quad \text{et} \quad \text{Cov}(X, aY + Z) = a\text{Cov}(X, Y) + \text{Cov}(X, Z).$$

3. Pour tout $a \in \mathbb{R}$,

$$\text{Var}(aX) = a^2\text{Var}X, \quad \text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2\text{Cov}(X, Y) \quad \text{et} \quad \text{Var}(X + a) = \text{Var}X.$$

4. $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. En particulier, $\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

La démonstration des propriétés élémentaires ci-dessus est laissée en exercice.

Remarque 7.1. Soient X, Y deux variables aléatoires à valeurs réelles et de carrés intégrables.

1. Des exemples de calculs de variance pour des lois classiques sont donnés dans la section suivante et les résultats sont résumés dans le tableau 7.1 en fin de chapitre.
2. Soit X une variable constante presque sûrement. Alors $X = \mathbb{E}(X)$ presque sûrement. Par suite, quel que soit Y , $\text{Cov}(X, Y) = 0$. Ainsi la variable X est décorrélée de toute autre variable de carré intégrable (et donc en particulier d'elle-même).
3. La covariance est déterminée par la loi du couple (X, Y) . En effet, notant $\mathbb{P}_{(X, Y)}$ la loi de la variable aléatoire (X, Y) ,

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

avec

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xy d\mathbb{P}_{(X, Y)}(x, y), \quad \mathbb{E}(X) = \int_{\mathbb{R}^2} x d\mathbb{P}_{(X, Y)}(x, y), \quad \text{et} \quad \mathbb{E}(Y) = \int_{\mathbb{R}^2} y d\mathbb{P}_{(X, Y)}(x, y).$$

4. Si deux variables aléatoires X et Y sont de carré intégrable et indépendantes, alors elles sont décorrélées. En effet,

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$$

Attention ! La réciproque est fausse.

5. Si $X_1, \dots, X_n$ sont toutes de carré intégrable, alors

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

7.2 Moments d'une variable aléatoire

Définition 7.1 (Moment d'une variable aléatoire).

Soit $X : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une variable aléatoire à valeurs dans $\overline{\mathbb{R}}$ et $p > 0$. La variable aléatoire X admet un **moment d'ordre p** si

$$\mathbb{E}(|X|^p) < +\infty,$$

On appelle **moment d'ordre p** de X la quantité $\mathbb{E}(X^p)$.

Les moments d'une variable gaussienne centrée réduite ont été donnés dans la proposition 4.12 du chapitre 4. Rappelons que ces moments ont été obtenus en dérivant la fonction caractéristique. N'oublions pas qu'un autre théorème utile pour calculer un moment d'une variable X est le théorème du transport, théorème permettant d'exprimer ce moment à l'aide d'une intégrale contre la loi de X . En effet

$$\mathbb{E}(|X|^p) = \int_{\mathbb{R}} |x|^p \mathbb{P}_X(dx) \quad \text{et, si } \mathbb{E}(|X|^p) < \infty, \text{ alors } \mathbb{E}(X^p) = \int_{\mathbb{R}} x^p \mathbb{P}_X(dx).$$

Définition 7.2 (Norme infinie).

Soit $X : (\Omega, \mathcal{F}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ une variable aléatoire à valeurs dans $\overline{\mathbb{R}}$. La variable aléatoire X est dite **uniformément bornée** si il existe une constante $C > 0$ telle que $|X| \leq C$ presque sûrement.

On appelle **norme infinie de X** , notée $\|X\|_\infty$, est définie par

$$\|X\|_\infty = \inf \{ C \in [0, +\infty] \mid |X| \leq C \text{ presque sûrement} \}.$$

Remarque 7.2.

1. Si $\|X\|_\infty < \infty$, alors $X \leq \|X\|_\infty$ presque sûrement.
2. La fonction $X \mapsto \|X\|_\infty$ est définie sur l'espace vectoriel des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$ dans $\mathbb{R}$ et c'est une norme, c'est-à-dire que, pour tout $\lambda \in \mathbb{R}$ et toutes variables aléatoires X, Y ,
 - (i) $\|\lambda X\|_\infty = |\lambda| \|X\|_\infty$,
 - (ii) $\|X + Y\|_\infty \leq \|X\|_\infty + \|Y\|_\infty$,
 - (iii) si $\|X\|_\infty = 0$, alors $X = 0$ presque sûrement.
3. On montre aisément que $\| |X|^p \|_\infty = \|X\|_\infty^p$. En particulier, si $\|X\|_\infty < \infty$, alors $|X|^p$ est bornée par $\|X\|_\infty^p$, donc X admet un moment d'ordre p borné par $\|X\|_\infty^p$.
4. Soit X une variable aléatoire de loi absolument continue sur $\mathbb{R}$, de densité f_X . La norme infinie de X peut être identifiée à partir de f_X de la façon suivante

$$\|X\|_\infty = \inf \left\{ C \in [0, \infty] / \int_{]C, \infty[} f_X(x) \lambda_1(dx) = 0 \right\}.$$

Nous donnons à présent des exemples de calculs de moments, récapitulés dans le tableau 7.1 en fin de chapitre.

7.2.1 Calcul pour les lois binomiales

Soit X une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$. Calculons $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$, qui sont bien définies car X et X^2 sont positives presque sûrement.

- Par définition

$$\mathbb{E}(X) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = n \sum_{k=1}^n C_{n-1}^{k-1} p^k (1-p)^{n-k}$$

car $n C_{n-1}^{k-1} = k C_n^k$ pour $1 \leq k \leq n$. Alors,

$$\mathbb{E}(X) = n \sum_{l=0}^{n-1} C_{n-1}^l p^{l+1} (1-p)^{n-l} = pn$$

d'après la formule du binôme de Newton.

- Au lieu de calculer l'espérance du carré de X on évalue $\mathbb{E}[X(X-1)]$. Alors, d'après le théorème du transport,

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^n k(k-1) C_n^k p^k (1-p)^{n-k} = n(n-1) \sum_{k=2}^n C_{n-2}^{k-2} p^k (1-p)^{n-k}$$

Si $n = 1$, alors $\mathbb{E}(X(X-1)) = 0$ et donc $\mathbb{E}(X^2) = \mathbb{E}(X) = p$ car $X^2 = X(X-1) + X$. Par conséquent, si $n = 1$

$$\text{Var}X = p - p^2 = p(1-p).$$

Supposons maintenant $n \geq 2$. Alors,

$$\mathbb{E}(X(X-1)) = n(n-1) \sum_{k=0}^{n-2} C_{n-2}^k p^{k+2} (1-p)^{n-k-2} = n(n-1)p^2.$$

Alors, $\mathbb{E}(X^2) = n(n-1)p^2 + pn$ et $\text{Var}X = n(n-1)p^2 + pn - p^2n^2 = np(1-p)$.

- Remarquons enfin que $\|X\|_\infty = n$. En effet,

$$\mathbb{P}(|X| \leq n) = \sum_{k \geq 0} C_n^k p^k (1-p)^{n-k} = 1$$

et, pour tout $0 < C < n$,

$$\mathbb{P}(|X| \leq C) = 1 - \mathbb{P}(|X| > C) \geq 1 - \mathbb{P}(X = n) = 1 - p^n < 1.$$

Ainsi, n est bien la plus petite valeur telle que $|X| \leq n$ presque sûrement.

7.2.2 Calcul pour les lois de Poisson

Soit X une variable aléatoire de loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$. Calculons $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$, qui sont bien définies car X et X^2 sont positives presque sûrement.

- Par définition,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{n \geq 0} n \mathbb{P}(Y = n) = \sum_{n \geq 0} n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \left(\sum_{n \geq 1} \frac{\lambda^n}{(n-1)!} \right) = \lambda e^{-\lambda} \left(\sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!} \right) \\ &= \lambda e^{-\lambda} e^{\lambda} = \lambda.\end{aligned}$$

- Au lieu de calculer l'espérance du carré de X on évalue $\mathbb{E}[X(X-1)]$.

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{n \geq 0} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \left(\sum_{n \geq 2} \frac{\lambda^n}{(n-2)!} \right) = \lambda^2 e^{-\lambda} \left(\sum_{n \geq 2} \frac{\lambda^{n-2}}{(n-2)!} \right) \\ &= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2.\end{aligned}$$

Mais $X^2 = X(X-1) + X$ donc $\mathbb{E}(X^2) = \lambda^2 + \mathbb{E}[X] = \lambda^2 + \lambda$ et $\text{Var}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.

- Remarquons enfin que $\|X\|_{\infty} = \infty$. En effet, pour tout $C \in [0, +\infty[$,

$$\begin{aligned}\mathbb{P}(|X| \leq C) &= 1 - \mathbb{P}(|X| > C) \\ &\geq 1 - \mathbb{P}(X = \lfloor C \rfloor + 1) \\ &= 1 - \frac{\lambda^{\lfloor C \rfloor + 1}}{(\lfloor C \rfloor + 1)!} e^{-\lambda} < 1.\end{aligned}$$

Ainsi, $C = +\infty$ est bien la plus petite valeur telle que $|X| \leq C$ presque sûrement.

7.2.3 Calcul pour les lois géométrique

Soit X une variable aléatoire de loi géométrique $\mathcal{G}(p)$ avec $p \in]0, 1[$. Calculons $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$, qui sont bien définies car X et X^2 sont positives presque sûrement. Pour le calcul de l'espérance et la variance de X on procède comme pour la loi de Poisson, on évalue directement $\mathbb{E}(X)$ puis $\mathbb{E}[X(X-1)]$. On utilise les identités suivantes :

$$\sum_{n \geq 0} x^n = \frac{1}{1-x}; \quad \text{pour tout } x \in]-1, 1[. \quad (7.1)$$

$$\frac{d}{dx} \left(\sum_{n \geq 0} x^n \right) = \sum_{n \geq 1} n x^{n-1} = \frac{1}{(1-x)^2}; \quad \text{pour tout } x \in]-1, 1[\quad (7.2)$$

$$\frac{d^2}{dx^2} \left(\sum_{n \geq 0} x^n \right) = \sum_{n \geq 2} n(n-1) x^{n-2} = \frac{2}{(1-x)^3}; \quad \text{pour tout } x \in]-1, 1[\quad (7.3)$$

On en déduit que

$$\mathbb{E}(X) = \frac{1}{p}, \quad \mathbb{E}[X(X-1)] = \frac{2(1-p)}{p^2}, \quad \mathbb{E}(X^2) = \frac{2}{p^2} - \frac{1}{p} \quad \text{et} \quad \text{Var}(X) = \frac{1-p}{p^2}.$$

De même que pour la loi de Poisson, on obtient $\|X\|_{\infty} = +\infty$.

7.2.4 Calcul pour la loi uniforme sur $[a, b]$

Soit X une variable aléatoire de loi uniforme sur $[a, b]$ (avec $a, b \in \mathbb{R}$ tel que $a < b$), c'est-à-dire une variable aléatoire de loi absolument continue ayant pour densité la fonction f définie sur $\mathbb{R}$ par

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

Calculons $\mathbb{E}(X)$ qui est bien définie car $|X|$ est bornée $\max(|a|, |b|)$ presque sûrement. D'après le théorème du transport,

$$\mathbb{E}(X) = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2} \quad \text{et} \quad \mathbb{E}(X^2) = \frac{1}{b-a} \int_{[a,b]} x^2 \lambda_1(dx) = \frac{b^2 + ab + a^2}{3}.$$

$$\text{D'où } \text{Var}(X) = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Comme mentionné plus haut, on a, d'après le théorème du transport,

$$\begin{aligned} \mathbb{P}(|X| \leq \max(|a|, |b|)) &= \mathbb{P}(-\max(|a|, |b|) \leq X \leq \max(|a|, |b|)) \\ &\geq \mathbb{P}(a \leq X \leq b) \\ &= \int_{[a,b]} \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) \lambda_1(dx) = 1, \end{aligned}$$

donc $|X| \leq \max(|a|, |b|)$ presque sûrement. De plus, pour tout $0 \leq C < \max(|a|, |b|)$,

$$\begin{aligned} \mathbb{P}(|X| \leq C) &= 1 - \mathbb{P}(|X| > C) = 1 - \mathbb{P}(X < -C \text{ ou } X > C) \\ &= 1 - \int_{\mathbb{R} \setminus [-C, C]} \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) \lambda_1(dx) \\ &= 1 - \int_{[a,b] \setminus [-C, C]} \frac{1}{b-a} \lambda_1(dx). \end{aligned}$$

Or $0 \leq C < \max(|a|, |b|)$, donc $[a, b] \setminus [-C, C]$ contient un intervalle de mesure non nulle. Par conséquent,

$$\mathbb{P}(|X| \leq C) < 1.$$

Finalement, nous en déduisons que $\|X\|_\infty = \max(|a|, |b|)$.

7.2.5 Calcul pour les lois exponentielles

Soit X_λ une variable aléatoire réelle de loi exponentielle de paramètre $\lambda > 0$, c'est-à-dire de loi absolument continue ayant pour densité la fonction f_λ définie sur $\mathbb{R}$ par

$$\forall x \in \mathbb{R}, f_\lambda(x) = \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{[0, +\infty[}(x).$$

Calculons $\mathbb{E}(X_\lambda)$ et $\mathbb{E}(X_\lambda^2)$, qui sont bien définies car X_λ et X_λ^2 sont positives presque sûrement. La variable aléatoire $Y_\lambda = X_\lambda/\lambda$ suit une loi exponentielle de paramètre un. De plus, $\mathbb{E}(X_\lambda) = \lambda \mathbb{E}(Y_\lambda)$ et $\mathbb{E}(X_\lambda^2) = \lambda^2 \mathbb{E}(Y_\lambda^2)$. Il suffit donc de calculer l'espérance et la variance lorsque $\lambda = 1$. Pour $\lambda = 1$,

$$\mathbb{E}(X_1) = \int_{[0, +\infty[} x e^{-x} \lambda_1(dx) \quad \text{et} \quad \mathbb{E}(X_1^2) = \int_{[0, +\infty[} x^2 e^{-x} \lambda_1(dx).$$

Les fonctions intégrées étant continues, en procédant par intégrations par parties, on constate que

$$\mathbb{E}(X_1) = \int_0^{+\infty} e^{-x} dx = 1 \quad \text{et} \quad \mathbb{E}(X_1^2) = 2 \int_0^{+\infty} x e^{-x} dx = 2.$$

On en déduit $\text{Var}(X_1) = 1$.

Enfin, $\|X\|_\infty = +\infty$, car, pour tout $C \in [0, +\infty[$,

$$\mathbb{P}(|X| \leq C) = 1 - \mathbb{P}(|X| > C) = 1 - e^{-\lambda C} < 1.$$

7.2.6 Calcul pour les lois gaussiennes

Soit X une variable de loi $\mathcal{N}(m, \sigma^2)$. Nous savons que $Y = \frac{X-m}{\sigma}$ suit une loi gaussienne centrée réduite, c'est-à-dire qu'elle a pour densité par rapport à λ_1 la fonction

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$

En utilisant la parité de f , on montre facilement que Y est centrée et donc par linéarité de l'espérance que $\mathbb{E}(X) = m$.

Par ailleurs,

$$\mathbb{E}(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^2 e^{-t^2/2} \lambda_1(dt) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t(te^{-t^2/2}) \lambda_1(dt).$$

En remarquant que l'intégrale précédente au sens de Lebesgue coïncide avec l'intégrale au sens de Riemann et en utilisant une intégration par parties,

$$\mathbb{E}(Y^2) = \frac{1}{\sqrt{2\pi}} \left(\left[-te^{-t^2/2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-t^2/2} dt \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} dt = 1.$$

Par conséquent, Y admet un moment d'ordre 2 et $\text{Var } Y = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = 1$. Étant donné que

$$X = \sigma Y + m,$$

X admet aussi un moment d'ordre 2 et $\text{Var } X = \sigma^2 \text{Var } Y = \sigma^2$.

On montre aisément que $\|X\|_{\infty} = +\infty$.

7.3 Inégalités classiques

Cette partie énonce plusieurs inégalités classiques.

Proposition 7.3 (Inégalité de Minkowski).

Soit $p \in [1, +\infty]$. Si X et Y sont deux variables aléatoires toutes deux à valeurs dans $[0, +\infty]$ presque sûrement ou admettent toutes deux des moments d'ordre p , alors

$$(\mathbb{E}(|X + Y|^p))^{1/p} \leq (\mathbb{E}(|X|^p))^{1/p} + (\mathbb{E}(|Y|^p))^{1/p}.$$

En particulier, si X et Y admettent des moments d'ordre p , alors $X + Y$ admet un moment d'ordre p .

Proposition 7.4 (Inégalité de Hölder).

Soient X et Y deux variables aléatoires à valeurs dans $\overline{\mathbb{R}}$.

1. Soient $p, q \in]1, +\infty[$ tels que $1/p + 1/q = 1$ (autrement dit $q = p/(1 - p)$). Alors

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{1/p} (\mathbb{E}(|Y|^q))^{1/q}$$

Par suite, si X admet un moment d'ordre p et si Y admet un moment d'ordre q , alors XY est intégrable.

2. Supposons $p = 1$. Alors

$$\mathbb{E}(|XY|) \leq \mathbb{E}(|X|) \|Y\|_{\infty}.$$

Par suite, si X admet un moment d'ordre p et si Y est uniformément bornée, alors XY est intégrable.

Corollaire 7.5.

Si $p = 2$ (et donc $q = 2$), l'inégalité de Hölder est encore nommée inégalité de Cauchy-Schwarz et s'écrit

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

L'inégalité de Hölder permet d'établir une comparaison entre le moment d'ordre p et celui d'ordre $p' > p$ lorsqu'ils existent.

Corollaire 7.6.

