

Probabilités

Master 1
Université de Paris

Notes de Pierre Youssef
reprises et complétées par Francis Comets puis Cyril Labbé

Année universitaire 2021/2022
Premier semestre

Table des matières

1	Espaces mesurés et probabilités.	5
I	Espaces mesurables	5
II	Mesure et probabilité	8
III	Classe monotone et unicité de la mesure engendrée	10
IV	Construction de mesure et mesure de Lebesgue	11
V	Espace de probabilité et événements	13
2	Intégration et mesure produit	17
I	Application mesurable	17
II	Intégration de fonctions positives	19
III	Intégrale dans le cas général	23
IV	Tribu et mesure produit	26
3	Variable aléatoire	31
I	Définition d'une variable aléatoire	31
II	Loi d'une variable aléatoire	33
III	Fonction de répartition	35
IV	Espérance et loi	38
V	Moments et inégalités classiques	39
VI	Fonction caractéristique	42
4	Indépendance	47
I	Événements indépendants et tribus indépendantes	47
II	Variables aléatoires indépendantes	49
III	Deuxième lemme de Borel-Cantelli	51
IV	Loi du tout ou rien	53
V	Processus de Poisson	54
5	Espérance conditionnelle	57
I	Conditionnement par rapport à un événement	57
II	Conditionnement par rapport à une variable discrète	58

Chapitre 1

Espaces mesurés et probabilités.

Rappelons que l'intégrale de Riemann d'une fonction $f : I \rightarrow \mathbb{R}$ est construite par approximations : on approxime les valeurs prises par f sur des "petits" sous-intervalles $[x_n, x_{n+1}[$ de I

$$\int f \approx \sum_n f(x_n) |x_{n+1} - x_n|.$$

Cela nécessite une certaine régularité de la fonction f afin que les approximations soient bonnes localement.

La théorie de l'intégration de Borel et Lebesgue repose sur un changement de point de vue fondamental : au lieu de découper l'intervalle I en petits sous-intervalles et d'évaluer la valeur de f sur ceux-ci, on découpe l'espace d'arrivée \mathbb{R} en petits sous-intervalles $[y_n, y_{n+1}[$ et on évalue la *mesure* de leurs pré-images par f :

$$\int f \approx \sum_n y_n \mu(f^{-1}([y_n, y_{n+1}[)).$$

Cela requiert d'avoir construit une mesure sur l'espace de départ (ce qui n'est pas une mince affaire en général), mais cela relâche très fortement la contrainte de régularité de la fonction f : celle-ci va devenir une contrainte de *mesurabilité*. Par ailleurs, cela permet de considérer des fonctions définies sur des espaces "quelconques" (pour peu qu'ils soient munis d'ensembles mesurables et d'une mesure) alors que l'intégrale de Riemann est intimement liée à l'ensemble \mathbb{R} .

Ce chapitre présente les notions sous-jacentes (tribus, mesures, etc.) à cette théorie de l'intégration, et le chapitre suivant présente la construction de l'intégrale.

I Espaces mesurables

Commençons par rappeler la notion de σ -algèbre (ou tribu) qui prescrit les ensembles mesurables.

Définition 1.1 (σ -algèbre)

Soit S un ensemble. On dit qu'une collection Σ de sous-ensembles de S est une σ -algèbre si les propriétés suivantes sont vérifiées :

- (i) $S \in \Sigma$.
- (ii) $A \in \Sigma \Rightarrow A^c := S \setminus A \in \Sigma$.
- (iii) $A_n \in \Sigma \forall n \in \mathbb{N} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \Sigma$.

Remarque 1.2

Considérons les assertions

(i') $\emptyset \in \Sigma$.

(iii') $A_n \in \Sigma \forall n \in \mathbb{N} \Rightarrow \bigcap_{n \in \mathbb{N}} A_n \in \Sigma$.

La définition de σ -algèbre reste inchangée si l'on impose (i') au lieu de (i) et/ou (iii') au lieu de (iii). En effet, (i) (resp. (i')) combinée avec (ii) implique (i') (resp. (i)). De même, si (ii) et (iii) sont vérifiés alors pour toute suite $A_n \in \Sigma$, on a $A_n^c \in \Sigma$ par (ii), et en combinant (ii) et (iii) avec l'identité

$$\bigcap_{n \in \mathbb{N}} A_n = \left(\bigcup_{n \in \mathbb{N}} A_n^c \right)^c,$$

on obtient (iii'). Un raisonnement analogue permet d'obtenir (iii) à partir de (ii) et (iii').

Finalement, notons que si la réunion dénombrable dans la propriété (iii) est remplacée par réunion finie, on dit dans ce cas que Σ est une algèbre. Ainsi une σ -algèbre est une algèbre mais la réciproque n'est pas nécessairement vraie.

On utilise parfois le terme *tribu* à la place de σ -algèbre.

Exemple 1.3

1. Etant donné un ensemble S , on peut définir $\{\emptyset, S\}$ qui vérifie les trois propriétés d'une σ -algèbre.
2. L'ensemble $\mathcal{P}(S)$ de toutes les parties de S est une σ -algèbre. On peut noter que toute σ -algèbre Σ sur S est située entre les deux précédentes i.e. $\{\emptyset, S\} \subset \Sigma \subset \mathcal{P}(S)$.

Définition 1.4 (Espace mesurable)

On dit que la paire (S, Σ) est un espace mesurable si S est un ensemble et Σ est une σ -algèbre sur S . On dit que les éléments de Σ sont les sous-ensembles de S qui sont Σ -mesurables.

Il semblerait naturel de toujours considérer la tribu $\mathcal{P}(S)$. Il se trouve que cette tribu est souvent trop grosse pour y construire des mesures intéressantes. On est alors amené à considérer des tribus engendrées par des classes d'ensembles simples (les intervalles de \mathbb{R} par exemple).

Définition 1.5 (σ -algèbre engendrée par une classe)

Soit \mathcal{C} une classe de sous-ensembles de S . La σ -algèbre engendrée par \mathcal{C} , notée $\sigma(\mathcal{C})$, est l'intersection de toutes les σ -algèbres contenant \mathcal{C} . C'est la plus petite σ -algèbre sur S qui contient \mathcal{C} .

Cette définition se base sur la propriété suivante : l'intersection d'un nombre quelconque (pas forcément fini ni dénombrable) de σ -algèbres est une σ -algèbre. La preuve est laissée en exercice.

Exemple 1.6

1. Soit $S = \{1, 2, 3, 4\}$ et $\mathcal{C} = \{\{1\}, \{2, 3\}\}$. Pour former $\sigma(\mathcal{C})$, on commence par mettre l'ensemble vide, S puis les éléments de \mathcal{C} . Après, on effectue les opérations de passage au complémentaire, l'union et l'intersection des éléments à disposition. Après avoir effectué toutes les opérations possibles, on vérifie que l'ensemble obtenu est bien une σ -algèbre. Dans cet exemple, on doit rajouter $\{1\}^c = \{2, 3, 4\}$, $\{2, 3\}^c = \{1, 4\}$, $\{1, 4\} \cap \{2, 3, 4\} = \{4\}$, et $\{4\}^c = \{1, 2, 3\}$ pour avoir

$$\sigma(\mathcal{C}) = \{\emptyset, S, \{1\}, \{2, 3\}, \{1, 2, 3\}, \{4\}, \{1, 4\}, \{2, 3, 4\}\}.$$

2. Un exemple important est la σ -algèbre Borelienne. Si S est un espace topologique, la σ -algèbre de Borel $\mathcal{B}(S)$ est la σ -algèbre engendrée par les ouverts de S . Ainsi $\mathcal{B} := \mathcal{B}(\mathbb{R})$ est la σ -algèbre de Borel sur \mathbb{R} i.e. la σ -algèbre engendrée par les ouverts de \mathbb{R} .

Au lieu de travailler avec tous les ouverts de \mathbb{R} , on sélectionnera un système générateur qui est plus explicite. Pour cela, on définit la classe suivante d'intervalles

$$I(\mathbb{R}) := \{(-\infty, x] : x \in \mathbb{R}\}$$

et on va vérifier qu'elle suffit pour obtenir tous les éléments de la σ -algèbre de Borel sur \mathbb{R} i.e. on va montrer que $\mathcal{B} = \sigma(I(\mathbb{R}))$. On procède par double inclusion, tout en notant d'abord que pour tout $x \in \mathbb{R}$ on a $(-\infty, x] = \bigcap_{n \in \mathbb{N}} (-\infty, x + 1/n)$ qui est une intersection dénombrable d'ouverts et ainsi $(-\infty, x] \in \mathcal{B}$. Puisque $\sigma(I(\mathbb{R}))$ est la plus petite σ -algèbre qui contient $I(\mathbb{R})$ alors $\sigma(I(\mathbb{R})) \subseteq \mathcal{B}$.

Pour l'autre inclusion, il suffit de montrer que tout intervalle ouvert de \mathbb{R} peut être généré à partir des éléments de $I(\mathbb{R})$. Soit donc $a < b$, et remarquons que puisque $(a, b) = \bigcup_{n \in \mathbb{N}} (a, b - 1/n]$ il nous suffit de montrer qu'un intervalle de la forme $(a, u]$ peut être généré à partir des éléments de $I(\mathbb{R})$. Mais on voit que

$$(a, u] = (-\infty, u] \cap (-\infty, a]^c \in \sigma(I(\mathbb{R})),$$

ce qui montre ce qu'on cherchait.

Terminons cette section en introduisant des définitions relatives aux suites d'ensembles. Pour une suite $(A_n)_{n \in \mathbb{N}} \in \Sigma$, on écrira $A_n \uparrow A$ si $A_n \subseteq A_{n+1}$ pour tout $n \in \mathbb{N}$ et $\bigcup_{n \in \mathbb{N}} A_n = A$. De même, on écrira $A_n \downarrow A$ si $A_{n+1} \subseteq A_n$ pour tout $n \in \mathbb{N}$ et $\bigcap_{n \in \mathbb{N}} A_n = A$.

On va maintenant introduire les notions de limite sup et limite inf. Rappelons les définitions dans le cas d'une suite de réels : étant donnée une suite de réels $(x_n)_{n \in \mathbb{N}}$, les limites inf et sup sont données par

$$\liminf x_n := \sup_{m \geq 0} \inf_{n \geq m} x_n = \lim_{m \rightarrow \infty} [\inf_{n \geq m} x_n] \quad \text{et} \quad \limsup x_n := \inf_{m \geq 0} \sup_{n \geq m} x_n = \lim_{m \rightarrow \infty} [\sup_{n \geq m} x_n].$$

On dit que la limite de la suite x_n existe lorsque les deux quantités précédentes coïncident. En faisant l'analogie entre inf et \cap d'une part, et entre sup et \cup de l'autre, on définit les limite sup et inf d'ensembles comme suit.

Définition 1.7

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'ensemble. La limite sup de A_n , notée $\limsup A_n$, est donnée par

$$\limsup A_n = \bigcap_{m \geq 0} \bigcup_{n \geq m} A_n.$$

De même, la limite inf de A_n , notée $\liminf A_n$, est donnée par

$$\liminf A_n = \bigcup_{m \geq 0} \bigcap_{n \geq m} A_n.$$

Remarque 1.8

1. Il est important de comprendre le sens de $\limsup A_n$. Dire que $\omega \in \limsup A_n$ équivaut à dire que

$$\forall m, \exists n \geq m \text{ tel que } \omega \in A_n.$$

C'est-à-dire ω appartient à une infinité d'ensembles A_n . On gardera donc en tête l'équivalence

$$\omega \in \limsup A_n \Leftrightarrow \omega \text{ appartient à une infinité des } A_n.$$

2. Dire que $\omega \in \liminf A_n$ équivaut à dire que

$$\exists m, \forall n \geq m \text{ on a } \omega \in A_n.$$

Ainsi

$$\omega \in \liminf A_n \Leftrightarrow \omega \text{ appartient à tous les } A_n \text{ à partir d'un certain rang.}$$

3. Exercice : Vérifier que $\liminf A_n \subseteq \limsup A_n$.
 4. Exercice : Montrer que

$$\mathbf{1}_{\limsup A_n} = \limsup \mathbf{1}_{A_n} \quad \text{et} \quad \mathbf{1}_{\liminf A_n} = \liminf \mathbf{1}_{A_n},$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de A .

II Mesure et probabilité

On est maintenant prêt à introduire la notion de mesure.

Définition 1.9 (Mesure)

Soit (S, Σ) un espace mesurable. Une mesure μ sur (S, Σ) est une fonction

$$\mu : \Sigma \rightarrow [0, \infty]$$

qui satisfait les propriétés suivantes :

- (i) $\mu(\emptyset) = 0$.
- (ii) Si $(A_n)_{n \in \mathbb{N}}$ est une suite d'ensembles disjoints dans Σ alors $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

On dit alors que le triplet (S, Σ, μ) est un espace mesuré.

Si de plus $\mu(S) = 1$ alors on dit que μ est une mesure de probabilité et que (S, Σ, μ) est un espace de probabilité.

Remarque 1.10

1. La propriété (ii) ci-dessus est connue comme la σ -additivité de la mesure. Si l'union dénombrable est remplacée par l'union finie, on parle alors tout simplement d'additivité.
2. Notons que le fait que Σ soit une σ -algèbre donne un sens aux propriétés (i) et (ii) puisque $\emptyset \in \Sigma$ et on peut donc lui appliquer μ et de même, si les A_n sont dans Σ alors l'union l'est aussi et on peut parler de $\mu(\bigcup_{n \in \mathbb{N}} A_n)$.
3. On dit que μ est finie si $\mu(S) < \infty$. Ainsi une mesure de probabilité est une mesure finie.
4. On dit que μ est σ -finie s'il existe une suite $(S_n)_{n \in \mathbb{N}}$ d'éléments de Σ telle que $\bigcup_{n \in \mathbb{N}} S_n = S$

et $\mu(S_n) < \infty$ pour tout $n \in \mathbb{N}$.

Exemple 1.11

1. (Mesure nulle) Soit (S, Σ) un espace mesurable. $\mu : \Sigma \rightarrow [0, \infty]$ définie par $\mu(A) = 0$ pour tout $A \in \Sigma$ est une mesure.
2. (Mesure de comptage) Le cardinal d'un ensemble est une mesure sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.
3. (Mesure de Dirac) Soit (S, Σ) un espace mesurable et $x \in S$. On définit $\delta_x : \Sigma \rightarrow [0, \infty]$ par $\delta_x(A) = 1$ si $x \in A$ et 0 sinon. On vérifie que δ_x est une mesure de probabilité qu'on appelle la mesure de Dirac.