Pour tous $1 \leq p \leq p' \leq +\infty$ et toute variable aléatoire X à valeurs dans $\overline{\mathbb{R}}$, on a

$$\mathbb{E}(|X|^p)^{1/p} \leq \mathbb{E}(|X|^{p'})^{1/p'}. \quad (7.4)$$

Démonstration. Soient $1 \leq p \leq p' \leq +\infty$ et X une variable aléatoire. D'après l'inégalité de Hölder,

$$\mathbb{E}(|X|^p) = \mathbb{E}(|X|^p 1) \leq \mathbb{E}(|X|^{p'})^{p/p'} \mathbb{E}(1^{1-\frac{1}{p/p'}})^{1-p/p'} = \mathbb{E}(|X|^{p'})^{p/p'}.$$

La fonction $x \mapsto x^{1/p}$ étant croissante sur $\mathbb{R}_+$, on en déduit que

$$\mathbb{E}(|X|^p)^{1/p} \leq \mathbb{E}(|X|^{p'})^{1/p'}.$$

Proposition 7.7 (Inégalité de Jensen).

1. Pour toute fonction convexe $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ et toute variable aléatoire X à valeurs dans $\mathbb{R}_+$,

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)). \quad (7.5)$$

2. Si X est intégrable et si $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction convexe telle que $\phi(X)$ est intégrable, alors

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)). \quad (7.6)$$

Proposition 7.8 (Inégalité de Markov).

Si X est une variable aléatoire à valeurs dans $[0, +\infty]$ presque sûrement et si $\lambda \in \mathbb{R}_+^*$, alors,

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

Démonstration. Soit $\lambda \in \mathbb{R}_+^*$. Remarquons que

$$X \geq \lambda \mathbf{1}_{\{X \geq \lambda\}} \geq 0$$

presque sûrement. Alors,

$$\mathbb{E}(X) \geq \mathbb{E}(\lambda \mathbf{1}_{\{X \geq \lambda\}}) = \lambda \mathbb{P}(X \geq \lambda),$$

ce qui prouve l'inégalité de Markov. ■

Proposition 7.9 (Inégalité de Bienaymée-Tchebytschef).

Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, pour tout $\lambda \in]0, +\infty[$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2},$$

où $\text{Var}(X) = \mathbb{E}(|X - \mathbb{E}(X)|^2)$.

Démonstration. Posons $Y = |X - \mathbb{E}(X)|$. Alors $Y \geq 0$ et

$$\{|X - \mathbb{E}(X)| \geq \lambda\} = \{Y \geq \lambda\} = \{Y^2 \geq \lambda^2\}.$$

D'après l'inégalité de Markov appliquée à la variable Y^2 ,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) = \mathbb{P}(Y^2 \geq \lambda^2) \leq \frac{\mathbb{E}(Y^2)}{\lambda^2},$$

ce qui prouve l'inégalité de Bienaymée-Tchebytschef car $\mathbb{E}(Y^2) = \text{Var}(X)$. ■

7.4 Les espaces L^p , $p \geq 1$

7.4.1 Définitions et premières propriétés

Définition 7.3 (Espace $L^p(\Omega, \mathcal{F}, \mathbb{P})$).

1. Pour tout $p \in]0, +\infty[$, $L^p(\Omega, \mathcal{F}, \mathbb{P})$ est l'ensemble des variables aléatoires X définies sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $\overline{\mathbb{R}}$ telles que

$$\mathbb{E}(|X|^p) < \infty.$$

2. $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ est l'ensemble des variables aléatoires X définies sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans $\overline{\mathbb{R}}$ vérifiant

$$\exists c \in \mathbb{R}_+, \text{ tel que } |X| \leq c \text{ } \mathbb{P}\text{-presque sûrement.}$$

Remarque 7.3. La notion d'espace L^p peut être généralisée en remplaçant dans la définition précédente

- la probabilité $\mathbb{P}$ par une mesure positive μ
- l'espérance $\mathbb{E}$ par l'intégrale contre la mesure μ
- le terme *variable aléatoire* par *fonction mesurable*.

Les espaces ainsi introduits sont notés $L^p(\Omega, \mathcal{F}, \mu)$, $p \in]0, +\infty[$. Tous les résultats de cette section, restent vrais dans ce cadre plus général.

Proposition 7.10.

Soit $p \in [1, +\infty]$. L'espace $L^p(\Omega, \mathcal{F}, \mathbb{P})$ muni de la norme $\|\cdot\|_p$ définie sur $L^p(\Omega, \mathcal{F}, \mathbb{P})$ par

$$\begin{cases} \|X\|_p = (\mathbb{E}(|X|^p))^{1/p} \text{ si } p \in [1, +\infty[, \\ \|X\|_\infty = \inf\{c \geq 0 \text{ tel que } |X| \leq c \text{ presque sûrement}\} \text{ si } p = +\infty, \end{cases}$$

est un espace vectoriel normé.

Remarque 7.4. Pour $p \in]0, 1[$, l'application

$$X \mapsto (\mathbb{E}(|X|^p))^{1/p}$$

est bien définie mais n'est pas une norme sur $L^p(\Omega, \mathcal{F}, \mathbb{P})$ car elle ne satisfait pas l'inégalité triangulaire.

Les espaces $L^p(\Omega, \mathcal{F}, \mathbb{P})$, $p \geq 1$, sont en fait des espaces de Banach, notion introduite dans la définition suivante. Pour ce chapitre, cette définition un peu technique peut être omise en première lecture, l'essentiel étant de retenir la définition des normes, le fait que L^p est un espace vectoriel et les inégalités classiques données en section 7.3. Néanmoins, cette notion d'espace de Banach est importante notamment lorsque l'on étudie la convergence d'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ dans L^p (cf. chapitre 8).

Définition 7.4.

Soit E un espace vectoriel muni d'une norme $\|\cdot\|$.

1. Une suite $(u_n)_{n \in \mathbb{N}}$ d'éléments de E est dite de Cauchy de E si

$$\forall \varepsilon > 0, \exists n_0 \geq 1 \text{ tel que } \forall n \geq n_0, \forall m \geq 0, \|u_n - u_{n+m}\| \leq \varepsilon.$$

2. L'espace vectoriel E muni de la norme $\|\cdot\|$ est un espace de Banach si toute suite de Cauchy de E converge.

Proposition 7.11.

Soit $p \in [1, +\infty]$. L'espace vectoriel $L^p(\Omega, \mathcal{F}, \mathbb{P})$ muni de la norme $\|\cdot\|_p$ est un espace de Banach.

Terminons par une comparaison entre les espaces L^p , conséquence immédiate du corollaire 7.6.

Proposition 7.12.

Pour tous $0 < p \leq p' \leq +\infty$ $L^{p'}(\Omega, \mathcal{F}, \mathbb{P}) \subset L^p(\Omega, \mathcal{F}, \mathbb{P})$. En particulier, si X admet un moment d'ordre $p' \geq 1$, alors elle admet un moment d'ordre $p \in]1, p']$.

Remarque 7.5. Cette proposition est fausse si on remplace $\mathbb{P}$ par une mesure positive quelconque μ .

7.4.2 L'espace L^2

Dans cette section, nous nous intéressons plus particulièrement à l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Nous savons que cet espace est un espace de Banach pour la norme $\|\cdot\|_2$ définie par

$$\forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \|X\|_2 = \mathbb{E}(X^2)^{1/2}.$$

Cette norme est en fait associée au produit scalaire $\langle \cdot, \cdot \rangle$ défini par

$$\forall X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \langle X, Y \rangle = \mathbb{E}(XY),$$

où XY est bien intégrable d'après l'inégalité de Cauchy-Schwartz. La définition suivante précise les notions de produit scalaire et d'espace de Hilbert.

Définition 7.5.

Soit E un $\mathbb{K}$ -espace vectoriel avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

1. Une application $\langle \cdot, \cdot \rangle : E \times E \longrightarrow \mathbb{K}$ est un produit scalaire sur E si elle satisfait les trois assertions suivantes :

- (i) elle est linéaire en sa première variable, c'est-à-dire que pour tous $a, b \in \mathbb{K}$ et tous $x, y, z \in E$,

$$\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$$

- (ii) pour tous $x, y \in E$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$

- (iii) $\langle x, x \rangle = 0 \implies x = 0$.

La norme $\| \cdot \|$ associée à un produit scalaire $\langle \cdot, \cdot \rangle$ est définie par

$$\|x\| = (\langle x, x \rangle)^{1/2}, \quad \forall x \in E$$

2. L'espace E muni du produit scalaire $\langle \cdot, \cdot \rangle$ est un espace de Hilbert si il est un espace de Banach pour la norme associée à ce produit scalaire.

Proposition 7.13.

L'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert.

La fin de cette partie sur l'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est consacrée à la notion de projection orthogonale.

Théorème 7.14 (Projection orthogonale).

Soit $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et F un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Alors il existe une unique variable aléatoire $P_F(X) \in F$ telle que $\|X - P_F(X)\|_2 = \inf_{Z \in F} \|X - Z\|_2$, ou, de manière équivalente, telle que

$$\mathbb{E}((X - P_F(X))^2) = \inf_{Z \in F} \mathbb{E}((X - Z)^2).$$

De plus, $P_F(X)$ est l'unique variable aléatoire de F telle que $\langle X - P_F(X), Z \rangle = 0$, $\forall Z \in F$, ou, de manière équivalente, telle que

$$\mathbb{E}((X - P_F(X))Z) = 0, \quad \forall Z \in F.$$

La variable aléatoire $P_F(X)$ est appelée le **projeté orthogonal de X sur F** . L'application $P_F : L^2 \rightarrow F$ est appelée la **projection orthogonale sur F** ou le **projecteur orthogonal sur F** . Il s'agit d'une application linéaire.

Exemple 7.1. Soient X et Y deux variables aléatoires de carré intégrable. L'ensemble

$$F = \{a + bY, a \in \mathbb{R}, b \in \mathbb{R}\}$$

est le sous-espace vectoriel fermé de L^2 engendré par 1 et Y . Dans cette situation, déterminer $P_F(x)$ revient à déterminer la variable aléatoire $a_0 + b_0Y$ avec $a_0, b_0 \in \mathbb{R}$ choisis de sorte que

$$\mathbb{E}((X - (a_0 + b_0Y))^2) = \min_{a, b \in \mathbb{R}} \mathbb{E}((X - (a + bY))^2).$$

Autrement dit, cela revient à déterminer la meilleure approximation de X (au sens de $\| \cdot \|_2$) par une fonction affine de Y .

Pour utiliser ce résultat en pratique, il nous faut mener le calcul de $P_F(x)$, sujet de la suite de cette section.

Proposition 7.15.

Soit $I \subset \mathbb{N}$. Considérons $(X_k)_{k \in I}$ un système orthonormal de $L^2(\Omega, \mathcal{F}, \mathbb{P})$, c'est-à-dire une famille de variables aléatoires de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ telle que

$$\forall k, l \in I, \mathbb{E}(X_k^2) = \|X_k\|_2^2 = 1 \text{ et } \mathbb{E}(X_k X_l) = \langle X_k, X_l \rangle = 0.$$

Considérons $F = \overline{\text{Vect}(X_k, k \in I)}$ le sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ engendré par $(X_k)_{k \in I}$. Alors, pour tout $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, la projection orthogonale de X sur F est

$$P_F(X) = \sum_{k \in I} \langle X, X_k \rangle X_k = \sum_{k \in I} \mathbb{E}(X X_k) X_k.$$

De plus,

$$\mathbb{E}(P_F(X)^2) = \sum_{k \in I} \langle X, X_k \rangle^2 = \sum_{k \in I} \mathbb{E}(X X_k)^2 \leq \mathbb{E}(X^2),$$

avec égalité si et seulement si $X = P_F(X) \in F$ presque sûrement.

Remarque 7.6. En pratique, on dispose souvent d'une base de F qui n'est pas orthonormale (comme dans l'exemple 7.1 ci-dessus). Dans ce cas, on utilise le procédé d'orthonormalisation de Gram-Schmidt pour poursuivre. Rappelons ce procédé sur un exemple : si $F = \text{Vect}(X_1, X_2)$ avec X_1 et X_2 non colinéaires, alors la famille (Y_1, Y_2) , définie par

$$Y_1 = \frac{X_1}{\|X_1\|_2} = \frac{X_1}{\sqrt{\mathbb{E}(X_1^2)}} \text{ et } Y_2 = \frac{X_2 - \langle X_2, Y_1 \rangle Y_1}{\|X_2 - \langle X_2, Y_1 \rangle Y_1\|_2} = \frac{X_2 - \mathbb{E}(X_2 Y_1) Y_1}{\sqrt{\mathbb{E}((X_2 - \mathbb{E}(X_2 Y_1) Y_1)^2)}}$$

est une base orthonormale de F .

Exemple 7.2. Reprenons l'exemple 7.1, où $F = \text{Vect}(1, Y)$. Dans ce cas, la famille (Y_1, Y_2) , définie par

$$Y_1 = \frac{1}{\sqrt{\mathbb{E}(1^2)}} = 1 \text{ et } Y_2 = \frac{Y - \mathbb{E}(Y) 1}{\sqrt{\mathbb{E}((Y - \mathbb{E}(Y) 1)^2)}} = \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var } Y}}$$

est une base orthonormale de F . D'après la proposition 7.15, le projeté orthogonal de X sur F est donc donné par

$$P_F(X) = \mathbb{E}(X 1) 1 + \mathbb{E} \left(X \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var } Y}} \right) \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var } Y}}.$$

En développant le calcul, on obtient donc

$$P_F(X) = \mathbb{E}(X) - \frac{\text{Cov}(X, Y)}{\text{Var } Y} \mathbb{E}(Y) + \frac{\text{Cov}(X, Y)}{\text{Var } Y} Y.$$

Remarque 7.7.

1. Les notions de base orthonormale et de projection orthogonale sont généralisables aux de espaces de Hilbert (en effet, les résultats précédents s'étendent à tous les espaces de Hilbert).
2. On peut généraliser l'exemple 7.2 ci-dessus au cas où l'on souhaite approximer X par une combinaison linéaire de plusieurs variables aléatoires $Y_1, Y_2, \dots, Y_n$. On parle alors de régression linéaire.

Nous terminons par la notion classique de coefficient de corrélation linéaire entre deux variables X, Y de carrés intégrables.

Définition 7.6 (Coefficient de corrélation linéaire).

Soient $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Supposons que X et Y ne sont pas presque sûrement constantes, c'est-à-dire supposons que

$$\mathbb{E}(X^2) < +\infty, \mathbb{E}(Y^2) < +\infty, \text{Var}(X) > 0 \quad \text{et} \quad \text{Var}(Y) > 0.$$

Le coefficient de corrélation linéaire entre X et Y est le réel

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

Remarque 7.8. L'exemple 7.2 et la linéarité de P_F montrent que nous avons, en posant $F = \text{Vect}(1, Y)$,

$$P_F \left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \right) = \rho_{XY} \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var}(Y)}}.$$

Autrement dit, $\rho_{XY} \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var}(Y)}}$ est la meilleure approximation de $\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}$ par une variable aléatoire de la forme $a + bY$ pour la norme $\|\cdot\|_2$.

La proposition suivante justifie la terminologie « corrélation linéaire ».

Proposition 7.16.

Soient X, Y deux variables aléatoires à valeurs dans $\overline{\mathbb{R}}$. Supposons que $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ et que X et Y ne sont pas constantes, c'est-à-dire supposons que

$$\mathbb{E}(X^2) < +\infty, \mathbb{E}(Y^2) < +\infty, \text{Var}(X) > 0 \quad \text{et} \quad \text{Var}(Y) > 0.$$

Notons ρ_{XY} le coefficient de corrélation linéaire entre les variables X et Y .

1. Alors, $|\rho_{XY}| \leq 1$.
2. Enfin, $\rho_{XY} = 1$ (respectivement $\rho_{XY} = -1$) si et seulement si il existe $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ tel que $Y = aX + b$ (respectivement $Y = -aX + b$).

Démonstration.

1. D'après l'inégalité de Cauchy-Schwarz,

$$|\text{Cov}(X, Y)| = |\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))| \leq \sqrt{\mathbb{E}((X - \mathbb{E}(X))^2)} \sqrt{\mathbb{E}((Y - \mathbb{E}(Y))^2)} = \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}.$$

Par conséquent, $|\rho_{XY}| \leq 1$.

2. Supposons que $|\rho_{XY}| = 1$. Alors

$$\mathbb{E} \left(\left(\frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var}(Y)}} \right)^2 \right) = 1 = \mathbb{E} \left(\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \right)^2 \right).$$

Donc, d'après la remarque 7.8,

$$\mathbb{E} \left(\left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \right)^2 \right) = \mathbb{E} \left(P_F \left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \right)^2 \right).$$

Nous déduisons donc de la dernière partie de la proposition 7.15 (cas d'égalité), que

$$\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var } X}} = P_F \left(\frac{X - \mathbb{E}(X)}{\sqrt{\text{Var } X}} \right) = \rho_{XY} \frac{Y - \mathbb{E}(Y)}{\sqrt{\text{Var } Y}}.$$

Par conséquent, si $\rho_{XY} = 1$ (respectivement $\rho_{XY} = -1$), alors il existe $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ tel que $Y = aX + b$ (respectivement $Y = -aX + b$).

Réciproquement, si il existe $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ tel que $Y = aX + b$, alors $\text{Cov}(X, Y) = a\text{Var}(Y)$ et $\text{Var}(X) = a^2\text{Var}(Y)$, donc

$$\rho_{XY} = \frac{a}{|a|},$$

qui vaut 1 si $a > 0$ et -1 si $a < 0$. ■

Loi de la v.a. X	Espérance	Variance
Loi binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$	np	$np(1 - p)$
Loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	λ	λ
Loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0, 1[$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
Loi Uniforme $\mathcal{U}([a, b])$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Loi Exponentielle $\mathcal{E}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	λ	λ^2
Loi Gaussienne $\mathcal{N}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$	m	σ^2

TABLE 7.1 – Espérance et variance de quelques lois classiques

Chapitre 8

Suites de variables aléatoires

Dans ce chapitre, nous nous intéressons au comportement asymptotique d'une suite de variables aléatoires définies sur un même espace $(\Omega, \mathcal{F}, \mathbb{P})$. Ces variables aléatoires sont supposées à valeurs dans $\mathbb{R}^d$, $d \geq 1$.

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités sur lequel seront définies toutes les variables aléatoires.
- Tout espace $E \subset \mathbb{R}^d$ est muni de sa tribu borélienne $\mathcal{B}(E)$. De plus, $\mathbb{R}^d$ est muni de la norme euclidienne notée $\|\cdot\|$.