Voici quelques propriétés élémentaires des mesures.

Proposition 1.12

Soit (S, Σ, μ) un espace mesuré. Alors on a

1. $\mu(A \cup B) \leq \mu(A) + \mu(B)$ pour tous $A, B \in \Sigma$.
2. $\mu(\bigcup_{i \geq 1} A_i) \leq \sum_{i \geq 1} \mu(A_i)$ pour tous $A_1, A_2, \dots \in \Sigma$.
3. Si $\mu(S) < \infty$ alors pour tous $A, B \in \Sigma$ on a

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

Plus généralement on a la formule d'inclusion-exclusion suivante

$$\mu\left(\bigcup_{i \leq n} A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mu\left(\bigcap_{j \leq k} A_{i_j}\right),$$

pour tous $A_1, \dots, A_n \in \Sigma$.

Démonstration Exercice. □

Une propriété utile des fonctions est la continuité : l'une de ses conséquences est de pouvoir affirmer que $\lim_n f(x_n) = f(\lim_n x_n)$. On aimerait une propriété similaire pour les mesures.

Lemme 1.13

Soit (S, Σ, μ) un espace mesuré et $(A_n)_{n \in \mathbb{N}} \in \Sigma$.

1. Si $A_n \uparrow A$ alors $\mu(A_n) \uparrow \mu(A)$.
2. Si $A_n \downarrow A$ et $\mu(A_k) < \infty$ pour un certain $k \in \mathbb{N}$, alors $\mu(A_n) \downarrow \mu(A)$.

Démonstration

1. Soit $B_1 = A_1$, $B_2 = A_2 \setminus A_1$ et plus généralement $B_n = A_n \setminus A_{n-1}$ pour tout $n \geq 2$. On remarque que les B_n sont disjoints et que $A_n = \bigcup_{i=1}^n B_i$. Ainsi

$$\mu(A_n) = \mu(B_1 \cup \dots \cup B_n) = \sum_{i=1}^n \mu(B_i) \xrightarrow{n \rightarrow \infty} \sum_{i=1}^{\infty} \mu(B_i) = \mu(A).$$

2. Soit $B_n = A_k \setminus A_{k+n}$ pour tout $n \in \mathbb{N}$. Alors $B_n \uparrow A_k \setminus \bigcap_{\ell \in \mathbb{N}} A_\ell$ et par la première partie on a

$$\mu(B_n) = \mu(A_k) - \mu(A_{k+n}) \xrightarrow{n \rightarrow \infty} \mu\left(A_k \setminus \bigcap_{\ell \in \mathbb{N}} A_\ell\right) = \mu(A_k) - \mu\left(\bigcap_{\ell \in \mathbb{N}} A_\ell\right).$$

Il suffit alors de retrancher des deux côtés $\mu(A_k)$, qui est finie.

□

III Classe monotone et unicité de la mesure engendrée

Etant donnée une tribu Σ engendrée par une classe \mathcal{C} , on peut se demander si toute mesure μ sur Σ est entièrement caractérisée par les valeurs qu'elle prend sur \mathcal{C} ? Autrement dit, si deux mesures μ et ν coïncident sur \mathcal{C} , coïncident-elles sur $\sigma(\mathcal{C})$?

Nous allons apporter une réponse à cette question en introduisant le Lemme de classe monotone (qui a de nombreuses applications). Commençons par définir la notion de classe monotone.

Définition 1.14 (Classe monotone)

Soit S un ensemble. On dit qu'une famille \mathcal{M} de sous-ensembles de S est une classe monotone si

- (i) $S \in \mathcal{M}$.
- (ii) Si $A, B \in \mathcal{M}$ et $A \subset B$, alors $B \setminus A \in \mathcal{M}$.
- (iii) Si $A_n \in \mathcal{M}$ et $A_n \subset A_{n+1}$ alors $\bigcup_n A_n \in \mathcal{M}$.

Remarquons que toute σ -algèbre est une classe monotone. En revanche, il existe des classes monotones qui ne sont pas des σ -algèbres. Par exemple sur $S = \{1, 2, 3, 4\}$, on peut considérer $\mathcal{M} = \{\emptyset, S, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$.

Par ailleurs, une intersection quelconque de classes monotones est une classe monotone. En conséquence, on peut parler de la plus petite classe monotone $\mathcal{M}(\mathcal{I})$ contenant une famille \mathcal{I} de sous-ensembles de S : il s'agit de l'intersection de toutes les classes monotones contenant \mathcal{I}

$$\mathcal{M}(\mathcal{I}) = \bigcap_{\mathcal{M} \text{ classe monotone, } \mathcal{I} \subset \mathcal{M}} \mathcal{M}.$$

Théorème 1.15 (Lemme de classe monotone)

Soit S un ensemble et $\mathcal{I} \subset \mathcal{P}(S)$ une famille de sous-ensembles de S stable par intersection finie. Alors toute classe monotone \mathcal{M} qui contient \mathcal{I} contient $\sigma(\mathcal{I})$. En d'autres termes, $\sigma(\mathcal{I}) = \mathcal{M}(\mathcal{I})$.

Remarque 1.16

Une famille stable par intersection finie est parfois appelée un π -système.

Démonstration Comme $\sigma(\mathcal{I})$ est une classe monotone (puisque c'est une σ -algèbre), alors $\mathcal{M}(\mathcal{I}) \subset \sigma(\mathcal{I})$. Pour montrer l'inclusion inverse, il nous suffit de montrer que $\mathcal{M}(\mathcal{I})$ est une σ -algèbre puisque $\sigma(\mathcal{I})$ est la plus petite σ -algèbre contenant \mathcal{I} .

On a $S \in \mathcal{M}(\mathcal{I})$ par définition. Soit $A \in \mathcal{M}(\mathcal{I})$ alors $A \in \mathcal{M}$ pour toute classe monotone \mathcal{M} contenant \mathcal{I} . Comme \mathcal{M} est une classe monotone et que $A \subset S$ appartient à \mathcal{M} , alors il en est de même pour $A^c = S \setminus A$.

Il reste donc à montrer que $\mathcal{M}(\mathcal{I})$ est stable par réunion dénombrable. Par la propriété (iii) des classes monotones, il nous suffit de montrer que $\mathcal{M}(\mathcal{I})$ est stable par réunion finie. Puisque $\mathcal{M}(\mathcal{I})$ est stable par passage au complémentaire, il nous suffit de vérifier que $\mathcal{M}(\mathcal{I})$ est stable par intersection finie. Pour cela, il suffit de montrer que si $A, B \in \mathcal{M}(\mathcal{I})$ alors $A \cap B \in \mathcal{M}(\mathcal{I})$.

Supposons d'abord que $A \in \mathcal{I}$. Soit $\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{I}) : A \cap B \in \mathcal{M}(\mathcal{I})\}$. Puisque \mathcal{I} est un π -système on a $\mathcal{I} \subset \mathcal{M}_A$. On va vérifier que \mathcal{M}_A est une classe monotone pour déduire que $\mathcal{M}(\mathcal{I}) \subset \mathcal{M}_A$. On a

- (i) $S \cap A = A \in \mathcal{I} \subset \mathcal{M}(\mathcal{I})$ donc $S \in \mathcal{M}_A$.
- (ii) Soit $B, B' \in \mathcal{M}_A$ avec $B \subset B'$. Alors $(B' \setminus B) \cap A = (B' \cap A) \setminus (B \cap A) \in \mathcal{M}(\mathcal{I})$ puisque $B' \cap A \in \mathcal{M}(\mathcal{I})$, $B \cap A \in \mathcal{M}(\mathcal{I})$ et $\mathcal{M}(\mathcal{I})$ est une classe monotone.
- (iii) Si B_n est une suite croissante de \mathcal{M}_A , alors $B_n \cap A$ est une suite croissante de $\mathcal{M}(\mathcal{I})$. Comme $\mathcal{M}(\mathcal{I})$ est une classe monotone, alors $\bigcup (B_n \cap A) \in \mathcal{M}(\mathcal{I})$. Puisque $\bigcup (B_n \cap A) = A \cap \bigcup B_n$ on en déduit que $\bigcup B_n \in \mathcal{M}_A$.

Soit maintenant

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

On vient de montrer ci-dessus que $\mathcal{I} \subset \mathcal{M}_1$. On vérifie comme avant que \mathcal{M}_1 est une classe monotone, ce qui implique que $\mathcal{M}(\mathcal{I}) \subset \mathcal{M}_1$. Ceci montre donc que pour tout $A, B \in \mathcal{M}(\mathcal{I})$, on a $A \cap B \in \mathcal{M}(\mathcal{I})$ et finit la preuve. \square

A l'aide de ce lemme, on peut apporter une réponse à la question posée plus haut.

Corollaire 1.17 (Unicité d'une mesure)

Soit S un ensemble, \mathcal{I} une famille de sous-ensembles de S stable par intersection finie et $\Sigma = \sigma(\mathcal{I})$. Si deux mesures finies μ_1, μ_2 sur (S, Σ) coïncident sur \mathcal{I} (i.e. $\mu_1(A) = \mu_2(A)$ pour tout $A \in \mathcal{I}$) et que $\mu_1(S) = \mu_2(S)$ alors elles coïncident sur Σ .

Le résultat reste vrai si μ et ν sont seulement σ -finies et sont telles que $\mu(E_n) = \nu(E_n) < \infty$ pour une certaine suite croissante E_n telle que $\cup_n E_n = S$.

Démonstration Nous allons montrer que $\mathcal{F} = \{A \in \Sigma : \mu_1(A) = \mu_2(A)\}$ est une classe monotone. En effet, ceci impliquera par le Lemme de classe monotone que $\sigma(\mathcal{I}) = \mathcal{M}(\mathcal{I}) \subset \mathcal{F}$ et que donc $\mathcal{F} = \Sigma$ pour finir la preuve.

Pour vérifier que c'est une classe monotone, on note d'abord que par hypothèse on a $\mu_1(S) = \mu_2(S)$. D'autre part, si $A, B \in \mathcal{F}$ et que $A \subset B$ alors $\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A)$. Notons que c'est ici qu'on a utilisé le fait que les mesures sont finies.

Il reste à montrer que \mathcal{F} est stable par réunion croissante. Soient alors $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, avec $A_n \subset A_{n+1}$. On a par le Lemma 1.13.

$$\mu_1\left(\bigcup_n A_n\right) = \lim_n \mu_1(A_n) = \lim_n \mu_2(A_n) = \mu_2\left(\bigcup_n A_n\right).$$

\square

On observera que la classe \mathcal{F} introduite dans la preuve précédente n'est pas une tribu en général : en effet, si μ et ν coïncident sur A et B , il n'y a aucune raison qu'elles coïncident sur $A \cup B$ (sauf si A et B sont disjoints, ou que l'un est inclus dans l'autre).

IV Construction de mesure et mesure de Lebesgue

Nous venons de voir un résultat d'unicité sur les mesures, la preuve repose sur le lemme de classe monotone qui est relativement simple à prouver. L'existence de mesure (vérifiant certaines propriétés sur des familles simples d'ensembles) est en général nettement plus difficile à établir. En

particulier, l'existence d'une mesure qui assigne aux intervalles leur longueur (la mesure de Lebesgue) n'est pas triviale. En effet, une telle mesure devra être définie sur la plus petite tribu contenant les intervalles (la tribu Borélienne!), or la plupart des ensembles de cette tribu sont "compliqués" et leur "longueur" n'est pas une quantité évidente à définir.

Nous allons présenter un résultat général assurant l'existence d'une mesure comme prolongement d'une "pré-mesure" (la longueur dans le cas de la mesure de Lebesgue). Commençons par introduire la notion d'algèbre.

Définition 1.18 (algèbre)

Soit S un ensemble. On dit qu'une collection Σ de sous-ensembles de S est une algèbre si les propriétés suivantes sont vérifiées :

- (i) $S \in \Sigma$.
- (ii) $A \in \Sigma \Rightarrow A^c := S \setminus A \in \Sigma$.
- (iii) $A, B \in \Sigma \Rightarrow A \cup B \in \Sigma$.

On a alors le résultat fondamental suivant.

Théorème 1.19 (Théorème d'extension de Carathéodory)

Soit S un ensemble, et soit Σ_0 une algèbre sur S . Si $\mu_0 : \Sigma_0 \rightarrow [0, \infty]$ satisfait :

- (i) $\mu_0(\emptyset) = 0$,
- (ii) $\mu_0(\cup_n A_n) = \sum_n \mu_0(A_n)$ pour toute suite $A_n \in \Sigma_0$ de sous-ensembles deux-à-deux disjoints tels que $\cup_n A_n \in \Sigma_0$,

alors il existe une extension de μ_0 à Σ i.e. il existe une mesure μ sur (S, Σ) telle que $\mu = \mu_0$ sur Σ_0 .

La morale de ce résultat est qu'il n'est pas nécessaire de définir une mesure sur toute la σ -algèbre, on peut se contenter de le faire sur une algèbre qui l'engendre. Sa preuve est délicate, elle repose sur l'introduction de la notion de mesure extérieure : on pourra consulter [BP00].

Par le corollaire précédent, si μ_0 est σ -finie alors l'extension est unique.

Prenons l'exemple de la mesure de Lebesgue sur $S = \mathbb{R}$ muni de la σ -algèbre de Borel $\mathcal{B}(\mathbb{R})$. Pour définir cette mesure, on va se contenter de la définir sur les réunions finies d'intervalles. On définit ainsi Σ_0 comme la collection de toutes les parties s'écrivant comme réunion finie d'intervalles disjoints. On peut vérifier que Σ_0 est une algèbre et que $\Sigma := \sigma(\Sigma_0) = \mathcal{B}(\mathbb{R})$. Au lieu de définir notre mesure λ sur tout Σ , on se contente de le faire sur Σ_0 en posant pour tout $A = I_1 \cup \dots \cup I_r \in \Sigma_0$

$$\lambda(A) = \sum_{k=1}^r |I_k|,$$

où $|I_k|$ est la longueur de l'intervalle I_k . On peut montrer que λ est σ -additive sur Σ_0 et le théorème de Carathéodory assure l'existence d'une extension à une mesure λ sur $\mathcal{B}(\mathbb{R})$. Cette mesure est appelée la *mesure de Lebesgue*. En outre, comme λ est σ -finie (on peut prendre $E_n := [-n, n]$) cette extension est unique.