8.1 Convergence presque sûre

8.1.1 Définition et propriétés

La notion de convergence presque sûre d'une suite de variables aléatoires relaxe celle de convergence simple en tant que suite de fonctions. Elle a déjà été rencontrée dans les théorèmes de convergence monotone ou dominée (cf. chapitre 3).

Définition 8.1 (Convergence presque sûre).

La suite de variables $(X_n)_{n \in \mathbb{N}}$ (à valeurs dans $\mathbb{R}^d$) **converge presque sûrement vers la variable aléatoire X** , ce que l'on note $\mathbf{X}_n \xrightarrow{\text{p.s.}} X$, si il existe N un ensemble négligeable tel que pour tout $\omega \in \Omega \setminus N$, la suite $(X_n(\omega))_{n \in \mathbb{N}}$ converge vers $X(\omega)$, c'est-à-dire tel que

$$\forall \omega \in \Omega \setminus N, \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega).$$

Exemple 8.1. Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Pour tout $n \in \mathbb{N}$,

$$X_n = \mathbf{1}_{[0, 1/n]}(X)$$

est une variable aléatoire à valeurs réelles. Par ailleurs, si $X(\omega) \neq 0$,

$$\lim_{n \rightarrow +\infty} X_n(\omega) = 0.$$

Or $\mathbb{P}(X \neq 0) = 1$ car X suit une loi uniforme sur $[0, 1]$. Par conséquent, $X_n \xrightarrow{\text{p.s.}} 0$.

Remarque 8.1. La convergence presque sûre possède les mêmes propriétés de stabilité que la convergence simple de fonctions. En particulier, si l'on considère $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ et $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^{d'}$ telles que

$$X_n \xrightarrow{\text{p.s.}} X \quad \text{et que} \quad Y_n \xrightarrow{\text{p.s.}} Y,$$

où X est une variable aléatoire à valeurs dans $\mathbb{R}^d$ et Y une variable aléatoire à valeurs dans $\mathbb{R}^{d'}$, alors le lecteur pourra vérifier les 3 assertions suivantes.

1. Si $d = d'$, alors pour tous $\lambda, \mu \in \mathbb{R}$, $\lambda X_n + \mu Y_n \xrightarrow{\text{p.s.}} \lambda X + \mu Y$.
2. Si $d' = 1$, alors $Y_n X_n \xrightarrow{\text{p.s.}} Y X$.
3. Si $d' = 1$ et si pour tout $n \in \mathbb{N}$, $\mathbb{P}(Y_n \neq 0) = \mathbb{P}(Y \neq 0) = 1$, alors $\frac{1}{Y_n} \xrightarrow{\text{p.s.}} \frac{1}{Y}$.

Les résultats de stabilité énoncés dans la remarque précédente peuvent être vu comme des conséquences des propriétés suivantes.

Proposition 8.1.

Considérons $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans $\mathbb{R}^d$ ainsi que $(Y_n)_{n \in \mathbb{N}}$ et Y des variables aléatoires à valeurs dans $\mathbb{R}^{d'}$.

1. *La suite de vecteurs aléatoires $((X_n, Y_n))_{n \in \mathbb{N}}$ à valeurs dans $\mathbb{R}^d \times \mathbb{R}^{d'}$ converge presque sûrement vers le vecteur aléatoire (X, Y) si et seulement si*

$$X_n \xrightarrow{\text{p.s.}} X \quad \text{et} \quad Y_n \xrightarrow{\text{p.s.}} Y.$$

2. *Supposons que les variables X_n , $n \in \mathbb{N}$ et X sont à valeurs presque sûrement dans $E \subset \mathbb{R}^d$. Alors si $X_n \xrightarrow{\text{p.s.}} X$ et si $f : E \rightarrow \mathbb{R}^{d'}$ est une fonction continue sur E , alors*

$$f(X_n) \xrightarrow{\text{p.s.}} f(X).$$

Démonstration.

1. La démonstration de l'assertion 1. est laissée en guise d'exercice.
2. Les variables X_n , $n \in \mathbb{N}$ et X étant à valeurs presque sûrement dans E , et étant en nombre dénombrable, il existe un négligeable N_1 tel que si $\omega \notin N_1$,

$$\forall n \in \mathbb{N}, X_n(\omega) \in E \quad \text{et} \quad X(\omega) \in E.$$

De plus, $X_n \xrightarrow{\text{p.s.}} X$ signifie qu'il existe un négligeable N_2 tel que

$$\forall \omega \notin N_2, \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega).$$

La fonction $f : E \rightarrow \mathbb{R}^{d'}$ étant continue, nous avons alors :

$$\forall \omega \notin (N_1 \cup N_2), \lim_{n \rightarrow +\infty} f(X_n(\omega)) = f(X(\omega)).$$

Les ensembles N_1 et N_2 étant négligeables, $N_1 \cup N_2$ l'est aussi, et donc $f(X_n) \xrightarrow{\text{p.s.}} f(X)$. ■

Énonçons à présent un critère, conséquence du lemme de Borel-Cantelli, pour établir la convergence presque sûre d'une suite de variables aléatoires. Pour $x \in \mathbb{R}^d$, $|x|$ désigne la norme euclidienne de x .

Corollaire 8.2.

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

1. Si pour tout $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty,$$

alors la suite $X_n \xrightarrow{p.s.} X$.

2. Si les variables aléatoires X_n , $n \in \mathbb{N}$, sont mutuellement indépendantes (cf. définition 6.2, page 63) alors $X_n \xrightarrow{p.s.} 0$ si et seulement si

$$\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) < +\infty.$$

Démonstration.

1. Supposons que pour tout $\varepsilon > 0$,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty,$$

Fixons $p \in \mathbb{N}^*$ et pour tout $n \in \mathbb{N}$, posons

$$A_{n,p} = \left\{ |X_n - X| \geq \frac{1}{p} \right\}.$$

Les variables X_n et X étant des fonctions mesurables, $A_{n,p} \in \mathcal{F}$ et par hypothèse,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_{n,p}) = \sum_{n \in \mathbb{N}} \mathbb{P}\left(|X_n - X| \geq \frac{1}{p}\right) < +\infty.$$

Par suite, d'après le lemme de Borel-Cantelli appliquée à la suite $(A_{n,p})_{n \in \mathbb{N}}$,

$$\mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_{n,p}\right) = 0.$$

Ceci étant vrai pour tout entier $p \in \mathbb{N}^*$, et la mesure $\mathbb{P}$ étant positive σ -sous-additive (cf. chapitre 1, proposition 1.1),

$$0 \leq \mathbb{P}\left(\bigcup_{p \in \mathbb{N}^*} \limsup_{n \rightarrow +\infty} A_{n,p}\right) \leq \sum_{p \in \mathbb{N}^*} \mathbb{P}\left(\limsup_{n \rightarrow +\infty} A_{n,p}\right) = 0.$$

Par suite,

$$\mathbb{P}\left(\bigcup_{p \in \mathbb{N}^*} \limsup_{n \rightarrow +\infty} A_{n,p}\right) = 0,$$

c'est-à-dire que pour presque tout ω , pour tout $p \in \mathbb{N}^*$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$,

$$|X_n(\omega) - X(\omega)| < \frac{1}{p}.$$

Autrement dit, la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X .

2. Supposons que les variables aléatoires X_n , $n \in \mathbb{N}$, sont indépendantes. Pour tout $n \in \mathbb{N}$, posons $A_n = \{|X_n| \geq \varepsilon\}$. Nous rappelons que $A_n \in \mathcal{F}$, car X_n est mesurable.

- D'après l'assertion 1.,

$$\left(\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) < +\infty \right) \implies X_n \xrightarrow{\text{p.s.}} 0.$$

- Pour établir la réciproque, nous montrons sa contraposée. Nous supposons donc qu'il existe $\varepsilon > 0$ tel que

$$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = +\infty.$$

Les variables aléatoires X_n , $n \in \mathbb{N}$, étant indépendantes, les événements $A_n = \{|X_n| \geq \varepsilon\}$, $n \in \mathbb{N}$, sont aussi indépendants. Dès lors, d'après le lemme de Borel-Cantelli,

$$\mathbb{P}\left(\limsup_{n \in \mathbb{N}} A_n\right) = 1.$$

Cela signifie que, pour presque tout ω , il existe une infinité d'indices $n \in \mathbb{N}$ tels que

$$|X_n(\omega)| > \varepsilon.$$

Par conséquent, pour presque tout ω , $(X_n(\omega))_{n \in \mathbb{N}}$ ne converge pas vers 0. ■

Nous pouvons déduire du corollaire précédent un critère basé sur la convergence d'une série de moments d'ordre p .

Corollaire 8.3 (Critère de convergence p.s. basé sur les moments).

Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans E . Si il existe $p \in \mathbb{R}_+^*$ tel que

$$\sum_{n \in \mathbb{N}} \mathbb{E}(|X_n|^p) < +\infty,$$

alors $X_n \xrightarrow{\text{p.s.}} 0$.

Démonstration Étant donné que $p \in \mathbb{R}_+^*$,

$$\forall \varepsilon > 0, \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(|X_n|^p > \varepsilon^p).$$

Alors, d'après l'inégalité de Markov,

$$\forall \varepsilon > 0, \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(|X_n|^p > \varepsilon^p) \leq \frac{\mathbb{E}(|X_n|^p)}{\varepsilon^p}.$$

La série de terme général $\mathbb{E}(|X_n|^p)$ étant convergente, on en déduit que

$$\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) < +\infty,$$

ce qui implique que $X_n \xrightarrow{\text{p.s.}} 0$ d'après le corollaire 8.2. ■

Exemple 8.2. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires gaussiennes centrées (c'est-à-dire de moyenne nulle) et de variances respectives $1/n$. On a alors

$$\mathbb{E}(|X|^2) = \frac{1}{n^2}.$$

Étant donné que la série de terme général $1/n^2$ est sommable, on déduit du corollaire précédent que la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers 0.

8.2 Convergence en probabilité

La convergence en probabilité est une notion plus faible que la convergence presque sûre : elle contrôle l'écart en probabilité entre X_n et sa limite.

Définition 8.2 (Convergence en probabilité).

La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge en probabilité vers la variable aléatoire X** , ce que l'on note $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}$, si, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Exemple 8.3. Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Pour tout $n \in \mathbb{N}^*$, posons

$$X_n = X + n^2 \mathbf{1}_{X \leq 1/n}.$$

Pour tout $n \in \mathbb{N}^*$, X_n est mesurable car X l'est. De plus,

$$\forall n \in \mathbb{N}^*, \mathbb{P}(|X - X_n| > \varepsilon) = \mathbb{P}(n^2 \mathbf{1}_{X \leq 1/n} > \varepsilon) = \begin{cases} 0 & \text{si } \varepsilon \geq n^2 \\ \mathbb{P}(X \leq 1/n) & \text{si } \varepsilon < n^2. \end{cases}$$

Étant donné que X suit une loi uniforme sur $[0, 1]$,

$$\forall n \in \mathbb{N}^*, \mathbb{P}(|X - X_n| > \varepsilon) = \begin{cases} 0 & \text{si } \varepsilon \geq n^2 \\ \frac{1}{n} & \text{si } \varepsilon < n^2. \end{cases}$$

Par conséquent,

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

et donc $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X .

Commençons tout d'abord par nous intéresser à l'unicité de la limite en probabilité d'une suite, unicité a priori non évidente en lisant la définition de la convergence.

Proposition 8.4 (Unicité p.s. de la limite).

Si la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la variable aléatoire X et vers la variable aléatoire Y , alors

$$X = Y \quad \text{presque sûrement.}$$

Démonstration. Pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - Y| > \varepsilon/2)$$

car

$$\{|X - Y| > \varepsilon\} \subset \{|X_n - X| > \varepsilon/2\} \cup \{|X_n - Y| > \varepsilon/2\}.$$

Alors, vu que $X_n \xrightarrow{\mathbb{P}} X$ et que $X_n \xrightarrow{\mathbb{P}} Y$,

$$\forall \varepsilon > 0, \mathbb{P}(|X - Y| > \varepsilon) = 0.$$

D'après la continuité monotone de $\mathbb{P}$ appliquée aux ensembles $A_p = \{|X - Y| > 1/p\}$, $p \in \mathbb{N}^*$,

$$\mathbb{P}(|X - Y| > 0) = \lim_{p \rightarrow +\infty} \mathbb{P}(\{|X - Y| > 1/p\}) = 0.$$

Les variables aléatoires X et Y sont donc égales presque sûrement. ■

Avant d'étudier les liens entre la convergence en probabilité et la convergence presque sûre, mentionnons que les propriétés de stabilité énoncées dans la proposition 8.1 pour la convergence presque sûre restent vraies pour la convergence en probabilité.

Proposition 8.5.

Considérons $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans $\mathbb{R}^d$ ainsi que $(Y_n)_{n \in \mathbb{N}}$ et Y des variables aléatoires à valeurs dans $\mathbb{R}^{d'}$.

1. *Alors, $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$ si et seulement si*

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{et} \quad Y_n \xrightarrow{\mathbb{P}} Y.$$

2. *Supposons que les variables X_n , $n \in \mathbb{N}$ et X sont à valeurs presque sûrement dans $E \subset \mathbb{R}^d$. Alors si $X_n \xrightarrow{\mathbb{P}} X$ et si $f : E \rightarrow \mathbb{R}^{d'}$ est une fonction continue sur E , alors*

$$f(X_n) \xrightarrow{\mathbb{P}} f(X).$$

Terminons à présent par les liens existants entre la convergence presque sûre et la convergence en probabilité.

Proposition 8.6 (Liens entre convergences p.s. et en probabilité).

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

1. *Si $X_n \xrightarrow{\text{p.s.}} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.*

2. *Si $X_n \xrightarrow{\mathbb{P}} X$, alors il existe une suite déterministe et strictement croissante d'entiers $(n_k)_{k \in \mathbb{N}}$ telle que la sous-suite $(X_{n_k})_{k \in \mathbb{N}}$ converge presque sûrement vers la variable aléatoire X .*

Démonstration.

1. Soit $\varepsilon > 0$. Supposons que $X_n \xrightarrow{\text{p.s.}} X$. Pour tout $n \in \mathbb{N}$,

$$Y_n = \mathbf{1}_{|X - X_n| > \varepsilon}$$

est une variable aléatoire à valeurs dans $\mathbb{R}$ et bornée par 1. De plus, comme $X_n \xrightarrow{\text{p.s.}} X$, $Y_n \xrightarrow{\text{p.s.}} 0$. Alors, d'après le théorème de convergence dominée,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X - X_n| > \varepsilon) = \lim_{n \rightarrow +\infty} \mathbb{E}(Y_n) = 0$$

et ce pour tout $\varepsilon > 0$. Par conséquent, $X_n \xrightarrow{\mathbb{P}} X$.

1. Supposons que $X_n \xrightarrow{\mathbb{P}} X$. En particulier, pour tout $k \in \mathbb{N}^*$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X - X_n| > 1/k) = 0.$$

On peut donc définir la suite d'entiers $(n_k)_{k \in \mathbb{N}}$ par $n_0 = 0$ et, pour tout $k \geq 1$,

$$n_k = \min \left\{ n \geq n_{k-1} + 1 / \mathbb{P}(|X - X_n| > 1/k) \leq 2^{-k} \right\}.$$

Par construction, la suite $(n_k)_{k \in \mathbb{N}}$ est une suite strictement croissante d'entiers telle que

$$\forall k \in \mathbb{N}^*, \mathbb{P}(|X - X_{n_k}| > 1/k) \leq 2^{-k}.$$

Fixons $\varepsilon > 0$. Alors, il existe $k_0 \in \mathbb{N}^*$ tel que pour tout entier $k \geq k_0$, $1/k < \varepsilon$. Notons qu'alors pour tout entier $k \geq k_0$,

$$\{|X - X_{n_k}| \geq \varepsilon\} \subset \{|X - X_{n_k}| > 1/k\}.$$

Par conséquent,

$$\sum_{k=k_0}^{+\infty} \mathbb{P}(|X - X_{n_k}| \geq \varepsilon) \leq \sum_{k=k_0}^{+\infty} \mathbb{P}(|X - X_{n_k}| > 1/k) \leq \sum_{k=k_0}^{+\infty} 2^{-k} < +\infty.$$

Étant donné que pour tout $k \leq k_0$, $\mathbb{P}(|X - X_{n_k}| \geq \varepsilon) \leq 1$, l'inégalité précédente implique que

$$\sum_{k \in \mathbb{N}} \mathbb{P}(|X - X_{n_k}| \geq \varepsilon) < +\infty.$$

Alors, d'après le corollaire 8.2, la sous-suite $(X_{n_k})_{k \in \mathbb{N}}$ converge presque sûrement vers X . ■

Remarque 8.2. En général, la convergence en probabilité n'implique pas la convergence presque sûre. Pour illustrer ce point, considérons une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ indépendantes et supposons que la loi de X_n est la loi de Bernoulli de paramètre $1/n$. Rappelons donc que la loi de X_n est

$$\mathbb{P}_{X_n} = \left(1 - \frac{1}{n}\right) \delta_0 + \frac{1}{n} \delta_1.$$

Alors, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n| > \varepsilon) = \begin{cases} 0 & \text{si } \varepsilon \geq 1 \\ \frac{1}{n} & \text{si } \varepsilon < 1 \end{cases}$$

Par suite, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n| > \varepsilon) = 0.$$

Autrement dit, $X_n \xrightarrow{\mathbb{P}} X = 0$. D'autre part, pour $0 < \varepsilon < 1$

$$\sum_{n \in \mathbb{N}^*} \mathbb{P}(|X_n| > \varepsilon) = \sum_{n \in \mathbb{N}^*} \frac{1}{n} = +\infty.$$

Les variables aléatoires X_n , $n \in \mathbb{N}^*$ étant indépendantes, vu l'égalité ci-dessus, d'après le lemme de Borel-Cantelli, la suite $(X_n)_{n \geq 1}$ ne converge pas presque sûrement vers $X = 0$.