Remarque 1.20

La σ -additivité sur Σ_0 n'est pas tout à fait triviale. Soit E_n une suite d'éléments disjoints de Σ_0 telle que $\cup_n E_n \in \Sigma_0$. Sans perte de généralité, on peut supposer que tous les E_n sont des intervalles, et qu'ils vivent dans un compact de \mathbb{R} (afin d'éviter la manipulation de quantités infinies). Des

considérations simples sur la longueur impliquent que pour tout $n \geq 1$

$$\sum_{i=1}^n \lambda(E_i) \leq \lambda(\cup_n E_n),$$

et ainsi $\sum_i \lambda(E_i) \leq \lambda(\cup_n E_n)$. L'autre inégalité est plus difficile. Par hypothèse, on peut écrire $\cup_n E_n$ comme la réunion finie d'intervalles disjoints I_1, \dots, I_N . On peut trouver des intervalles fermés K_1, \dots, K_N tels que $K_i \subset I_i$ et $\lambda(K_i) \geq \lambda(I_i) - \epsilon 2^{-i}$. Ainsi $K = \cup_{i=1}^N K_i$ est un compact tel que

$$\lambda(K) \geq \lambda(\cup_n E_n) - \epsilon.$$

Par ailleurs, on peut trouver des intervalles ouverts $O_n \supset E_n$ tels que $\lambda(O_n) \leq \lambda(E_n) + \epsilon 2^{-n}$. On a donc construit un recouvrement ouvert du compact K , dont on peut extraire un sous-recouvrement fini O_1, \dots, O_m . On trouve alors

$$\lambda(K) \leq \lambda(\cup_{i=1}^m O_i) \leq \sum_{i=1}^m \lambda(O_i) \leq \sum_{i=1}^m \lambda(E_n) + \epsilon \leq \sum_{i \geq 1} \lambda(E_n) + \epsilon.$$

Ainsi

$$\lambda(\cup_n E_n) \leq \sum_{i \geq 1} \lambda(E_n) + 2\epsilon.$$

V Espace de probabilité et événements

On se focalise dorénavant sur le cas des espaces de probabilité. Rappelons qu'un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace de probabilité si (Ω, \mathcal{A}) est un espace mesurable i.e. \mathcal{A} est une σ -algèbre sur Ω , et \mathbb{P} est une mesure de probabilité.

Dans ce cas, on appelle Ω l'univers, un élément $\omega \in \Omega$ un échantillon et \mathcal{A} la famille des événements.

Exemple 1.21

1. On lance une pièce de monnaie à deux reprises. Le résultat de notre expérience consiste en les paires de caractères *Pile*, *Face*. On pose donc

$$\Omega = \{PP, FF, PF, FP\}$$

et on prend $\mathcal{A} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties de Ω . On peut définir \mathbb{P} comme étant la probabilité uniforme sur Ω . Ainsi, l'événement "obtenir au moins une *Face*" est donné par le sous-ensemble $\{FF, PF, FP\}$ qui a une probabilité $3/4$.

2. Si notre expérience consiste à choisir un nombre réel au hasard entre 0 et 1. Alors l'univers est $\Omega = [0, 1]$ qu'on munit de la tribu borelienne $\mathcal{A} = \mathcal{B}([0, 1])$ et de la mesure de Lebesgue. Ainsi la probabilité que le nombre qu'on a choisi soit plus grand que $2/3$ est égale à $1/3$.

Définition 1.22 (*Presque sûrement*)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On dit qu'un énoncé Γ est vrai \mathbb{P} -presque-sûrement (ou simplement *p.s.*) si $C = \{\omega \in \Omega : \Gamma \text{ est vrai sur } \omega\} \in \mathcal{A}$ et $\mathbb{P}(C) = 1$.

Exemple 1.23

1. On reprend l'exemple de lancer deux fois une pièce. On considère l'énoncé Γ : "On obtient au moins une Pile ou une Face". On voit que $C = \{\omega \in \Omega : \Gamma \text{ est vrai sur } \omega\} = \{FF, PP, FP, PF\} = \Omega \in \mathcal{A}$ et que $\mathbb{P}(C) = 1$. Ainsi Γ est vrai p.s.
2. On reprend l'exemple de choisir un nombre au hasard dans $[0, 1]$. On fait l'énoncé Γ : "On obtient un nombre différent de $0, 1/2, 1$ ". On voit que $C = \{\omega \in [0, 1] : \Gamma \text{ est vrai sur } \omega\} = (0, 1) \setminus \{1/2\} \in \mathcal{B}([0, 1])$ et est de mesure de Lebesgue égale à 1. Ainsi Γ est vrai p.s.

Proposition 1.24

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Si les $A_n \in \mathcal{A}$ sont tels que $\mathbb{P}(A_n) = 1$ pour tout $n \in \mathbb{N}$, alors $\mathbb{P}(\bigcap_{n \in \mathbb{N}} A_n) = 1$. En d'autres termes, l'intersection dénombrable d'événements presque sûrs est presque sûre.

Démonstration On a $\mathbb{P}(A_n^c) = 0$ pour tout $n \in \mathbb{N}$ et ainsi $\mathbb{P}(\bigcup_{k=1}^n A_k^c) = 0$. Or $\bigcup_{k=1}^n A_k^c \uparrow \bigcup_{\ell \in \mathbb{N}} A_\ell^c$, ainsi $\mathbb{P}(\bigcup_{\ell \in \mathbb{N}} A_\ell^c) = \lim_{n \rightarrow \infty} \mathbb{P}(\bigcup_{k=1}^n A_k^c) = 0$ et donc $\mathbb{P}(\bigcap_{n \in \mathbb{N}} A_n) = 1$. \square

Proposition 1.25

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements, on a alors :

- (i) (Premier lemme de Borel-Cantelli) Si $\sum_n \mathbb{P}(A_n) < \infty$, alors

$$\mathbb{P}(\limsup A_n) = 0.$$

- (ii)

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n).$$

- (iii)

$$\limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

Démonstration

- (i) Posons $B = \limsup A_n$ et $B_m = \bigcup_{n \geq m} A_n$. Par les propriétés des mesures, on a $\mathbb{P}(B_m) \leq \sum_{n \geq m} \mathbb{P}(A_n) < \infty$. De plus, les B_m sont décroissants et $\bigcap_m B_m = B$, donc

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcap_m B_m\right) = \lim_m \mathbb{P}(B_m) \leq \lim_m \sum_{n \geq m} \mathbb{P}(A_n) = 0.$$

- (ii) Soit $C = \liminf A_n$ et $C_m = \bigcap_{n \geq m} A_n$. Les C_m étant croissants et $\bigcup_m C_m = C$, on a

$$\mathbb{P}(C) = \lim_m \mathbb{P}(C_m).$$

Or $\mathbb{P}(C_m) \leq \mathbb{P}(A_n)$ pour tout $n \geq m$ car $\bigcap_{n \geq m} A_n \subseteq A_n$. Ainsi

$$\mathbb{P}(C) = \lim_m \mathbb{P}(C_m) \leq \lim_m \inf_n \mathbb{P}(A_n) = \liminf \mathbb{P}(A_n).$$

- (iii) Posons $B = \limsup A_n$ et $B_m = \bigcup_{n \geq m} A_n$. Les B_m sont décroissants et \mathbb{P} est une mesure finie, alors

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_m B_m\right) = \lim_m \mathbb{P}(B_m).$$

Mais $B_m \supseteq A_n$ pour tout $n \geq m$, donc $\mathbb{P}(B_m) \geq \mathbb{P}(A_n)$ pour tout $n \geq m$. Ainsi

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcap_m B_m\right) = \lim_m \mathbb{P}(B_m) \geq \lim_m \sup_{n \geq m} \mathbb{P}(A_n) = \limsup \mathbb{P}(A_n).$$

□

On peut également prouver (iii) en utilisant (ii) sur les complémentaires.