8.3 Convergence L^p

Soit $p \geq 1$. Dans le chapitre 7, nous avons vu la construction des espaces L^p pour les variables aléatoires à valeurs réelles. Dans la présente section, nous généralisons cette notion aux variables aléatoires à valeurs dans $\mathbb{R}^d$ en posant

$$L^p_{\mathbb{R}^d}(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X : \Omega \rightarrow \mathbb{R}^d \text{ mesurable} / |X|^p \text{ est intégrable} \right\},$$

où $|X|$ désigne la norme euclidienne de X . En absence d'ambiguïté sur la dimension d , nous omettrons la dépendance en $\mathbb{R}^d$ et nous noterons donc $L^p(\Omega, \mathcal{F}, \mathbb{P})$ cet espace. De plus, nous le munissons de la norme $\|\cdot\|_p$ définie par

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

Définition 8.3 (Convergence dans L^p).

Soit $p \in [1, +\infty[$. Supposons que pour tout $n \in \mathbb{N}$, $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ **converge dans L^p vers la variable aléatoire X** , ce que l'on note $X_n \xrightarrow{L^p} X$, si

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0,$$

c'est-à-dire si $\lim_{n \rightarrow +\infty} \|X_n - X\|_p = 0$.

Remarque 8.3. Le lecteur pourra vérifier que $X_n \xrightarrow{L^p} X$ si et seulement si chaque coordonnée $X_n^{(i)}$ de X_n converge dans $L^p_{\mathbb{R}}(\Omega, \mathcal{F}, \mathbb{P})$ vers la coordonnée $X^{(i)}$ du vecteur aléatoire X . Il en déduira aussi que l'assertion 1. de la proposition 8.1 (respectivement de la proposition 8.5) est vraie en remplaçant la convergence presque sûre (respectivement en probabilité) par la convergence dans L^p .

Notons que la convergence au sens L^p implique celle du moment d'ordre p .

Proposition 8.7.

Soit $p \in [1, +\infty[$. Soient $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans $\mathbb{R}^d$. Supposons que pour tout $n \in \mathbb{N}$, $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. Si $X_n \xrightarrow{L^p} X$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n|^p) = \mathbb{E}(|X|^p).$$

De plus, si $X_n \xrightarrow{L^1} X$, alors $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Démonstration. La première convergence est une simple conséquence de la continuité de la norme $\|\cdot\|_p$ sur l'ensemble L^p . La seconde convergence provient de l'inégalité $|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}(|X_n - X|)$. ■

Comparons à présent la convergence dans L^p avec la convergence dans L^q , pour $p \leq q$.

Proposition 8.8 (Comparaison des convergences dans L^p et L^q).

Soient $p, q \in [1, +\infty[$ tels que $p \leq q$. Soient $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans $\mathbb{R}^d$. Supposons que, pour tout $n \in \mathbb{N}$, $X_n \in L^q(\Omega, \mathcal{F}, \mathbb{P})$ et $X \in L^q(\Omega, \mathcal{F}, \mathbb{P})$. Si $X_n \xrightarrow{L^q} X$, alors $X_n \xrightarrow{L^p} X$.

Démonstration. Soient $p, q \in [1, +\infty[$ tels que $p \leq q$. Alors $L^p(\Omega, \mathcal{F}, \mathbb{P}) \subset L^q(\Omega, \mathcal{F}, \mathbb{P})$, donc, sous nos hypothèses,

$$X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{et} \quad X \in L^p(\Omega, \mathcal{F}, \mathbb{P}).$$

D'après le corollaire 7.6 du chapitre 7,

$$(\mathbb{E}(|X_n - X|^p))^{1/p} \leq \mathbb{E}(|X_n - X|^q)^{1/q}.$$

Par conséquent, si $(X_n)_{n \in \mathbb{N}}$ converge dans L^q vers X , c'est-à-dire si

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^q)^{1/q} = 0$$

alors,

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p)^{1/p} = 0,$$

c'est-à-dire que $(X_n)_{n \in \mathbb{N}}$ converge aussi dans L^p vers X . ■

La proposition suivante lie la convergence dans L^p et la convergence en probabilité.

Proposition 8.9 (Convergence L^p /Convergence en probabilité).

Soit $p \in [1, +\infty[$. Soient $(X_n)_{n \in \mathbb{N}}$ et X des variables aléatoires à valeurs dans E .

1. Supposons que, pour tout $n \in \mathbb{N}$, $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. Si $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
2. Supposons qu'il existe $a \in \mathbb{R}_+$ tel que

$$\forall n \in \mathbb{N}, |X_n| \leq a \quad \text{presque sûrement.}$$

Alors, pour tout $p \geq 1$,

$$X_n \xrightarrow{\mathbb{P}} X \iff X_n \xrightarrow{L^p} X$$

Remarque 8.4. En toute généralité, la convergence en probabilité n'implique pas la convergence dans L^p .

Démonstration.

1. Supposons que $X_n \xrightarrow{L^p} X$ et considérons $\varepsilon > 0$. Étant donné que $p \in \mathbb{R}_+^*$,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p).$$

Alors, d'après l'inégalité de Markov,

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \frac{\mathbb{E}(|X_n - X|^p)}{\varepsilon^p}.$$

La suite $(X_n)_{n \in \mathbb{N}}$ convergeant vers X dans L^p ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

et ce pour tout $\varepsilon > 0$. Par conséquent, la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité.

2. Soit $p \geq 1$. D'après la proposition 8.9, si $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{\mathbb{P}} X$. Réciproquement, supposons que $X_n \xrightarrow{\mathbb{P}} X$. Étant donné que $|X - X_n| \geq |X| - |X_n|$ et que $|X_n| \leq a$ presque sûrement,

$$\mathbb{P}(|X| > a + 1) \leq \mathbb{P}(|X_n - X| > 1).$$

En faisant tendre $n \rightarrow +\infty$, on déduit de la convergence en probabilité de $(X_n)_{n \in \mathbb{N}}$ vers X que

$$\mathbb{P}(|X| > a + 1) = 0.$$

Alors, pour tout $n \in \mathbb{N}$, $|X_n - X| \leq |X_n| + |X| \leq 2a + 1$ presque sûrement et donc

$$\mathbb{E}(|X_n - X|^p \mathbf{1}_{|X_n - X| > \varepsilon}) \leq (2a + 1)^p \mathbb{P}(|X_n - X| > \varepsilon).$$

Par suite, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}(|X_n - X|^p) &= \mathbb{E}(|X_n - X|^p \mathbf{1}_{|X_n - X| > \varepsilon}) + \mathbb{E}(|X_n - X|^p \mathbf{1}_{|X_n - X| \leq \varepsilon}) \\ &\leq (2a + 1)^p \mathbb{P}(|X_n - X| > \varepsilon) + \varepsilon^p \end{aligned}$$

Ainsi, pour tout $\varepsilon > 0$,

$$\limsup_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) \leq \varepsilon^p$$

car $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. Étant donné que $\varepsilon > 0$ est arbitraire et que pour tout n , $\mathbb{E}(|X_n - X|^p) \geq 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = \limsup_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0,$$

c'est-à-dire que $X_n \xrightarrow{L^p} X$. ■

Terminons cette partie en donnant le lien entre la convergence dans L^p et la convergence presque sûre.

Proposition 8.10 (Convergence p.s./Convergence L^p).

Soient $p \in [1, +\infty[$, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires dans L^p et X une variable aléatoire. Si

(i) $X_n \xrightarrow{p.s.} X$

(ii) et si il existe $Z \in L^p$ une variable aléatoire à valeurs dans $[0, +\infty]$ telle que

$$\forall n \in \mathbb{N}, |X_n| \leq Z \quad \text{presque sûrement,}$$

alors $X \in L^p$ et $X_n \xrightarrow{L^p} X$.

Démonstration. Étant donné que $|X_n| \leq Z$ presque sûrement pour tout n et que $X_n \xrightarrow{p.s.} X$, nous avons :

$$|X| \leq Z \quad \text{presque sûrement}$$

Par conséquent, $\mathbb{E}(|X|^p) \leq \mathbb{E}(Z^p) < +\infty$ et donc $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. De plus, pour tout $n \in \mathbb{N}$,

$$|X_n - X|^p \leq (|X_n| + |X|)^p \leq 2^p Z^p \quad \text{presque sûrement}$$

Or $2^p Z^p$ est une variable aléatoire intégrable car $Z \in L^p$, donc, d'après le théorème de convergence dominée appliqué à la suite de variables aléatoires $(|X_n - X|^p)_{n \in \mathbb{N}}$ qui converge presque sûrement vers 0, nous avons :

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0.$$

Par suite, $X_n \xrightarrow{L^p} X$. ■

8.4 Convergence en loi

8.4.1 Définition et premières propriétés

La convergence en loi est la convergence la plus faible présentée dans ce chapitre. Plutôt que d'étudier l'écart entre X_n et une variable X en tant que fonction, elle s'intéresse à l'écart entre leur loi.

Définition 8.4 (Convergence en loi).

La suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ à valeurs dans $\mathbb{R}^d$ **converge en loi vers la variable aléatoire** X à valeurs dans $\mathbb{R}^d$, ce que l'on note $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ continue bornée,

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

Remarque 8.5.

Il y a unicité de la loi limite mais pas de la variable aléatoire limite. Plus précisément, si la suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire X alors elle converge aussi en loi vers la variable aléatoire Y si et seulement si X et Y ont même loi. La convergence en loi n'affirme pas que X et X_n sont proches dans E mais que la loi de X_n est proche de celle de X pour n grand.

Exemple 8.4. Soit X_n une variable aléatoire de loi $\mathbb{P}_{X_n} = \frac{1}{2}\delta_{1/n} + \frac{1}{2}\delta_{n/(n+1)}$. Alors, pour toute fonction f continue bornée sur $\mathbb{R}$,

$$\mathbb{E}(f(X_n)) = \frac{1}{2}f\left(\frac{1}{n}\right) + \frac{1}{2}f\left(\frac{n}{n+1}\right).$$

Par conséquent, pour toute fonction f continue bornée sur $\mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}(f(X_n)) = \frac{1}{2}f(0) + \frac{1}{2}f(1) = \mathbb{E}(f(X))$$

avec X de loi de Bernoulli de paramètre $p = 1/2$. Par conséquent, $X_n \xrightarrow{\mathcal{L}} X$.

La proposition suivante est une simple conséquence de la définition.

Proposition 8.11 (Image par une fonction continue).

Soient X et $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires à valeurs presque sûrement dans $E \subset \mathbb{R}^d$. Si $X_n \xrightarrow{\mathcal{L}} X$ et si $f : E \rightarrow \mathbb{R}^{d'}$ est une fonction continue, alors

$$f(X_n) \xrightarrow{\mathcal{L}} f(X).$$

Remarque 8.6. Appliquant la proposition ci-dessus, si $X_n \xrightarrow{\mathcal{L}} X$, alors pour tout $1 \leq i \leq d$, la i^{e} coordonnée $X_n^{(i)}$ de X_n converge en loi vers la i^{e} coordonnée $X^{(i)}$ de X . La réciproque est par contre fausse.

La proposition suivante lie la convergence en probabilité et la convergence en loi.

Proposition 8.12 (Convergence en probabilité $\Rightarrow$ Convergence en loi).

Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.

Remarque 8.7. La convergence presque sûre et la convergence L^p ($p \geq 1$) impliquant la convergence en probabilité, elles impliquent aussi la convergence en loi. La convergence en loi est donc la plus faible de toutes les convergences étudiées jusqu'à présent.

Démonstration. Supposons $X_n \xrightarrow{\mathbb{P}} X$. Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction continue bornée. Alors, d'après la proposition 8.5, $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. De plus, f étant bornée, il existe $a \in \mathbb{R}$ tel que

$$\forall n \in \mathbb{N}, |f(X_n)| \leq a \quad \text{partout.}$$

Alors, d'après l'assertion 2. de la proposition 8.9, $f(X_n) \xrightarrow{L^1} f(X)$. Par conséquent, d'après la proposition 8.7,

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

Ceci étant vrai pour toute fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$, nous concluons que $X_n \xrightarrow{\mathcal{L}} X$. ■

8.4.2 Caractérisations

Nous pouvons caractériser la convergence en loi d'une suite de variables aléatoires réelles à partir du comportement asymptotique de la suite de leur fonction de répartition.

Proposition 8.13 (Suite des fonctions de répartition).

Soient $X_n, n \in \mathbb{N}$ et X des variables aléatoires réelles de fonctions de répartition respectives F_{X_n} et F_X .

Alors, $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

pour tout $x \in \mathbb{R}$ point de continuité de F_X .

Remarque 8.8. Dans la proposition précédente, la convergence la suite $(F_{X_n})_{n \in \mathbb{N}}$ des fonctions de répartition n'a a priori pas lieu sur tout $\mathbb{R}$.

La convergence en loi peut aussi se déduire de l'asymptotique d'une suite de fonctions caractéristiques.

Théorème 8.14 (Théorème de Paul Lévy).

Soient $X_n, n \in \mathbb{N}$ et X des variables aléatoires à valeurs dans $\mathbb{R}^d$ de fonctions caractéristiques respectives φ_{X_n} et φ_X . La suite de fonctions $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge simplement vers φ_X si et seulement si $X_n \xrightarrow{\mathcal{L}} X$.

1. Si $X_n \xrightarrow{\mathcal{L}} X$, alors la suite de fonctions $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge simplement vers φ_X .

2. Si $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge simplement vers une fonction φ continue en 0, alors il existe une variable aléatoire Y à valeurs dans $\mathbb{R}^d$ dont la fonction caractéristique est φ et $X_n \xrightarrow{\mathcal{L}} Y$.

En particulier, la réciproque de la première assertion est vraie : si la suite de fonctions $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge simplement vers φ_X , alors $X_n \xrightarrow{\mathcal{L}} X$.

Remarque 8.9. En fait, si $X_n \xrightarrow{\mathcal{L}} X$, alors la suite de fonctions $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge uniformément sur tout compact vers φ_X .

8.5 Somme de variables aléatoires indépendantes

Dans cette section, étant donné une suite de variables aléatoires à valeurs réelles $(X_n)_{n \in \mathbb{N}}$, on pose

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \forall n \geq 1.$$

Des variables aléatoires X_n , $n \in \mathbb{N}$ sont dites *indépendantes identiquement distribuées* (en abrégé *i.i.d.*), si elles sont mutuellement indépendantes et si elles ont toutes même loi.

8.5.1 Loi des grands nombres

Proposition 8.15 (Loi faible des grands nombres).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles de même loi, deux à deux indépendantes et de carré intégrable. Alors

$$\overline{X_n} \xrightarrow[n \rightarrow +\infty]{L^2} \mathbb{E}(X_1).$$

En renforçant les hypothèses, la convergence de $\overline{X_n}$ a lieu aussi presque sûrement.

Proposition 8.16 (Loi forte des grands nombres).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d.

1. Si $\mathbb{E}(|X_1|) < +\infty$, c'est-à-dire si les variables aléatoires X_n sont intégrables, alors

$$\overline{X_n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}(X_1).$$

2. Si $\mathbb{E}(|X_1|) = +\infty$, c'est-à-dire si les variables aléatoires X_n ne sont pas intégrables, alors la suite $(\overline{X_n})_{n \in \mathbb{N}^*}$ diverge presque sûrement.

8.5.2 Théorème central limite

Théorème 8.17 (Théorème central limite).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. de carré intégrable. Considérons Z une variable aléatoire de loi $\mathcal{N}(0, \sigma^2)$ avec $\sigma^2 = \text{Var } X_1$. Alors,

$$\sqrt{n}(\overline{X_n} - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Z.$$

Démonstration. Notons φ_{X_1} la fonction caractéristique de X_1 et φ_{Y_n} celle de $Y_n = \sqrt{n}(\overline{X_n} - \mathbb{E}(X_1))$.

Par indépendance mutuelle des X_k , $k \in \mathbb{N}$, pour tout $t \in \mathbb{R}$,

$$\varphi_{Y_n}(t) = \mathbb{E}\left(e^{it \sum_{k=1}^n X_k / \sqrt{n}}\right) e^{-it\sqrt{n}\mathbb{E}(X_1)} = \prod_{k=1}^n \mathbb{E}\left(e^{itX_k / \sqrt{n}}\right) e^{-it\sqrt{n}\mathbb{E}(X_1)}.$$

Les variables aléatoires X_k , $k \in \mathbb{N}$ ayant même loi, pour tout $t \in \mathbb{R}$,

$$\varphi_{Y_n}(t) = e^{-it\sqrt{n}\mathbb{E}(X_1)} \left(\varphi_{X_1}(t/\sqrt{n})\right)^n.$$

Étant donné que X_1 admet un moment d'ordre 2, φ_{X_1} est $\mathcal{C}^2$ et

$$\varphi_{X_1}(t) = 1 + \varphi'_{X_1}(0)t + \frac{1}{2}\varphi''_{X_1}(0)t^2 + o(t^2) = 1 + i\mathbb{E}(X_1)t - \frac{\mathbb{E}(X_1^2)}{2}t^2 + o(t^2).$$

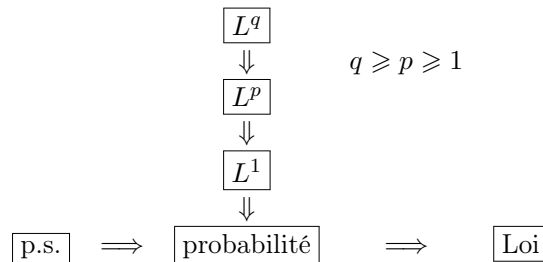
Alors, en utilisant un développement limité à l'ordre 2 de $x \mapsto \ln(1+x)$ en 0, on obtient

$$\begin{aligned} \lim_{n \rightarrow +\infty} \varphi_{Y_n}(t) &= \lim_{n \rightarrow +\infty} e^{-it\sqrt{n}\mathbb{E}(X_1)} \left(1 + \frac{i\mathbb{E}(X_1)t}{\sqrt{n}} - \frac{\mathbb{E}(X_1^2)t^2}{2n} + o(t^2)\right)^n \\ &= \exp\left(-t^2\mathbb{E}(X_1^2)/2 + t^2(\mathbb{E}(X_1))^2/2\right) = \exp(-t^2\text{Var } X_1/2) = \varphi_Z(t) \end{aligned}$$

où $Z \sim \mathcal{N}(0, \text{Var } X_1)$. Alors, d'après le théorème de Paul Lévy, $\sqrt{n}(\overline{X_n} - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Z$. ■

8.6 Tableau récapitulant les liens entre les convergence

Le tableau ci-dessous récapitule les liens entre les divers types de convergence.



Chapitre 9

Vecteurs gaussiens

Les vecteurs aléatoires gaussiens sont très utilisés pour modéliser conjointement plusieurs phénomènes réels considérés comme gaussiens dans leur globalité. Comme dans le cas unidimensionnel (variables aléatoires réelles gaussiennes), utiliser un vecteur gaussien en modélisation est souvent justifié par le fait qu'ils apparaissent comme limites de sommes de vecteurs (théorème central limite). Enfin, travailler sous l'hypothèse de vecteur gaussien est un cadre dans lequel il est souvent plus simple d'établir des propriétés (utilisation de propriétés non vraies a priori sans cette hypothèse, calculs simplifiés se ramenant souvent au calcul de moments d'ordre 1 et 2, etc.).

Hypothèses/Notations pour ce chapitre

- $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités sur lequel seront définies toutes les vecteurs aléatoires.
- Tout espace $E \subset \mathbb{R}^d$ est muni de sa tribu borélienne $\mathcal{B}(E)$.
- Pour toute matrice A , A' représente la transposée de A .
- Tout élément $x \in \mathbb{R}^d$ est noté sous la forme d'un vecteur colonne, c'est-à-dire que $x = (x_1, \dots, x_d)'$.
- **Attention !** Dans la suite du cours, une variable aléatoire réelle égale à une constante $a \in \mathbb{R}$ presque sûrement est dite **gaussienne de variance nulle et de moyenne a** . Sa loi, qui est la masse de Dirac δ_a , est encore notée $\mathcal{N}(a, 0)$.

9.1 Définition et première propriétés

9.1.1 Définition

La définition d'un vecteur gaussien généralise la notion de variable aléatoire réelle gaussienne. Remarquons que, dans cette définition, nous avons besoin de considérer que les variable aléatoires constantes presque sûrement sont gaussiennes.

Définition 9.1 (Vecteurs gaussiens).

Un vecteur aléatoire $X = (X_1, \dots, X_d)'$ à valeurs dans $\mathbb{R}^d$ est dit **gaussien** si toute combinaison linéaire de ses coordonnées est une variable gaussienne réelle, c'est-à-dire si pour tout $a = (a_1, \dots, a_d)' \in \mathbb{R}^d$, la variable aléatoire $\sum_{i=1}^d a_i X_i$ est une variable réelle gaussienne.

Remarque 9.1. Un vecteur gaussien à valeurs dans $\mathbb{R}$ est simplement une variable aléatoire réelle gaussienne.

Comme nous le verrons plus tard, la loi d'un vecteur gaussien dépend de deux quantités importantes : son vecteur moyenne et sa matrice de covariance, définis ci-après.

Définition 9.2.

Soit $X = (X_1, \dots, X_d)'$ un vecteur gaussien à valeurs dans $\mathbb{R}^d$. Alors toute coordonnée X_i , $1 \leq i \leq d$, est une variable aléatoire gaussienne. En particulier, chacune d'elle admet un moment d'ordre 2 (et donc d'ordre 1). Par suite, on peut définir son **vecteur moyenne**

$$m_X = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))'$$

et sa **matrice de covariance**

$$\Gamma_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}.$$

Exemple 9.1.

1. Si U est une variable aléatoire gaussienne centrée réduite $\mathcal{N}(0, 1)$, alors le vecteur $X = (U, U)'$ est un vecteur aléatoire gaussien, de vecteur moyenne et matrice de covariance

$$m_X = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{et} \quad \Gamma_X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

2. Si $X_1, \dots, X_d$ sont des variables aléatoires réelles gaussiennes centrées réduites (soit de loi $\mathcal{N}(0, 1)$) et mutuellement indépendantes, alors le vecteur $X = (X_1, \dots, X_d)'$ est un vecteur aléatoire gaussien, de vecteur moyenne et matrice de covariance

$$m_X = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{et} \quad \Gamma_X = Id_d = \begin{pmatrix} 1 & & (0) \\ & \ddots & \\ (0) & & 1 \end{pmatrix}.$$

Attention ! Les marginales d'un vecteur gaussien peuvent être gaussiennes sans que, pour autant, le vecteur X soit gaussien. Ainsi, demander qu'un vecteur X soit gaussien est plus fort que de demander que toutes ses coordonnées soient des variables aléatoires réelles gaussiennes. Un contre-exemple sera abordé en TD.

9.1.2 Fonction caractéristique d'un vecteur gaussien

La proposition suivante est très utile puisqu'elle permet de caractériser la loi d'un vecteur gaussien à l'aide de sa fonction caractéristique. Avant de l'énoncer, rappelons la notation du produit scalaire sur $\mathbb{R}^d$: pour tous vecteurs $u, v \in \mathbb{R}^d$,

$$\langle u, v \rangle := \sum_{i=1}^d u_i v_i.$$

Proposition 9.1.

Si X est un vecteur gaussien à valeurs dans $\mathbb{R}^d$ de vecteur moyenne m_X et de matrice de covariance Γ_X , alors sa fonction caractéristique φ_X est donnée sur $\mathbb{R}^d$ par

$$\forall u \in \mathbb{R}^d, \varphi_X(u) = \mathbb{E} \left(e^{i \langle u, X \rangle} \right) = \exp \left(i \langle u, m_X \rangle - \frac{u' \Gamma_X u}{2} \right).$$

Démonstration.

Le vecteur X étant gaussien, la variable aléatoire réelle $\langle u, X \rangle = \sum_{i=1}^d u_i X_i$ est gaussienne. De plus, sa moyenne et sa variance sont données par

$$\mathbb{E}(\langle u, X \rangle) = \mathbb{E} \left(\sum_{i=1}^d u_i X_i \right) = \sum_{i=1}^d u_i \mathbb{E}(X_i) = \langle u, m_X \rangle$$

et, par bilinéarité de la fonction covariance,

$$\text{Var}(\langle u, X \rangle) = \text{Var} \left(\sum_{i=1}^d u_i X_i \right) = \sum_{1 \leq i, j \leq d} \text{Cov}(u_i X_i, u_j X_j) = \sum_{1 \leq i, j \leq d} u_i \text{Cov}(X_i, X_j) u_j = u' \Gamma_X u,$$

par définition de Γ_X . Par conséquent, la fonction caractéristique de $\langle u, X \rangle$ est la fonction caractéristique d'une variable aléatoire gaussienne réelle (cf. chapitre 4, Tableau 4.1 page 48) donnée par

$$t \in \mathbb{R} \mapsto \mathbb{E} \left(e^{it \langle u, X \rangle} \right) = \exp \left(it \langle u, m_X \rangle - t^2 \frac{u' \Gamma_X u}{2} \right).$$

En appliquant ce dernier résultat à $t = 1$, on obtient que pour tout $u \in \mathbb{R}^d$,

$$\varphi_X(u) = \mathbb{E} \left(e^{i \langle u, X \rangle} \right) = \exp \left(i \langle u, m_X \rangle - \frac{u' \Gamma_X u}{2} \right)$$

ce qu'il fallait démontrer. ■

La fonction caractéristique caractérisant la loi d'une variable aléatoire, nous déduisons de la proposition précédente le résultat suivant.

Corollaire 9.2.

La loi d'un vecteur gaussien est caractérisée par son vecteur moyenne et sa matrice de covariance.

Cette caractérisation justifie la définition suivante.

Définition 9.3.

La loi d'un vecteur gaussien à valeurs dans $\mathbb{R}_d$ de vecteur moyenne m et de matrice de covariance Γ est notée $\mathcal{N}_d(m, \Gamma)$.

9.2 Construction et existence des vecteurs gaussiens

Nous avons déjà rencontré un premier exemple non trivial de vecteur gaussien (cf. Exemple 9.1) : un vecteur aléatoire dont les composantes sont des variables réelles i.i.d. de loi gaussienne centrée réduite est un vecteur gaussien de matrice de covariance égale à l'identité. La proposition suivante donne un moyen pour construire d'autres vecteurs gaussiens, à partir de celui donné dans l'exemple 9.1 par exemple. La démonstration de cette proposition est laissée en exercice.

Proposition 9.3.

Soit $X = (X_1, \dots, X_d)'$ un vecteur gaussien de moyenne m_X et de matrice de covariance Γ_X . Soit $Y = AX + b$, où A est une matrice de taille $d' \times d$ et b est un vecteur colonne de $\mathbb{R}^{d'}$. Alors Y est un vecteur gaussien à valeurs dans $\mathbb{R}^{d'}$, de vecteur moyenne m_Y et de matrice de covariance Γ_Y donnés par

$$m_Y = Am_X + b \text{ et } \Gamma_Y = A\Gamma_X A'.$$

Exemple 9.2. Soient $X_1, \dots, X_d$ des variables aléatoires réelles i.i.d. de loi gaussienne centrée réduite. Alors $X = (X_1, \dots, X_d)'$ est un vecteur aléatoire de loi $\mathcal{N}(0, Id_d)$ (c'est-à-dire un vecteur gaussien de moyenne nulle et de matrice de covariance égale à la matrice identité Id_d de taille $d \times d$). Par conséquent, $Y = AX$ est un vecteur aléatoire de loi $\mathcal{N}_{d'}(0, AA')$ (c'est-à-dire un vecteur gaussien de moyenne $m_Y = 0$ et matrice de covariance $\Gamma_Y = AA'$).

Cette proposition nous permet en particulier d'établir une condition nécessaire et suffisante pour l'existence de la loi $\mathcal{N}(m, \Gamma)$. Avant d'énoncer cette nouvelle propriété, rappelons ce que l'on appelle une matrice réelle symétrique positive.

Définition 9.4.

Soit Γ une matrice symétrique réelle de taille $d \times d$. Elle est alors dite **positive** si l'une des deux propriétés suivantes sont vérifiées (nous admettrons que ces deux propriétés sont équivalentes) :

- (i) les valeurs propres de Γ sont toutes positives (cf. Exemple 9.5) ;
- (ii) la forme quadratique

$$\begin{aligned} Q_\Gamma : \mathbb{R}^d &\longrightarrow \mathbb{R} \\ x &\longmapsto x'\Gamma x = \sum_{i=1}^d \sum_{j=1}^d \Gamma_{ij} x_i x_j. \end{aligned}$$

est positive, c'est-à-dire qu'elle est à valeurs dans $\mathbb{R}_+$ (cf. Exemple 9.6).

Exemple 9.3. Soit $X = (X_1, \dots, X_d)'$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Alors sa matrice de covariance Γ_X est réelle symétrique positive. En effet, pour tous $i, j \in \{1, \dots, d\}$, par symétrie de la covariance,

$$[\Gamma_X]_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = [\Gamma_X]_{ji}$$

et, pour tout $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$, nous obtenons par bilinéarité de la covariance,

$$x'\Gamma_X x = \sum_{i=1}^d \sum_{j=1}^d \text{Cov}(X_i, X_j) x_i x_j = \text{Cov}\left(\sum_{i=1}^d x_i X_i, \sum_{j=1}^d x_j X_j\right) = \text{Var}\left(\sum_{i=1}^d x_i X_i\right) \geq 0,$$

car la variance d'une variable aléatoire réelle est positive. Ainsi, toute matrice de covariance est symétrique positive.

Proposition 9.4.

Soient $m \in \mathbb{R}^d$ et Γ une matrice réelle de taille $d \times d$. Alors il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et un vecteur gaussien défini sur $(\Omega, \mathcal{F}, \mathbb{P})$ de vecteur moyenne m et de matrice de covariance Γ si et seulement si la matrice Γ est symétrique positive.

Démonstration.

- Si il existe un vecteur gaussien de matrice de covariance Γ , alors Γ est symétrique positive (comme toute matrice de covariance d'après l'exemple 9.3).
- Supposons à présent que Γ est symétrique positive et montrons qu'il est possible de construire un vecteur gaussien de matrice de covariance Γ . Pour cela, on admet l'existence d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel sont définies des variables aléatoires réelles $X_1, \dots, X_d$ i.i.d. de loi gaussienne centrée réduite. Alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien centré de matrice de covariance Id_d . La matrice Γ étant supposée symétrique positive, elle admet une racine carrée symétrique, c'est-à-dire une matrice A de taille $d \times d$ telle que $A = A'$ et $\Gamma = A^2 = AA'$. Mais alors le vecteur $Y = m + AX$ est, d'après la proposition 9.3, un vecteur gaussien de moyenne m_Y et de matrice de covariance Γ_Y donnés par

$$m_Y = A0 + m = m \text{ et } \Gamma_Y = AId_dA' = AA' = \Gamma,$$

ce qui conclue la preuve. ■

Exemple 9.4. Soit $m = (m_1, m_2, m_3) \in \mathbb{R}^3$. Considérons la matrice réelle symétrique

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

La matrice Γ admet les valeurs propres 0, 1 et 2, donc il s'agit d'une matrice réelle symétrique positive. Par conséquent, il existe un vecteur gaussien de vecteur moyenne $m \in \mathbb{R}^3$ et de matrice de covariance Γ .

Exemple 9.5. Soit $m = (m_1, m_2)' \in \mathbb{R}^2$. Considérons la matrice réelle symétrique

$$\Gamma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \text{ avec } a, b, c \in \mathbb{R}.$$

La matrice Γ étant carrée de dimension 2, ses valeurs propres sont positives si et seulement si

$$\begin{cases} \text{trace}(\Gamma) \geq 0 \\ \det \Gamma \geq 0 \end{cases} \iff \begin{cases} a + c \geq 0 \\ ac \geq b^2 \end{cases} \iff \begin{cases} a \geq 0, d \geq 0 \\ |b| \leq \sqrt{ad}. \end{cases}$$

Par conséquent, il existe un vecteur gaussien de vecteur moyenne $m \in \mathbb{R}^2$ et de matrice de covariance Γ si et seulement si $a \geq 0$, $d \geq 0$ et $|b| \leq \sqrt{ad}$.

Exemple 9.6. Considérons la matrice réelle symétrique

$$\Gamma = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

Nous souhaitons montrer qu'il existe un vecteur gaussien $X = (X_1, X_2, X_3)'$ de moyenne nulle et de matrice de covariance Γ . Pour cela, nous allons montrer que la forme quadratique $x \mapsto x'\Gamma x$ est positive.

Nous avons, pour tout $x = (x_1, x_2, x_3)' \in \mathbb{R}^3$,

$$x'\Gamma x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1^2 - 2x_1x_3 + 2x_2^2 + 4x_2x_3 + 5x_3^2.$$

Nous pouvons à présent réduire la forme quadratique, c'est-à-dire l'écrire sous la forme d'une combinaison linéaire de termes au carré :

$$\begin{aligned} x'\Gamma x &= (x_1 - x_3)^2 - x_3^2 + 2x_2^2 + 4x_2x_3 + 5x_3^2 = (x_1 - x_3)^2 + 2x_2^2 + 4x_2x_3 + 4x_3^2 \\ &= (x_1 - x_3)^2 + 2(x_2 + x_3)^2 - 2x_3^2 + 4x_3^2 = (x_1 - x_3)^2 + 2(x_2 + x_3)^2 + 2x_3^2. \end{aligned}$$

On en déduit que $x'\Gamma x \geq 0$ pour tout $x \in \mathbb{R}^3$ et, par conséquent, que Γ est positive. D'après la proposition 9.4, il existe donc un vecteur gaussien à valeurs dans $\mathbb{R}^3$ de moyenne nulle et de matrice de covariance Γ .

9.3 Propriétés de la loi d'un vecteur gaussien

Nous avons vu dans la section 9.1 que la loi d'un vecteur gaussien est caractérisée par son vecteur moyenne et sa matrice de covariance. A présent, nous donnons des informations supplémentaires sur cette loi, à savoir un critère nécessaire et suffisant pour l'indépendance des coordonnées d'un vecteur gaussien et un critère nécessaire et suffisant pour l'existence d'une densité par rapport à la mesure de Lebesgue.

9.3.1 Vecteurs gaussiens et indépendance

Nous avons vu que, en général, la non corrélation de deux variables aléatoires n'implique pas leur indépendance. C'est cependant le cas si l'on considère les coordonnées d'un vecteur gaussien.

Proposition 9.5.

Soit $X = (X_1, \dots, X_d)'$ un vecteur gaussien à valeurs dans $\mathbb{R}^d$. Les variables aléatoires X_i et X_j sont indépendantes si et seulement si elles sont décorrélées, c'est-à-dire si $\text{Cov}(X_i, X_j) = 0$. Plus généralement, $X_{i_1}, \dots, X_{i_n}$ sont mutuellement indépendantes si et seulement si $\text{Cov}(X_{i_l}, X_{i_k}) = 0$ pour tout $i_l \neq i_k$ et $k, l = 1, \dots, n$.

Remarque 9.2. Pour les vecteurs gaussiens, l'indépendance mutuelle et l'indépendance deux à deux sont donc des propriétés équivalentes !

Démonstration.

- Si $X_{i_1}, \dots, X_{i_n}$ sont mutuellement indépendantes, alors elles sont indépendantes deux à deux et, par conséquent, elles sont décorrélées, c'est-à-dire $\text{Cov}(X_{i_k}, X_{i_l}) = 0$ pour tout $i_l \neq i_k$ et $k, l = 1, \dots, n$.
- Supposons à présent que $\text{Cov}(X_{i_k}, X_{i_l}) = 0$ pour tout $i_l \neq i_k$ et $k, l = 1, \dots, n$. Alors le vecteur $(X_{i_1}, \dots, X_{i_n})$ est un vecteur gaussien qui a pour moyenne le vecteur $(\mathbb{E}(X_{i_1}), \dots, \mathbb{E}(X_{i_n}))'$ et pour matrice de covariance la matrice diagonale

$$\Gamma = \begin{pmatrix} \text{Var}(X_{i_1}) & & (0) \\ & \ddots & \\ (0) & & \text{Var}(X_{i_n}) \end{pmatrix}.$$

Soient $Y_1, \dots, Y_n$ des variables réelles i.i.d. gaussiennes centrées réduites. Alors le vecteur $Z = (\mathbb{E}(X_{i_1}) + \text{Var}(X_{i_1})Y_1, \dots, \mathbb{E}(X_{i_n}) + \text{Var}(X_{i_n})Y_n)$ est un vecteur gaussien de même vecteur moyenne et de même matrice de covariance que $(X_{i_1}, \dots, X_{i_n})$. On en déduit qu'ils ont même loi. Or les composantes de Z sont mutuellement indépendantes, c'est donc également le cas des composantes de $(X_{i_1}, \dots, X_{i_n})$. ■

Exemple 9.7.