Chapitre 2

Intégration et mesure produit

I Application mesurable

Dans le chapitre précédent, on est parti d'un ensemble quelconque que l'on a muni d'une σ -algèbre pour former un espace mesurable. En algèbre par exemple, on part d'un ensemble et on l'équipe d'une loi d'addition ou de produit. En topologie, on identifie les ouverts et fermés de l'espace (en le munissant d'une métrique par exemple). Dans tous ces domaines, l'étape suivante est de définir les applications entre ces différents espaces. Il s'agit d'avoir une application au vrai sens du terme mais qui tient compte de la structure sous-jacente. Par exemple, en algèbre, on parle de morphisme de groupes qui respecte la loi d'addition ou de produit en imposant que l'image de la somme (ou du produit) est la somme (ou produit) des images par cette application.

En topologie, on dit qu'une application est continue lorsqu'elle respecte le transfert de topologie entre l'espace de départ et l'espace d'arrivée. On demande ainsi que l'image réciproque d'un ouvert de l'ensemble d'arrivée soit un ouvert de l'ensemble de départ.

Définition 2.1 (Application mesurable)

Soient (S, Σ) et (F, \mathcal{F}) deux espaces mesurables. On dit que $h : S \rightarrow F$ est (Σ, \mathcal{F}) -mesurable si pour tout $A \in \mathcal{F}$, on a $h^{-1}(A) \in \Sigma$.

Ainsi, une application est mesurable si l'image réciproque de tout ensemble mesurable est mesurable.

Remarque 2.2

1. Dans la suite du cours, on prendra fréquemment $F = \mathbb{R}$ muni de la tribu Borélienne $\mathcal{F} = \mathcal{B}(\mathbb{R})$.
2. On dira qu'une application est mesurable, sans préciser les tribus sur les espaces d'arrivée et de départ, lorsqu'il n'y aura pas d'ambiguïté sur les tribus considérées. Parfois, on se contentera de préciser la tribu de départ : par exemple, $h : S \rightarrow \mathbb{R}$ est Σ -mesurable signifiera (la plupart du temps) que h est $(\Sigma, \mathcal{B}(\mathbb{R}))$ -mesurable.
3. Parfois, on sera amené à considérer des tribus sur $\overline{\mathbb{R}}$, le complété de \mathbb{R} , muni de sa topologie usuelle. Rappelons que dans cette topologie, les intervalles $]a, +\infty]$ et $[-\infty, b[$ sont des ouverts pour tout $a, b \in \mathbb{R}$. La tribu Borélienne sur $\overline{\mathbb{R}}$, notée $\mathcal{B}(\overline{\mathbb{R}})$, est la plus petite tribu contenant les ouverts de $\overline{\mathbb{R}}$.

Exemple 2.3

1. Soit $c \in F$ et $h : S \rightarrow F$ définie par $h(x) = c$ pour tout $x \in S$. Alors h est (Σ, \mathcal{F}) -mesurable,

quelles que soient les tribus Σ et \mathcal{F} considérées. En effet, pour tout ensemble $A \subset F$ on a : $h^{-1}(A) = S$ si $c \in A$, et $h^{-1}(A) = \emptyset$ si $c \notin A$, or S et \emptyset appartiennent à toute tribu sur S .

2. Soit $B \in \Sigma$, alors $h := \mathbf{1}_B$ est $(\Sigma, \mathcal{B}(\mathbb{R}))$ -mesurable. En effet, soit A un borélien de \mathbb{R} : si 0 et 1 appartiennent à A , alors $h^{-1}(A) = S \in \Sigma$; si $0 \in A$ et $1 \notin A$ alors $h^{-1}(A) = B^c \in \Sigma$; si $0 \notin A$ et $1 \in A$ alors $h^{-1}(A) = B \in \Sigma$; et finalement si 0 et 1 n'appartiennent pas à A alors $h^{-1}(A) = \emptyset \in \Sigma$.
3. On pose $S = F = \mathbb{R}$, $\Sigma = \{\emptyset, \mathbb{R}\}$, et $\mathcal{F} = \mathcal{B}(\mathbb{R})$. On définit h par $h(x) = 0$ si $x \geq 0$ et $h(x) = 3$ sinon. Alors h n'est pas $(\Sigma, \mathcal{B}(\mathbb{R}))$ -mesurable puisque $h^{-1}(\{0\}) = [0, \infty) \notin \Sigma$.

Les opérations de somme, produit, composition préservent la continuité ; on va montrer qu'il en est de même pour la mesurabilité. Avant cela, énonçons des faits élémentaires à propos de l'image réciproque.

Proposition 2.4

Soit $h : S \rightarrow F$.

- (i) On a $h^{-1}(\bigcup_n A_n) = \bigcup_n h^{-1}(A_n)$, $h^{-1}(A^c) = (h^{-1}(A))^c$, $h^{-1}(\bigcap A_n) = \bigcap h^{-1}(A_n)$. En d'autres termes, l'image réciproque préserve les opérations courantes sur les ensembles.
- (ii) Soit Σ une tribu sur S et \mathcal{C} une classe d'ensembles sur F . Si $h^{-1}(A) \in \Sigma$ pour tout $A \in \mathcal{C}$ alors h est $(\Sigma, \sigma(\mathcal{C}))$ -mesurable.

Démonstration

- (i) Exercice.
- (ii) On pose $\mathcal{G} := \{A \in \mathcal{P}(F) : h^{-1}(A) \in \Sigma\}$. En utilisant (i), il est facile de vérifier que \mathcal{G} est une tribu. Si \mathcal{G} contient \mathcal{C} , alors elle contient la plus petite tribu contenant \mathcal{C} c'est-à-dire $\sigma(\mathcal{C})$, ce qui conclut la preuve. □

Le point (ii) est très utile pour vérifier la mesurabilité d'applications. Par exemple, si $F = \mathbb{R}$ et $\mathcal{F} = \mathcal{B}(\mathbb{R})$, alors $h : S \rightarrow \mathbb{R}$ est mesurable dès que $\{h \leq c\} \in \Sigma$ pour tout $c \in \mathbb{R}$. (On utilise ici la notation $\{h \leq c\} := \{x \in S : h(x) \leq c\}$). Cela s'applique aussi lorsque $F = \overline{\mathbb{R}}$ et $\mathcal{F} = \mathcal{B}(\overline{\mathbb{R}})$. Par ailleurs, (ii) implique que si $S = F = \mathbb{R}$ alors toute fonction continue est $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ -mesurable.