1. Soit X un vecteur gaussien dont la matrice de covariance est diagonale. Alors les coordonnées de X sont mutuellement indépendantes.
2. Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien de matrice de covariance

$$\Gamma_X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix}.$$

Alors X_1 est indépendant de X_2 et de X_3 . Cependant, X_2 et X_3 ne sont pas indépendants.

9.3.2 Vecteur gaussien et densité

La proposition suivante caractérise l'existence d'une densité pour la loi d'un vecteur gaussien par rapport à la mesure de Lebesgue. Elle précise le cas échéant une densité pour cette loi.

Proposition 9.6.

Soit $X = (X_1, \dots, X_d)'$ un vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^d$ de vecteur moyenne m_X et de matrice de covariance Γ_X .

1. La loi du vecteur X est absolument continue par rapport à la mesure de Lebesgue si et seulement si sa matrice de covariance Γ_X est inversible.
2. De plus, si Γ_X est inversible, alors le vecteur aléatoire X admet pour densité par rapport à la mesure de Lebesgue λ_d la fonction $f_X : \mathbb{R}^d \rightarrow \mathbb{R}_+$ définie par

$$\forall x \in \mathbb{R}^d, f_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma_X}} \exp \left(-\frac{(x - m_X)' \Gamma_X^{-1} (x - m_X)}{2} \right).$$

Définition 9.5 (Loi gaussien dégénérée).

Tout vecteur gaussien dont la matrice de covariance n'est pas inversible est dit dégénéré. La loi d'un tel vecteur est appelée une loi gaussienne dégénérée.

Remarque 9.3. Dans le cas $d = 1$, nous retrouvons le fait qu'une variable gaussienne réelle n'admet une densité que si sa variance est non nulle. Dans le cas où sa variance est nulle, c'est-à-dire dans le cas où la variable est égale à une constante presque sûrement, la variable gaussienne est dégénérée au sens de la définition précédente.

9.3.3 Cas $d = 2$

Soit $X = (X_1, X_2)'$ un vecteur gaussien de moyenne m_X et de matrice de covariance Γ_X . Alors,

$$m_X = \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \end{pmatrix} \quad \text{et} \quad \Gamma_X = \begin{pmatrix} \text{Var} X_1 & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \text{Var} X_2 \end{pmatrix}.$$

Nous notons dans la suite

$m_1 = \mathbb{E}(X_1),$	$m_2 = \mathbb{E}(X_2),$
$\sigma_{x_1} = \sqrt{\text{Var}X_1},$	$\sigma_{x_2} = \sqrt{\text{Var}X_2}$
$\sigma_{x_1 x_2} = \text{Cov}(X_1, X_2).$	

Commençons par quelques remarques.

- Le déterminant de Γ_x est donné par $\det \Gamma_x = \sigma_{x_1}^2 \sigma_{x_2}^2 - \sigma_{x_1 x_2}^2$.
- De plus, d'après l'inégalité de Cauchy-Schwarz,

$$\sigma_{x_1 x_2}^2 \leq \sigma_{x_1}^2 \sigma_{x_2}^2,$$

ce qui confirme la condition liant a, b, c établie dans l'exemple 9.5 page 105.

- Par conséquent, si $\sigma_{x_1}^2 = 0$ ou si $\sigma_{x_2}^2 = 0$, alors $\det \Gamma_x = 0$.

Donnons la densité par rapport à λ_2 du vecteur X lorsqu'elle existe.

Proposition 9.7.

1. Si $\sigma_{x_1 x_2}^2 = \sigma_{x_1}^2 \sigma_{x_2}^2$, la loi de X n'est pas absolument continue.
2. Supposons $\sigma_{x_1 x_2}^2 \neq \sigma_{x_1}^2 \sigma_{x_2}^2$. Alors, le coefficient de corrélation

$$\rho = \frac{\sigma_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}}$$

entre X_1 et X_2 est bien défini. De plus, la loi de X admet pour densité par rapport à λ_2 la fonction $f_x : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ définie par

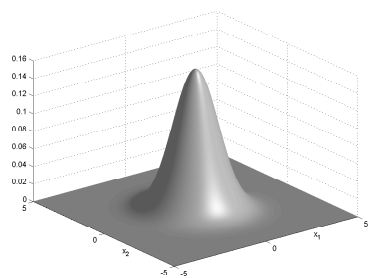
$$f_x(x) = \frac{1}{2\pi \sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2 (1 - \rho^2)}} \exp \left(\frac{-1}{1 - \rho^2} \left(\frac{(x_1 - m_1)^2}{2\sigma_{x_1}^2} - \frac{\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_{x_1} \sigma_{x_2}} + \frac{(x_2 - m_2)^2}{2\sigma_{x_2}^2} \right) \right).$$

Démonstration.

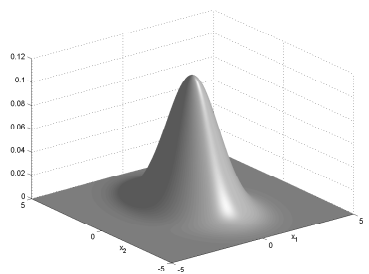
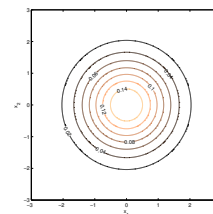
- La loi de X est absolument continue si et seulement si $\det \Gamma_x \neq 0$, c'est-à-dire si et seulement si $\sigma_{x_1 x_2}^2 \neq \sigma_{x_1}^2 \sigma_{x_2}^2$.
- Supposons à présent $\sigma_{x_1 x_2}^2 \neq \sigma_{x_1}^2 \sigma_{x_2}^2$. Dans ce cas, $\sigma_{x_1} \sigma_{x_2} \neq 0$ et le coefficient de corrélation ρ est bien défini. De plus, Γ_x est inversible et

$$\Gamma_x^{-1} = \frac{1}{\det \Gamma_x} \begin{pmatrix} \sigma_{x_2}^2 & -\sigma_{x_1 x_2} \\ -\sigma_{x_1 x_2} & \sigma_{x_1}^2 \end{pmatrix} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1/\sigma_{x_1}^2 & -\rho/(\sigma_{x_1} \sigma_{x_2}) \\ -\rho/(\sigma_{x_1} \sigma_{x_2}) & 1/\sigma_{x_2}^2 \end{pmatrix}$$

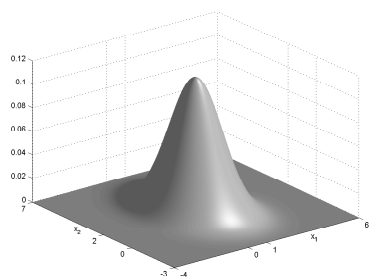
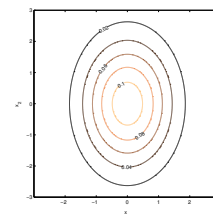
car $\det \Gamma_x = \sigma_{x_1}^2 \sigma_{x_2}^2 (1 - \rho^2)$ et $\sigma_{x_1 x_2} = \rho \sigma_{x_1} \sigma_{x_2}$. La proposition 9.6 permet de conclure. ■



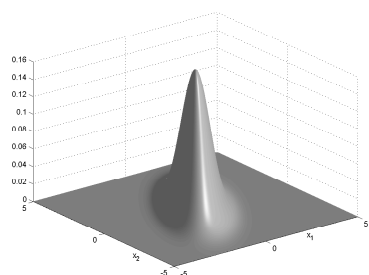
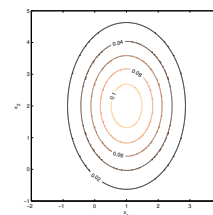
Loi gaussienne standard



Loi $\mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right)$



Loi $\mathcal{N}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right)$



Loi $\mathcal{N}\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$

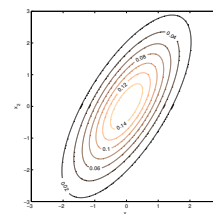


FIGURE 9.1 – Densité d'un vecteur gaussien et ses courbes de niveau

Considérons un vecteur gaussien $X = (X_1, X_2)'$ à valeurs dans $\mathbb{R}^2$ de loi absolument continue et commentons à présent l'allure des courbes de niveau de sa densité. La figure 9.1 page 109 illustre les remarques suivantes.

- Les courbes de niveau de la densité de X par rapport à λ_2 sont des ellipses concentriques dont le centre a mêmes coordonnées que le vecteur moyenne de X .
- Par ailleurs, la dépendance ou non des coordonnées de X se lit sur le graphique : elles sont indépendantes si et seulement si un des axes est parallèle à l'axe des abscisses et l'autre est parallèle à l'axe des ordonnées.
- Enfin, si X_1 et X_2 sont indépendantes, les lignes de niveau sont des cercles si et seulement si X_1 et X_2 ont

même variance.

Intéressons nous à présent au cas d'un vecteur dégénéré.

Proposition 9.8.

Supposons $\sigma_{X_1, X_2}^2 = \sigma_{X_1}^2 \sigma_{X_2}^2$.

1. Si $\sigma_{X_i} = 0$, avec $i \in \{1, 2\}$, alors les variables X_1 et X_2 sont indépendantes et X_i est presque sûrement constante.
2. Si $\sigma_{X_1, X_2} = \sigma_{X_1} \sigma_{X_2} \neq 0$, alors presque sûrement $X_2 = \frac{\sigma_{X_2}}{\sigma_{X_1}}(X_1 - m_1) + m_2$.
3. Si $\sigma_{X_1, X_2} = -\sigma_{X_1} \sigma_{X_2} \neq 0$, alors presque sûrement $X_2 = -\frac{\sigma_{X_2}}{\sigma_{X_1}}(X_1 - m_1) + m_2$.

Démonstration.

1. La première propriété se déduit immédiatement de la proposition 9.5.
2. Le vecteur aléatoire $Y = \sigma_{X_1}(X_2 - m_2) - \sigma_{X_2}(X_1 - m_1)$ est un vecteur gaussien de moyenne nulle et de variance nulle. En effet, en utilisant la bilinéarité de la covariance, on obtient

$$\begin{aligned} \text{Var}(Y) &= \text{Cov}(Y, Y) = \sigma_{X_1}^2 \text{Var}(X_2) + \sigma_{X_2}^2 \text{Var}(X_1) - 2\sigma_{X_1} \sigma_{X_2} \text{Cov}(X_1, X_2) \\ &= \sigma_{X_1}^2 \sigma_{X_2}^2 + \sigma_{X_2}^2 \sigma_{X_1}^2 - 2\sigma_{X_1} \sigma_{X_2} \sigma_{X_1, X_2} = 0, \end{aligned}$$

sous l'hypothèse $\sigma_{X_1, X_2} = \sigma_{X_1} \sigma_{X_2}$. Ainsi, Y est constant égal à 0 presque sûrement, ce qui nous permet de conclure.

3. Le même calcul, avec la variable aléatoire $Y = \sigma_{X_1}(X_2 - m_2) + \sigma_{X_2}(X_1 - m_1)$, permet de conclure. ■

9.4 Vecteurs gaussiens et convergence en loi

9.4.1 Stabilité de l'ensemble des vecteurs gaussiens

L'ensemble des vecteurs gaussiens est stable pour la convergence en loi et la convergence en loi d'une suite de vecteurs gaussiens est équivalent à la convergence des suites des moyennes et des covariances.

Proposition 9.9.

Soit $(X^{(n)})_{n \in \mathbb{N}}$ une suite de vecteurs gaussiens à valeurs dans $\mathbb{R}^d$. Pour tout $n \in \mathbb{N}$, posons $m_n = \mathbb{E}(X^{(n)})$ et notons Γ_n la matrice de covariance de $X^{(n)}$.

1. La suite de vecteurs aléatoires $(X^{(n)})_{n \in \mathbb{N}}$ converge en loi si et seulement si les deux suites $(m_n)_{n \in \mathbb{N}}$ et $(\Gamma_n)_{n \in \mathbb{N}}$ convergent (respectivement dans $\mathbb{R}^d$ et $\mathcal{M}_{d,d}(\mathbb{R})$).
2. Par ailleurs, si la suite $(X^{(n)})_{n \in \mathbb{N}}$ converge en loi, alors elle converge en loi vers une variable aléatoire X de loi gaussienne $\mathcal{N}(m, \Gamma)$, avec

$$m = \lim_{n \rightarrow +\infty} m_n = \lim_{n \rightarrow +\infty} \mathbb{E}(X^{(n)})$$

et, pour tout $1 \leq i, j \leq d$,

$$\Gamma_{ij} = \lim_{n \rightarrow +\infty} (\Gamma_n)_{ij} = \lim_{n \rightarrow +\infty} \text{Cov}(X_i^{(n)}, X_j^{(n)}).$$

9.4.2 Théorème central limite

Nous terminons cette partie en généralisant le théorème central limite au cadre des vecteurs aléatoires.

Théorème 9.10.

Soit $(X^{(n)})_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^d$. Supposons que la norme du vecteur $X^{(1)}$ est de carré intégrable, c'est-à-dire supposons que

$$\mathbb{E} \left(|X^{(1)}|^2 \right) < \infty.$$

Notons alors Γ la matrice de covariance du vecteur $X^{(1)}$. Alors

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X^{(i)} - \mathbb{E} \left(X^{(1)} \right) \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z,$$

où Z est un vecteur gaussien à valeurs dans $\mathbb{R}^d$ de loi $\mathcal{N}(0, \Gamma)$.

9.5 Vecteurs gaussiens et projection orthogonale dans L^2

Dans cette partie, nous nous donnons un vecteur gaussien $X = (X_1, \dots, X_d)'$ et nous cherchons à donner la meilleure approximation (au sens de la norme L^2) de X_1 à partir d'une fonction mesurable de $X_2, \dots, X_d$. Dans la suite, $B(\mathbb{R}^p, \mathbb{R}^n)$ désigne l'ensemble des fonction boréliennes de $\mathbb{R}^p$ dans $\mathbb{R}^n$.

Proposition 9.11.

Soit $(X_1, X_2)'$ un vecteur gaussien à valeurs dans $\mathbb{R}^2$. Alors

$$\inf_{f \in B(\mathbb{R}, \mathbb{R})} \mathbb{E} \left((X_1 - f(X_2))^2 \right) = \inf_{a_1, a_2 \in \mathbb{R}} \mathbb{E} \left((X_1 - (a_1 + a_2 X_2))^2 \right) = \mathbb{E} \left((X_1 - W)^2 \right)$$

où

$$W = \begin{cases} \mathbb{E}(X_1) & \text{si } \text{Var}(X_2) = 0 \\ \mathbb{E}(X_1) + \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)}(X_2 - \mathbb{E}(X_2)) & \text{si } \text{Var}(X_2) \neq 0. \end{cases}$$

Dans le cas $d = 2$, nous observons donc que la meilleure approximation de X_1 par une fonction mesurable de X_2 est donnée par une fonction affine de X_2 . Ce résultat se généralise aux vecteurs gaussiens de plus grande dimension.

Proposition 9.12.

Soient $Y = (Y_1, \dots, Y_p)'$ un vecteur à valeurs dans $\mathbb{R}^p$ et $Z = (Z_1, \dots, Z_n)'$ un vecteur à valeurs dans $\mathbb{R}^n$. Supposons que $X = (Y_1, \dots, Y_p, Z_1, \dots, Z_n)'$ est un vecteur gaussien de matrice de covariance

$$\Gamma_X = \begin{pmatrix} \Gamma_Y & \Gamma_{Y,Z} \\ \Gamma_{Y,Z}' & \Gamma_Z \end{pmatrix},$$

où Γ_Y est la matrice de covariance de Y et Γ_Z celle de Z .

Si Γ_Z est inversible, alors

$$\inf_{f \in B(\mathbb{R}^n, \mathbb{R}^p)} \mathbb{E} (\|Y - f(Z)\|^2) = \mathbb{E} (\|Y - W\|^2),$$

où W est un vecteur gaussien à valeurs dans $\mathbb{R}^p$ donné par

$$W = \mathbb{E}(Y) + (\Gamma_{Y,Z} \Gamma_Z^{-1}) (Z - \mathbb{E}(Z)).$$

Exemple 9.8. Soit $X = (X_1, X_2, X_3)'$ un vecteur gaussien de moyenne $m_X = (1, 2, -1)$ et de matrice de covariance Γ_X donnée par

$$\Gamma_X = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

Notre but est d'utiliser la proposition précédente pour déterminer W tel que

$$\inf_{f \in B(\mathbb{R}^2, \mathbb{R})} \mathbb{E} (|X_1 - f(X_2, X_3)|^2) = \mathbb{E} (|X - W|^2).$$

Dans notre situation, $Y = X_1$, $Z = (X_2, X_3)$ et

$$\Gamma_Y = 1, \Gamma_Z = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} \text{ et } \Gamma_{Y,Z} = \begin{pmatrix} 0 & -1 \end{pmatrix}.$$

Par conséquent, le vecteur W vérifie

$$\begin{aligned} W &= 1 + \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 5/6 & -1/3 \\ -1/3 & 1/3 \end{pmatrix} \left(\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right) \\ &= 1 + \begin{pmatrix} 1/3 & -1/3 \end{pmatrix} \left(\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right) \\ &= \frac{X_2 - X_3}{3}. \end{aligned}$$

Nous avons ainsi calculé la meilleure approximation de X_1 par une fonction mesurable de X_2 et X_3 .