Proposition 2.5

- (i) Soit $f : S \rightarrow F$ (Σ, \mathcal{F}) -mesurable et $g : F \rightarrow G$ $(\mathcal{F}, \mathcal{G})$ -mesurable. Alors $g \circ f$ est (Σ, \mathcal{G}) -mesurable.
- (ii) Soit $h, h_1, h_2 : S \rightarrow \mathbb{R}$ des applications Σ -mesurables et soit $\lambda \in \mathbb{R}$. Alors λh , $h_1 + h_2$, $h_1 h_2$ sont Σ -mesurables.
- (iii) Soit $(h_n)_{n \in \mathbb{N}}$ une suite d'applications Σ -mesurables. Alors $\inf h_n$ (resp. $\sup h_n$) et $\liminf h_n$ (resp. $\limsup h_n$) sont Σ -mesurables (notons que ces applications prennent leurs valeurs dans $\overline{\mathbb{R}}$ et non \mathbb{R} , il faut donc travailler avec la tribu Borelienne de $\overline{\mathbb{R}}$).

Démonstration

- (i) Pour tout $A \in \mathcal{G}$, on a $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$. Or $B := g^{-1}(A) \in \mathcal{F}$ par mesurabilité de g , donc $f^{-1}(B) \in \Sigma$ par mesurabilité de f et l'on conclut.

(ii) On commence par observer que la fonction $f : x \mapsto \lambda x$ de \mathbb{R} dans \mathbb{R} est mesurable car continue. Ainsi par (i), on en déduit que $\lambda h = f \circ h$ est Σ -mesurable. De même, les fonctions $(x, y) \mapsto x + y$ et $(x, y) \mapsto xy$ de \mathbb{R}^2 dans \mathbb{R} sont mesurables car continues. Par ailleurs l'application (h_1, h_2) de S dans \mathbb{R}^2 est $(\Sigma, \mathcal{B}(\mathbb{R}^2))$ -mesurable (ici on anticipe la Proposition 2.26), donc par (i) on en déduit que $h_1 + h_2$ et $h_1 h_2$ sont Σ -mesurables.

(iii) Pour le premier point, il suffit de remarquer que pour tout $c \in \mathbb{R}$, on a $\{\inf_n h_n \geq c\} = \bigcap_n \{h_n \geq c\}$. Comme h_n est Σ -mesurable alors $\{h_n \geq c\} \in \Sigma$ et $\{\inf_n h_n \geq c\} \in \Sigma$ comme intersection dénombrable d'ensemble mesurables.

Pour le deuxième point, notons que $\liminf h_n = \lim_{k \rightarrow \infty} \inf_{n \geq k} h_n$. Par l'argument ci-dessus, on a que pour tout k , $\inf_{n \geq k} h_n$ est Σ -mesurable. Remarquons également que $\inf_{n \geq k} h_n$ est croissante en k . Ainsi, il suffit de montrer que la limite d'une suite croissante $(g_k)_{k \in \mathbb{N}}$ d'applications mesurables est mesurable. Pour ceci, on remarque que $\lim_k g_k = \sup_k g_k$ et que par l'argument précédent cette application est Σ -mesurable.

□

II Intégration de fonctions positives

L'idée est de définir l'intégrale sur des objets simples puis d'étendre la définition à un cadre général par des arguments d'approximation.

Définition 2.6 (Fonctions étagées)

Soit (S, \mathcal{A}) un espace mesurable et $f : S \rightarrow \mathbb{R}$ une fonction mesurable. On dit que f est étagée si elle ne prend qu'un nombre fini de valeurs. Plus précisément, on dit que f est étagée s'il existe un entier $n \geq 1$, des réels $a_1 < a_2 < \dots < a_n$ et des parties mesurables $A_i \in \mathcal{A}$, $1 \leq i \leq n$ tels que

$$f(x) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}(x), \quad x \in \mathbb{R}.$$

Définition 2.7 (Intégrale d'une fonction étagée positive)

Soit (S, \mathcal{A}, μ) un espace mesuré et soit f une fonction étagée positive. L'intégrale de f par rapport à μ est définie par

$$\int f d\mu = \sum_{i=1}^n a_i \mu(A_i).$$

On suit la convention $0 \cdot \infty = 0$ lorsque $a_i = 0$ et $\mu(A_i) = \infty$.

Exemple 2.8

On considère $([0, 1], \mathcal{B}([0, 1]), \lambda)$ avec λ la mesure de Lebesgue. On définit $f(x) = 1$ si $x \leq \frac{1}{3}$, $f(x) = 99$ si $x \in (\frac{1}{3}, \frac{2}{3}]$ et 0 sinon. Clairement c'est une fonction étagée puisqu'elle ne prend que 3 valeurs et on peut écrire $f(x) = \mathbf{1}_{[0, \frac{1}{3}]} + 99\mathbf{1}_{(\frac{1}{3}, \frac{2}{3}]}$. Ainsi $\int f d\lambda = \lambda([0, \frac{1}{3}]) + 99\lambda((\frac{1}{3}, \frac{2}{3}]) = 33 + \frac{1}{3}$.

L'espace des fonctions étagées positives est stable par addition et multiplication par un réel positif. On vérifie alors aisément les propriétés suivantes.

Proposition 2.9

Soient f et g deux fonctions étagées positives.

(i) Pour tous $\alpha, \beta \geq 0$, on a

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

(ii) Si $f \leq g$ alors $\int f d\mu \leq \int g d\mu$.

Démonstration Exercice. □

Pour étendre notre définition de l'intégrale aux fonctions mesurables positives quelconques, on va s'appuyer sur le résultat de densité suivant.

Proposition 2.10

Soit f une fonction mesurable à valeurs dans $[0, \infty]$. Alors il existe une suite croissante (f_n) de fonctions étagées positives telle que $f = \lim_n f_n$.

Démonstration L'idée est de tronquer f à un seuil élevé puis de faire une approximation dyadique pour le reste des valeurs. Pour ceci on pose pour tout $n \geq 1$

$$A_n = \{x \in S : f(x) \geq n\},$$

et pour tout $i \in \{0, 1, \dots, n2^n - 1\}$ on définit

$$B_{n,i} = \{x \in S : i2^{-n} \leq f(x) < (i+1)2^{-n}\}.$$

On prend alors

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbf{1}_{B_{n,i}} + n \mathbf{1}_{A_n},$$

qui est une fonction étagée. Vérifions que $\lim_n \uparrow f_n(x) = f(x)$ pour tout $x \in S$. Soit $x \in S$. Si $f(x) = \infty$ alors $x \in A_n$ et $f_n(x) = n \uparrow \infty = f(x)$. Sinon, pour tout $n > f(x)$ on a $x \in B_{n,i}$ avec $i = \lfloor 2^n f(x) \rfloor$. On a donc $f_n(x) = \frac{\lfloor 2^n f(x) \rfloor}{2^n}$ qui tend de façon croissante vers $f(x)$ lorsque n tend vers l'infini. □

Définition 2.11 (Intégrale d'une fonction mesurable positive)

Soit (S, \mathcal{A}, μ) un espace mesuré et $f : S \rightarrow [0, \infty]$ une fonction mesurable. On définit l'intégrale de f par rapport à μ par

$$\int f d\mu = \sup_{\{h \text{ étagée positive, } h \leq f\}} \int h d\mu.$$

Remarque 2.12

1. Cette définition coïncide avec celle qu'on a donnée pour les fonctions étagées puisque si h est étagée et $h \leq f$ alors $\int h d\mu \leq \int f d\mu$.
2. On peut vérifier que si f et g sont deux fonctions mesurables positives et $f \leq g$ alors $\int f d\mu \leq \int g d\mu$.