Annexe A

Lois classiques

Loi de la v.a. X	$\mathbb{P}_X$	Espérance	Variance	Fonction Caractéristique
Loi de Bernoulli $\mathcal{B}(p)$ avec $p \in [0, 1]$	$\mathbb{P}_X = (1-p)\delta_0 + p\delta_1$	p	$p(1-p)$	$\varphi_X(t) = 1 - p + pe^{it}$
Loi Binomiale $\mathcal{B}(n, p)$ avec $n \in \mathbb{N}^*$ et $p \in [0, 1]$	$\mathbb{P}_X = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \delta_k$	np	$np(1-p)$	$\varphi_X(t) = (1 - p + pe^{it})^n$
Loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	$\mathbb{P}_X = \sum_{k=0}^{+\infty} \frac{e^{-\lambda} \lambda^k}{k!} \delta_k$	λ	λ	$\varphi_X(t) = e^{\lambda(e^{it}-1)}$
Loi Géométrique $\mathcal{G}(p)$ avec $p \in]0, 1[$	$\mathbb{P}_X = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\varphi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}$

Loi de la v.a. X	Densité de $\mathbb{P}_X$	Espérance	Variance	Fonction de répartition	Fonction Caractéristique
Loi Gaussienne $\mathcal{N}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$F_X(x) = \int_{-\infty}^x f_X(t) \lambda_1(dt)$	$\varphi_X(x) = e^{imx - \frac{\sigma^2 x^2}{2}}$
Loi Uniforme $\mathcal{U}([0, 1])$	$f_X(x) = \mathbf{1}_{[0,1]}(x)$	$\frac{1}{2}$	$\frac{1}{12}$	$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \in [0, 1] \\ 1 & \text{si } x > 1 \end{cases}$	$\varphi_X(x) = \begin{cases} \frac{e^{ix}-1}{ix} & \text{si } x \neq 0 \\ 1 & \text{si } x = 0 \end{cases}$
Loi Exponentielle $\mathcal{E}(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$	$f_X(x) = \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{\mathbb{R}_+^*}(x)$	λ	λ^2	$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x/\lambda} & \text{si } x \geq 0 \end{cases}$	$\varphi_X(x) = \frac{1}{1 - i\lambda x}$
Loi de Cauchy	$f_X(x) = \frac{1}{\pi(1+x^2)}$	n'existe pas	n'existe pas	$F_X(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$	$\varphi_X(x) = e^{- x }$

Annexe B

Simulation d'une variable aléatoire

Ce chapitre est dédié à la simulation numérique de variables aléatoires. Dans la plupart des cas, nous apprendrons à simuler une variable aléatoire de loi donnée à l'aide de variables aléatoires de loi uniforme sur $[0, 1]$.

Le texte est accompagné d'exemples de code en Matlab et en Python¹. Dans toute la suite nous supposons que le code en Python comporte l'entête suivante, qui nous permettra d'utiliser la fonction `random()` du module `random` et les fonctions mathématiques usuelles du module `math`.

Python
<pre># Entête des extraits de code en Python from random import random from math import *</pre>

Hypothèses/Notations pour ce chapitre

- Toutes les variables aléatoires considérées sont définies sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.
- Tout sous-ensemble de $\mathbb{R}^d$ est muni de sa tribu borélienne.
- Tout ensemble E fini ou dénombrable est muni de sa tribu discrète $\mathcal{P}(E)$.

B.1 Simulation d'une variable aléatoire de loi uniforme en Matlab et en Python

Voici comment simuler une variable aléatoire U de loi uniforme sur $[0, 1]$ dans Matlab et Python.

Matlab
<pre>U=rand;</pre>
Python
<pre>U=random()</pre>

1. Matlab et le module `stats` de la bibliothèque SciPy de Python proposent également des solutions pré-implémentées pour simuler des variables aléatoires. Cependant ce n'est pas le cas pour tous les langages de programmation/toutes les configurations. Par exemple, les méthodes vues dans ce chapitre se révèlent utiles avec PyPy (un interpréteur de Python avec compilation en temps réel) qui ne supporte pas la bibliothèque SciPy.

Remarque B.1. En théorie, une variable aléatoire U de loi uniforme sur $[0, 1]$ est différente de 0 ou 1 presque sûrement. En effet, par définition de la loi uniforme sur $[0, 1]$,

$$\mathbb{P}(U \in \{0, 1\}) = \int_{\{0,1\}} \mathbf{1}_{[0,1]} d\lambda_1 = \lambda_1(\{0, 1\}) = 0.$$

En Matlab, la fonction `rand` ne renvoie que des valeurs différentes de 0 et 1. En Python cependant, la fonction `random()` peut retourner la valeur 0 (mais pas la valeur 1). Pour obtenir uniquement des valeurs dans $]0, 1[$, une solution possible consiste à utiliser la fonction `myRandom` définie ci-dessous.

Python
<pre>def myRandom(): U=random() while U==0 or U==1: U=random() return U</pre>

Lorsque l'on fait appel plusieurs fois de suite à ces fonctions, le programme retourne des valeurs que l'on peut supposer indépendantes entre elles. Ainsi, pour simuler un vecteur aléatoire V contenant N variables aléatoires indépendantes de loi uniforme sur $[0, 1]$, nous pouvons utiliser le code suivant.

Matlab
<pre>N=1000; V=rand(1,N);</pre>
Python
<pre>N=1000 V=[random() for _ in xrange(N)]</pre>

B.2 Simulation d'une loi gaussienne par la méthode de Box-Muller

La simulation d'une variable aléatoire X de loi normale centrée réduite par la méthode de Box-Muller se déduit de la proposition suivante.

Proposition B.1.

Soient U_1 et U_2 deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Alors la variable aléatoire X définie par

$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

est une variable aléatoire de loi gaussienne centrée réduite, c'est-à-dire la loi gaussienne $\mathcal{N}(0, 1)$.

Idées de la démonstration.

- Tout d'abord les hypothèses sur U_1, U_2 reviennent à supposer que le couple (U_1, U_2) est de loi uniforme sur $[0, 1]^2$ (cf. chapitre 6). Autrement dit, sa loi est absolument continue de densité $f_{(U_1, U_2)} = \mathbf{1}_{[0,1]^2}$.

- Nous allons utiliser la technique décrite dans le paragraphe 2.5.2 du second chapitre, i.e. nous allons utiliser un changement de variables. Notons tout d'abord que comme le support de la densité de (U_1, U_2) est $[0, 1]^2$,

$$\mathbb{P}((U_1, U_2) \in]0, 1[^2) = 1.$$

Afin d'effectuer un changement de variables licite, nous introduisons la variable aléatoire

$$Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

qui est, tout comme X , définie presque sûrement. Nous nous intéressons alors à la loi du couple (X, Y) . Par définition, $(X, Y) = g(U_1, U_2)$ presque sûrement avec $g :]0, 1[^2 \rightarrow \mathbb{R}^2$ définie par

$$g(u_1, u_2) = \left(\sqrt{-2 \ln u_1} \cos(2\pi u_2), \sqrt{-2 \ln u_1} \sin(2\pi u_2) \right).$$

Considérons pour cela une fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue positive. Dès lors, d'après le théorème du transport,

$$\mathbb{E}(\varphi(X, Y)) = \mathbb{E}(\varphi(g(U_1, U_2))) = \int_{[0,1]^2} \varphi(g(u_1, u_2)) \lambda_2(du_1, du_2) = \int_{]0,1[^2} \varphi(g(u_1, u_2)) \lambda_2(du_1, du_2)$$

car la loi de (U_1, U_2) est la loi uniforme sur $[0, 1]^2$. Le lecteur vérifiera que le changement de variable $(x, y) = g(u_1, u_2)$ est licite et qu'en l'effectuant, il vient :

$$\mathbb{E}(\varphi(X, Y)) = \mathbb{E}(\varphi(g(U_1, U_2))) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \varphi(x, y) e^{-\frac{x^2+y^2}{2}} \lambda_2(dx, dy).$$

Cette égalité étant vraie pour toute fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ continue positive, la loi du couple (X, Y) est absolument continue de densité $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$ donnée par

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

Il s'agit de la loi gaussienne standard sur $\mathbb{R}^2$ (cf. chapitre 9).

- Enfin, la loi de X s'obtient en utilisant le théorème de Fubini (cf. chapitre 5, Tableau 5.1). Ce théorème permet d'établir que la loi de X est aussi absolument continue et que sa densité $f_X : \mathbb{R} \rightarrow \mathbb{R}$ est donnée par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) \lambda_2(dy) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Vu l'expression de sa densité, la loi de X est bien la loi gaussienne centrée réduite. ■

Remarque B.2. On peut remplacer le *cosinus* par un *sinus* dans la proposition sans changer le résultat, c'est-à-dire que sous les mêmes hypothèses, la variable aléatoire

$$Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

suit aussi la loi gaussienne centrée réduite.

On déduit immédiatement de cette proposition les portions de code suivantes pour simuler une variable aléatoire réelle X de loi normale centrée réduite.

Matlab
<pre>U=rand(1,2); X=sqrt(-2*log(U(1)))*cos(2*pi*U(2)); % En Matlab, tableaux indicés à partir de 1</pre>
Python
<pre>U=[1-random(),random()] X=sqrt(-2*log(U[0]))*cos(2*pi*U[1]) # En Python, tableaux indicés à partir de 0</pre>

Remarque B.3. Dans Python, la fonction `random()` peut renvoyer la valeur 0, ce qui pose problème pour le calcul de $\ln U_1$. Nous pallions ici cette difficulté en simulant U_1 avec l'expression `1-random()`, qui peut retourner la valeur 1, mais pas la valeur 0.

B.3 Simulation d'une variable aléatoire discrète

Dans cette section, nous apprenons à simuler une variable aléatoire discrète lorsque l'on dispose de sa loi et d'une variable aléatoire U de loi uniforme sur $[0, 1]$.

B.3.1 Un premier exemple : v.a. de Bernoulli

Nous souhaitons simuler une variable aléatoire X de loi de Bernoulli de paramètre $p \in [0, 1]$, c'est-à-dire une variable aléatoire égale à 0 avec probabilité $1 - p$ et égale à 1 avec probabilité p .

Considérons U une variable aléatoire de loi uniforme sur $[0, 1]$. Alors la loi de la variable aléatoire $X = \mathbf{1}_{U \leq p}$ est la loi de Bernoulli de paramètre p . Nous rappelons que

$$\mathbb{P}(X = 1) = \mathbb{P}(U \leq p) = p \quad \text{et} \quad \mathbb{P}(X = 0) = \mathbb{P}(U > p) = 1 - p.$$

En Matlab ou en Python, nous pourrions donc simuler une variable aléatoire de loi de Bernoulli de paramètre $p \in [0, 1]$ de la façon suivante. Nous avons choisi ici $p = 0.2$.

Matlab
<pre>p=0.2; u=rand; X=(u<=p);</pre>
Python
<pre>p=0.2 u=random() X=(u<=p)</pre>

B.3.2 Variables aléatoires discrètes générales

En suivant une méthode similaire, nous apprenons à présent à simuler une variable aléatoire discrète, c'est-à-dire une variable aléatoire prenant ses valeurs, presque sûrement, dans un ensemble fini ou dénombrable, noté ici

$$E = \{x_i / i \in I\} \quad \text{avec } I = \mathbb{N} \text{ ou } I = \{0, \dots, n\}.$$

Soit μ une mesure de probabilité sur $\{x_i / i \in I\}$. Cette loi s'écrit sous la forme

$$\mu = \sum_{i \in I} p_i \delta_{x_i}$$

où $p_i = \mu(\{x_i\}) \in [0, 1]$. Notre but est de définir une variable aléatoire X de loi μ , c'est-à-dire telle que

$$\mathbb{P}(X = x_i) = \mu(\{x_i\}) = p_i, \quad \forall i \in I,$$

à l'aide d'une variable aléatoire U de loi uniforme sur $[0, 1]$. La proposition suivante répond à cette question.

Proposition B.2.

La variable aléatoire X définie par

$$X = x_0 \mathbf{1}_{U \leq p_0} + \sum_{i \in I \setminus \{0\}} x_i \mathbf{1}_{\sum_{j=0}^{i-1} p_j < U \leq \sum_{j=0}^i p_j} = \begin{cases} x_0 & \text{si } U \leq p_0 \\ x_1 & \text{si } p_0 < U \leq p_0 + p_1 \\ x_2 & \text{si } p_0 + p_1 < U \leq p_0 + p_1 + p_2 \\ \dots & \\ x_i & \text{si } p_0 + \dots + p_{i-1} < U \leq p_0 + \dots + p_i, \quad \forall i \in I, \quad i \geq 1 \\ \dots & \end{cases}$$

est une variable aléatoire de loi μ , c'est-à-dire qu'elle vérifie

$$\mathbb{P}(X = x_i) = \mu(\{x_i\}) = p_i, \quad \forall i \in I.$$

Remarque B.4. L'ensemble $A = \{\sum_{j=0}^i p_j / i \in I\}$ est dénombrable donc de mesure de Lebesgue nulle. Par conséquent, une variable aléatoire U de loi uniforme sur $[0, 1]$ est telle que

$$\mathbb{P}(U \in A) = 0$$

car sa loi est absolument continue. Autrement dit,

$$U \notin \left\{ \sum_{j=0}^i p_j / i \in I \right\} \quad \text{presque sûrement}$$

Par suite, l'ordre des inégalités strictes/non-strictes dans l'écriture de X ne modifie pas sa loi.

Remarque B.5. Il pourra être intéressant de définir une famille de nombre réels $Q = (q_i)_{i \in I}$ par

$$q_{-1} = 0, \quad q_0 = p_0, \quad q_1 = p_0 + p_1, \quad \dots, \quad q_i = \sum_{k=0}^i p_k, \quad i \in I.$$

De cette façon, étant donné U de loi uniforme sur $[0, 1]$, la variable aléatoire X s'écrit sous la forme

$$X = x_{i_0} \quad \text{où } i_0 \text{ est l'unique indice tel que } q_{i_0-1} < U \leq q_{i_0}$$

ou encore $X = x_{i_0}$ où i_0 est le plus petit indice tel que $U \leq q_{i_0}$.

Démonstration. La variable aléatoire X définie dans la proposition est à valeurs dans l'ensemble

$$E = \{x_i / i \in I\}.$$

Cet ensemble étant fini ou dénombrable, la loi de X est discrète et s'écrit :

$$\mathbb{P}_X = \sum_{i \in I} \mathbb{P}(X = x_i) \delta_{x_i}.$$

De plus par définition,

$$\{X = x_0\} = \{U \leq p_0\} \quad \text{et pour tout } i \geq 1, \quad \{X = x_i\} = \left\{ \sum_{j=0}^{i-1} p_j < U \leq \sum_{j=0}^i p_j \right\}.$$

Par suite,

$$\mathbb{P}(X = x_i) = \begin{cases} \mathbb{P}(U \leq p_0) = p_0 & \text{si } i = 0 \\ \mathbb{P}\left(\sum_{j=0}^{i-1} p_j < U \leq \sum_{j=0}^i p_j\right) = \sum_{j=0}^i p_j - \sum_{j=0}^{i-1} p_j = p_i & \text{si } i \geq 1 \end{cases}$$

car la loi de U est la loi uniforme sur $[0, 1]$ et car $\sum_{j=0}^k p_j \in [0, 1]$ pour tout $k \in I$. Ainsi, la loi de X est : $\mathbb{P}_X = \sum_{i \in I} \mathbb{P}(X = x_i) \delta_{x_i} = \mu$. ■

Exemple B.1. Dans ce premier exemple, nous considérons la mesure de probabilité

$$\mu = \frac{1}{10} \delta_0 + \frac{4}{10} \delta_1 + \frac{2}{10} \delta_2 + \frac{3}{10} \delta_3.$$

Ainsi, étant donnée une variable aléatoire U de loi uniforme sur $[0, 1]$, on peut définir une variable aléatoire X de loi μ par

$$X = \begin{cases} 0 & \text{si } U \leq 1/10 \\ 1 & \text{si } 1/10 < U \leq 5/10 \\ 2 & \text{si } 5/10 < U \leq 7/10 \\ 3 & \text{si } 7/10 < U \leq 1. \end{cases}$$

En Matlab ou en Python, cela peut-être interprété de la façon suivante.

Matlab
<pre>mu=[0.1,0.4,0.2,0.3]; Q=cumsum(mu); U=rand; X=find(U<=Q,1,'first')-1</pre>
Python
<pre>mu=[0.1,0.4,0.2,0.3] U=random() i,q=0,mu[0] while not U<=q: i=i+1,q+=mu[i] X=i</pre>

Exemple B.2. Nous souhaitons à présent simuler une variable aléatoire de loi géométrique de paramètre $p \in [0, 1[$ donné, c'est-à-dire de loi

$$\mu = \sum_{i \in \mathbb{N}^*} (1-p)^{i-1} p \delta_i.$$

Alors, étant donnée une variable aléatoire U de loi uniforme sur $[0, 1]$, on peut définir une variable aléatoire X de loi μ par

$$X = \begin{cases} 1 & \text{si } U \leq p \\ 2 & \text{si } p < U \leq p + (1-p)p \\ 3 & \text{si } p + (1-p)p < U \leq p + (1-p)p + (1-p)^2 p \\ \dots & \\ i & \text{si } \sum_{k=1}^{i-1} (1-p)^{k-1} p < U \leq \sum_{k=1}^i (1-p)^{k-1} p, \quad \forall i \geq 1, \end{cases}$$

en respectant la convention $\sum_{k=1}^0 (1-p)^{k-1} p = 0$.

On peut donc simuler X en faisant une boucle sur $\{1, 2, \dots\}$ jusqu'à atteindre un indice i vérifiant

$$\sum_{k=1}^{i-1} (1-p)^{k-1} p < U \leq \sum_{k=1}^i (1-p)^{k-1} p.$$

Toutefois, dans les situations où le nombre d'indices possibles est infini, comme ici, on essaiera d'éviter cette démarche, car la longueur de la boucle pourrait prendre des valeurs excessivement grandes pour peu que U soit proche de 1 et le temps de calcul être très important.

Dans notre situation, nous pouvons remarquer que

$$\begin{aligned} X = i &\iff \sum_{k=1}^{i-1} (1-p)^{k-1} p \leq U < \sum_{k=1}^i (1-p)^{k-1} p \\ &\iff 1 - (1-p)^{i-1} \leq U < 1 - (1-p)^i \\ &\iff (1-p)^i < 1 - U \leq (1-p)^{i-1} \\ &\iff i-1 \leq \frac{\ln(1-U)}{\ln(1-p)} < i, \end{aligned}$$

donc

$$X = 1 + \left\lceil \frac{\ln(1-U)}{\ln(1-p)} \right\rceil$$

où $[t]$ désigne la partie entière de $t \in \mathbb{R}$.

En Matlab ou en Python, on obtient donc le code suivant pour simuler une variable aléatoire de loi géométrique.

Matlab
<pre>p=0.01; U=rand; X=1+floor(log(1-U)/log(1-p));</pre>
Python
<pre>p=0.01 U=random() X=1+floor(log(1-U)/log(1-p))</pre>

Remarque B.6. Dans Matlab et Python, les commandes `U=rand` et `U=random()` ne renvoient jamais la valeur 1. Par conséquent, dans le code précédent, la définition de la variable X ne produira pas d'erreur de définition.

Exemple B.3. Voyons à présent le cas de la loi binomiale. Nous souhaitons simuler une variable aléatoire X de loi binomiale de paramètre $p \in [0, 1]$ et $n \in \{1, 2, \dots\}$, c'est-à-dire de loi $\mu = \sum_{i=0}^n p_i \delta_i$ avec

$$p_i = \mathbb{P}(X = i) = C_n^i p^i (1-p)^{n-i}, \quad \forall i \in \{0, 1, \dots, n\}.$$

Nous pouvons ici aussi précéder en utilisant la proposition B.2. Cependant, nous sommes dans une situation défavorable : X peut prendre une infinité de valeurs (comme dans l'exemple précédent), mais il n'y a pas d'expression simple pour $p_0 + \dots + p_i$, ce qui nous oblige donc à utiliser une boucle sur un nombre potentiellement infini d'indices et conduit à un temps de calcul potentiellement important. De plus, le calcul des coefficients binomiaux est aussi assez coûteux en temps, ce qui ralentit encore le calcul de $p_0 + \dots + p_i$, pour chaque valeur de i .

Cependant, en utilisant la définition de la loi binomiale, on se rend compte qu'il s'agit également de la loi de la somme de n variables aléatoires indépendantes de Bernoulli de paramètre p . Ainsi, pour simuler X , on préférera simuler n variables aléatoires $X_1, \dots, X_n$ indépendantes, de loi de Bernoulli de paramètre p , puis poser

$$X = X_1 + \dots + X_n.$$

En Matlab ou en Python, nous pouvons donc utiliser le code suivant.

Matlab
<pre>p=0.2 n=1000000 X=sum(rand(1,n)<=p)</pre>
Python
<pre>p=0.2 n=1000000 X=sum([random()<=p for _ in xrange(n)])</pre>

B.4 Simulation d'une v.a à partir de sa fonction de répartition

B.4.1 Fonction de répartition réciproque généralisée

Étant donnée une variable aléatoire Y de loi μ sur $\mathbb{R}$, rappelons que la fonction de répartition de Y est définie par

$$F_Y(t) = \mathbb{P}(Y \leq t) = \mu([-\infty, t]), \quad \forall t \in \mathbb{R}.$$

Définissons la fonction réciproque généralisée de F_Y .

Définition-Proposition B.3.

Pour tout $u \in]0, 1[$,

$$G_Y(u) = \inf\{t \in \mathbb{R}, F_Y(t) \geq u\}, \quad (\text{B.1})$$

est un réel et la fonction $G_Y :]0, 1[\rightarrow \mathbb{R}$ est appelée fonction réciproque généralisée de F_Y .

Démonstration. Pour tout $u \in]0, 1[$, nous pouvons poser

$$G_Y(u) = \inf\{t \in \mathbb{R}, F_Y(t) \geq u\},$$

avec par la convention classique $\inf \emptyset = +\infty$. Il s'agit ici d'établir que $G_Y(u) \in \mathbb{R}$.

D'après la proposition 4.2 du chapitre 4, la fonction de répartition F_Y est une fonction croissante telle que

$$\lim_{t \rightarrow -\infty} F_Y(t) = 0 \text{ et } \lim_{t \rightarrow +\infty} F_Y(t) = 1.$$

Dès lors, pour tout $u \in]0, 1[$, $\{t \in \mathbb{R}, F_Y(t) \geq u\}$ est un ensemble minoré non vide. Par conséquent, $G_Y(u) = \inf\{t \in \mathbb{R}, F_Y(t) \geq u\} \in \mathbb{R}$. ■

Dans le cas où F_Y est strictement croissante, sa fonction réciproque généralisée est simplement son inverse.

Proposition B.4.

Si la fonction de répartition F_Y définit une bijection d'un intervalle $]a, b[$ (avec $-\infty \leq a < b \leq +\infty$) sur l'intervalle $]0, 1[$, alors sa fonction réciproque généralisée $G_Y :]0, 1[\rightarrow]a, b[$ définie par (B.1), est l'inverse de la restriction F_Y à $]a, b[$, i.e. pour $u \in]0, 1[$ et $t \in]a, b[$

$$G_Y(u) = t \iff F_Y(t) = u.$$

En particulier, si F_Y est strictement croissante sur $\mathbb{R}$, alors elle est bijective de $\mathbb{R}$ sur $]0, 1[$ et $G_Y = F_Y^{-1}$.

Notons enfin, que contrairement à ce que la notation pourrait laisser penser, la fonction de répartition d'une variable aléatoire Y et *a fortiori* sa réciproque généralisée ne dépendent que de la loi μ de Y .

Exemple B.4. Soit Y une variable aléatoire de loi uniforme sur $[a, b]$, avec $-\infty < a < b < +\infty$, c'est-à-dire de densité

$$f_Y(x) = \frac{1}{b-a} \mathbf{1}_{x \in [a, b]}, \quad \forall x \in \mathbb{R}.$$

Alors, d'après l'exemple 3.1 du chapitre 3, pour tout $t \in \mathbb{R}$,

$$F_Y(t) = \int_{]-\infty, t]} f_Y(x) d\lambda_1(x) = \frac{t-a}{b-a} \mathbf{1}_{t \in [a, b]} + \mathbf{1}_{t \in [b, +\infty[}.$$

Par conséquent, F_Y définit une bijection de $]a, b[$ sur $]0, 1[$. De plus, pour tout $u \in]0, 1[$ et $t \in]a, b[$,

$$F_Y(t) = u \iff \frac{t-a}{b-a} = u \iff t = a + (b-a)u.$$

Nous en déduisons que pour tout $u \in]0, 1[$,

$$G_Y(u) = a + (b-a)u.$$

Exemple B.5. Soit Y une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$, c'est-à-dire de densité

$$f_Y(x) = \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{x > 0}, \quad \forall x \in \mathbb{R}.$$

Alors, pour tout $t \in \mathbb{R}$,

$$F_Y(t) = \int_{]-\infty, t]} f_Y(x) d\lambda_1(x) = \int_{]-\infty, t]} \frac{e^{-x/\lambda}}{\lambda} \mathbf{1}_{x>0} d\lambda_1(x) = (1 - e^{-t/\lambda}) \mathbf{1}_{t \geq 0}.$$

Par conséquent, F_Y définit une bijection de $]0, +\infty[$ sur $]0, 1[$. De plus, pour tout $u \in]0, 1[$ et $t \in]0, +\infty[$,

$$F_Y(t) = u \iff 1 - e^{-t/\lambda} = u \iff t = -\lambda \ln(1 - u).$$

Nous en déduisons que pour tout $u \in]0, 1[$,

$$G_Y(u) = -\lambda \ln(1 - u).$$

Exemple B.6. Soit Y une variable aléatoire discrète à valeurs dans $\mathbb{N}$ telle que

$$\mathbb{P}(Y = i) = p_i, \quad \forall i \in \mathbb{N},$$

où les p_i sont des nombres positifs tels que $\sum_{i=0}^{\infty} p_i = 1$. Alors, d'après le chapitre 3, la fonction de répartition de Y est donnée, pour tout $t \geq 0$, par

$$F_Y(t) = \sum_{i=0}^{[t]} p_i$$

où l'on rappelle que $[t]$ désigne la partie entière de t . On en déduit que sa fonction de répartition inverse G_Y est donnée par : pour tout $u \in]0, 1[$

$$G_Y(u) = \begin{cases} 0 & \text{si } u \leq p_0 \\ 1 & \text{si } p_0 < u \leq p_0 + p_1 \\ 2 & \text{si } p_0 + p_1 < u \leq p_0 + p_1 + p_2 \\ \dots & \\ i & \text{si } \sum_{k=0}^{i-1} p_k < u \leq \sum_{k=0}^i p_k. \end{cases}$$

Exercice B.1. Considérons X une variable aléatoire de Cauchy, c'est-à-dire de loi absolument continue de densité $f_X : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$f_X(y) = \frac{1}{\pi(1 + y^2)}.$$

Montrer que pour tout $u \in]0, 1[$, $G_Y(u) = \tan(\pi(u - 1/2))$.

Remarque B.7. Il n'existe pas de formule explicite pour la fonction réciproque généralisée d'une variable aléatoire de loi gaussienne.

B.4.2 Méthode de la transformée réciproque généralisée

La méthode de simulation d'une variable aléatoire que nous proposons à présent est basée sur la connaissance de la fonction de répartition réciproque généralisée définie dans la partie précédente. Elle s'appuie sur la proposition suivante.

Proposition B.5.

Soit Y une variable aléatoire à valeurs réelles et de loi μ . Alors, étant donnée une variable aléatoire U de loi uniforme sur $[0, 1]$,

$$X = G_Y(U) \text{ est une variable aléatoire de loi } \mu,$$

où G_Y est la fonction de répartition réciproque généralisée de Y . En particulier, X et Y ont même loi.

Remarque B.8. En général, la fonction G_Y n'est pas définie en $u = 0$ et en $u = 1$. Cependant, comme une variable aléatoire U de loi uniforme sur $[0, 1]$ vérifie $\mathbb{P}(U \in \{0, 1\}) = 0$, la variable aléatoire $X = G_Y(U)$ est bien définie presque sûrement. Rappelons qu'en Python par exemple, la fonction `random()` peut renvoyer la valeur 0 (mais pas la valeur 1). Pour éviter une erreur en cas de non définition de G_Y en 0 (ce qui n'est pas le cas dans les exemples ci-dessous), on pourra par exemple utiliser la fonction `myRandom()` définie dans la section B.1.

Démonstration. Notons $F_X : \mathbb{R} \rightarrow \mathbb{R}$ la fonction de répartition de X et supposons pour simplifier que la fonction de répartition F_Y de Y est bijective de $\mathbb{R}$ sur $]0, 1[$. Dans ce cas, G_Y est la réciproque de la restriction de F_Y . De plus, la fonction F_Y est croissante, pour tout $u \in]0, 1[$ et $t \in \mathbb{R}$,

$$G_Y(u) \leq t \iff F_Y(t) \geq u.$$

Par suite, pour tout $t \in \mathbb{R}$,

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(G_Y(U) \leq t) = \mathbb{P}(U \leq F_Y(t)) = F_Y(t)$$

car U suit la loi uniforme sur $]0, 1[$ et $F_Y(t) \in]0, 1[$. Ainsi, X a même fonction de répartition que Y et donc même loi μ . ■

Exemple B.7. Soient $-\infty < a < b < +\infty$. Nous avons vu que la fonction de répartition réciproque généralisée d'une variable aléatoire Y de loi uniforme sur $[a, b]$ est donnée par

$$G_Y(u) = a + (b - a)u, \quad \forall u \in]0, 1[.$$

Par conséquent, étant donnée une variable aléatoire U de loi uniforme sur $[0, 1]$, la variable aléatoire

$$X = a + (b - a)U$$

est de loi uniforme sur $[a, b]$.

Voici le code correspondant en Matlab et en Python.

Matlab
<pre>a=5;b=10; U=rand; X=a+(b-a)*U;</pre>
Python
<pre>a,b=5,10 U=random() X=a+(b-a)*U</pre>

Exemple B.8. Soit $\lambda > 0$. Nous avons vu que la fonction de répartition d'une variable aléatoire Y de loi exponentielle de paramètre $\lambda > 0$ s'écrit

$$G_Y(u) = -\lambda \ln(1 - u), \quad \forall u \in]0, 1[.$$

Par conséquent, étant donnée une variable aléatoire U de loi uniforme sur $[0, 1]$, la variable aléatoire

$$X = -\lambda \ln(1 - U)$$

est de loi exponentielle de paramètre λ .

Voici le code correspondant en Matlab et en Python.

Matlab
<pre>lambda=1; U=rand; X=-lambda*log(1-U);</pre>
Python
<pre>lambda=1 #lambda est un mot-clef en Python U=random() X=-lambda*log(1-U)</pre>

Exemple B.9. Dans le cas d'une loi discrète μ , on s'aperçoit aisément que la méthode décrite dans la proposition B.2 de la section B.3.2 consiste exactement à simuler une variable aléatoire X de loi μ en posant

$$X = G_Y(U),$$

où U est une variable aléatoire de loi uniforme sur $[0, 1]$ et G_Y est la fonction de répartition réciproque généralisée d'une variable aléatoire Y de loi μ .

B.5 Méthode de rejet

Nous proposons ici une méthode pour simuler une variable aléatoire X de densité f sur $\mathbb{R}^d$ à partir d'une autre variable aléatoire Y de densité g . En pratique, nous supposons savoir comment simuler Y et calculer f et g .

Remarque B.9. On aura en tête que f et g sont les densités des lois de variables aléatoires X et Y par rapport à la mesure de Lebesgue. Toutefois, les méthodes décrites dans cette partie fonctionnent également si les lois de X et Y admettent une densité par rapport à une même mesure de référence ν .

Pour pouvoir utiliser cette méthode, on suppose qu'il existe une constante $c > 0$ telle que

$$f(x) \leq cg(x) \quad \forall x \in \mathbb{R}^d.$$

On pose alors pour tout $x \in \mathbb{R}^d$ tel que $g(x) \neq 0$,

$$h(x) = \frac{f(x)}{cg(x)}.$$

Étant donnée une variable aléatoire Y absolument continue de densité g ,

$$\mathbb{P}(g(Y) = 0) = \int_{\mathbb{R}^d} \mathbf{1}_{g(y)=0} g(y) \lambda_d(dy) = \int_{\mathbb{R}^d} 0 d\lambda_d = 0.$$

Ainsi, $g(Y) \neq 0$ presque sûrement et dans l'étape 2 donnée ci-après, la variable aléatoire $h(Y)$ est bien définie presque sûrement.

Descriptif de la méthode.

L'algorithme de la méthode du rejet est le suivant.

- **Étape 1** : on simule deux variables aléatoires indépendantes Y et U , avec Y de densité g et U une variable aléatoire de loi uniforme sur $[0, 1]$.
- **Étape 2** : si $U \leq h(Y)$, on pose $X = Y$. Sinon, on revient à l'étape 1 (les nouvelles variables simulées devront être indépendantes de celles déjà simulées).

Exemple B.10. Nous souhaitons simuler une variable aléatoire X à valeurs dans $\mathbb{R}^d$ de loi uniforme sur le disque de centre $(0, 0)$ et de rayon 1, c'est-à-dire de loi absolument continue de densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ donnée par

$$f(x, y) = \frac{1}{\pi} \mathbf{1}_{x^2+y^2 \leq 1}, \quad \forall (x, y) \in \mathbb{R}^2.$$

Pour cela nous allons utiliser des variables aléatoires dans $\mathbb{R}^2$ de loi uniforme sur le carré $[0, 1] \times [0, 1]$, loi de densité $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ donnée par

$$g(x, y) = \frac{1}{4} \mathbf{1}_{|x| \leq 1, |y| \leq 1}, \quad \forall (x, y) \in \mathbb{R}^2.$$

Comme $f(x, y) \leq \frac{4}{\pi} g(x, y)$, on pose, pour tout $(x, y) \in [0, 1] \times [0, 1]$,

$$h(x, y) = \frac{\pi f(x, y)}{4g(x, y)} = \mathbf{1}_{x^2+y^2 \leq 1}.$$

Enfin, on remarque qu'une variable aléatoire Y de loi uniforme sur le carré $[0, 1] \times [0, 1]$ peut être simulée par $Y = (Y_1, Y_2)$, où Y_1 et Y_2 sont deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Ce qui nous permet de mettre au point le code suivant pour simuler une variable aléatoire de loi uniforme sur le disque.

Matlab

```
h= @(x,y) (x^2+y^2)<=1;
U=rand;
Y1=rand;Y2=rand;
while U>h(Y1,Y2)
    U=rand;
    Y1=rand;Y2=rand;
end
X=[Y1,Y2];
```

Python

```
h=lambda x,y : x**2+y**2 <= 1
U=random()
Y1,Y2=random(), random()
while U>h(Y1,Y2):
    U=random()
    Y1,Y2=random(), random()
X=[Y1,Y2]
```

Voici une généralisation de ce calcul si l'on souhaite simuler une variable aléatoire de loi uniforme sur la boule de $\mathbb{R}^d$ de centre 0 et de rayon 1, avec $d \geq 1$. Remarquons que, si l'on change h en la fonction indicatrice d'un autre ensemble (par exemple d'une ellipse ou de l'intersection de deux boules), alors ce code permet de simuler une variable aléatoire uniforme sur cet ensemble.

Matlab

```
h= @(Y) sum(Y.^2)<=1;  
d=10; %d=dimension de l'espace  
U=rand;  
Y=rand(1,d);  
while U>h(Y)  
    U=rand;  
    Y=rand(1,d);  
end  
X=Y;
```

Python

```
h=lambda Y : sum([y**2 for y in Y])<=1  
d=10 #d=dimension de l'espace  
U=random()  
Y=[random() for _ in xrange(d)]  
while U>h(Y):  
    U=random()  
    Y=[random() for _ in xrange(d)]  
X=Y
```