3. On a donné une définition de l'intégrale des fonctions mesurables à valeurs dans $[0, \infty]$ et pas seulement dans $[0, \infty)$. Cela permet d'intégrer tout sup de fonctions mesurables positives (voir Proposition 2.5).

Théorème 2.13 (Théorème de convergence monotone)

Soit $(f_n)_{n \in \mathbb{N}}$ une suite croissante de fonctions mesurables à valeurs dans $[0, \infty]$. Alors

$$\int (\lim_n f_n) d\mu = \lim_n \int f_n d\mu.$$

Démonstration On note $f = \lim_n f_n$. Puisque f_n est une suite croissante, on a $f_n \leq f_{n+1} \leq f$ pour tout $n \in \mathbb{N}$ de sorte que la remarque précédente assure l'inégalité $\int f_n d\mu \leq \int f_{n+1} d\mu \leq \int f d\mu$ pour tout $n \in \mathbb{N}$. La suite des intégrales est donc croissante, ainsi sa limite existe, et l'on obtient

$$\lim_n \int f_n d\mu \leq \int f d\mu.$$

Il nous reste à montrer l'inégalité dans le sens inverse. Soit $h = \sum_{i=1}^m a_i \mathbf{1}_{A_i}$ une fonction étagée positive avec $h \leq f$. Soit $\alpha < 1$ et

$$B_n = \{x \in S : \alpha h(x) \leq f_n(x)\}.$$

Remarquons que $f_n \geq \alpha \mathbf{1}_{B_n} h$ et donc

$$\int f_n d\mu \geq \alpha \int \mathbf{1}_{B_n} h d\mu = \alpha \sum_{i=1}^m a_i \mu(A_i \cap B_n).$$

Par ailleurs la suite B_n est croissante et $S = \cup_n B_n$. En effet, pour tout $x \in S$, soit $f(x) = 0$ auquel cas $h(x) = f_n(x) = 0$ et $x \in B_n$ pour tout $n \geq 1$; soit $f(x) > 0$ et dans ce cas la suite croissante $f_n(x)$ passe au-dessus de $\alpha h(x) < f(x)$ à partir d'un certain rang. Ainsi pour tout $i \leq m$, on a $(A_i \cap B_n) \uparrow A_i$ et par le Lemme 1.13 on trouve $\mu(A_i \cap B_n) \uparrow \mu(A_i)$. Ainsi on obtient

$$\lim_n \int f_n d\mu \geq \alpha \sum_{i=1}^m a_i \mu(A_i) = \alpha \int h d\mu.$$

Ceci étant vrai pour n'importe quel $\alpha < 1$, on fait tendre α vers 1 pour trouver

$$\lim_n \int f_n d\mu \geq \int h d\mu.$$

Ceci étant vrai pour n'importe quelle fonction étagée $h \leq f$, on prend le sup dans le membre de droite pour obtenir

$$\lim_n \int f_n d\mu \geq \sup_{\{h \text{ étagée positive, } h \leq f\}} \int h d\mu = \int f d\mu.$$

□

Remarque 2.14

| On a déjà vu un cas particulier de ce résultat dans le chapitre précédent au Lemme 1.13. En

effet, si A_n est une suite croissante d'événements, alors $(\mathbf{1}_{A_n})$ est une suite croissante de fonctions mesurables positives.

Ce résultat prend toute sa valeur lorsqu'il est combiné avec la proposition qui le précède : étant donné une fonction mesurable positive, la proposition nous assure qu'on peut l'écrire comme limite croissante de fonctions étagées positives, et le théorème précédent nous garantit la commutation entre limite et intégrale de sorte que l'intégrale de notre fonction est la limite des intégrales des fonctions étagées positives.

On peut ainsi déduire les propriétés suivantes de leurs analogues pour les fonctions étagées.

Proposition 2.15

(i) Soient f et g deux fonctions mesurables à valeurs dans $[0, \infty]$ et soient $\alpha, \beta \geq 0$. On a

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

(ii) Si $(f_n)_{n \in \mathbb{N}}$ est une suite de fonctions mesurables à valeurs dans $[0, \infty]$ alors

$$\int \sum_n f_n d\mu = \sum_n \int f_n d\mu.$$

(iii) Si f est une fonction mesurable à valeurs dans $[0, \infty]$ alors $\int f d\mu = 0$ si et seulement si $f = 0$ μ -presque partout.

(iv) Soient f, g deux fonctions mesurables à valeurs dans $[0, \infty]$. Si $f = g$ μ -presque partout alors $\int f d\mu = \int g d\mu$.

Démonstration

(i) Soient $(f_n)_{n \in \mathbb{N}}$ et $(g_n)_{n \in \mathbb{N}}$ deux suites croissantes de fonctions étagées positives dont les limites respectives sont f et g . On a évidemment que $\alpha f_n + \beta g_n$ est une suite croissante de fonctions étagées positives dont la limite est $\alpha f + \beta g$. Par le théorème de convergence monotone, on obtient que

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

Cependant, on a déjà vu que l'intégrale sur les fonctions étagées est linéaire, ainsi

$$\int (\alpha f_n + \beta g_n) d\mu = \alpha \int f_n d\mu + \beta \int g_n d\mu.$$

En passant à la limite et en regroupant les équations précédentes, on obtient le résultat.

(ii) On remarque d'abord que (i) implique que

$$\int \sum_{k=1}^n f_k d\mu = \sum_{k=1}^n \int f_k d\mu.$$

On a donc

$$\sum_{n \in \mathbb{N}} \int f_n d\mu = \lim_{n \rightarrow \infty} \sum_{k=1}^n \int f_k d\mu = \lim_{n \rightarrow \infty} \int \sum_{k=1}^n f_k d\mu.$$

Or $\sum_{k=1}^n f_k$ est croissante car les f_k sont positives, donc par le théorème de convergence monotone, on a

$$\lim_{n \rightarrow \infty} \int \sum_{k=1}^n f_k d\mu = \int \lim_{n \rightarrow \infty} \sum_{k=1}^n f_k d\mu = \int \sum_{n \in \mathbb{N}} f_n d\mu$$

- (iii) Soit f_n une suite croissante de fonctions étagées positives qui converge vers f . Par le Théorème de Convergence Monotone, on a $\int f d\mu = \lim_n \int f_n d\mu$. Si $f = 0$ μ -presque partout, alors, comme $0 \leq f_n \leq f$, on a $f_n = 0$ μ -presque partout et l'on déduit alors facilement que $\int f_n d\mu = 0$ ce qui suffit à conclure. Si $\int f d\mu = 0$ alors $\int f_n d\mu = 0$ pour tout n , et ainsi la définition de l'intégrale des fonctions étagées assure que $f_n = 0$ μ -presque partout.
- (iv) On remarque que $u := \max(f - g, 0)$ est mesurable positive. Si $f = g$ μ -presque partout alors $u = 0$ μ -presque partout et l'on peut appliquer (iii) pour obtenir $\int u d\mu = 0$. De même, on obtient $\int \max(g - f, 0) d\mu = 0$. Or $f + \max(g - f, 0) = g + \max(f - g, 0)$ de sorte que

$$\int (f + \max(g - f, 0)) d\mu = \int (g + \max(f - g, 0)) d\mu$$

ce qui implique le résultat voulu. □

Corollaire 2.16

Soit f une fonction mesurable positive. Pour tout $A \in \mathcal{A}$, on définit $\nu(A) = \int f \mathbf{1}_A d\mu$ (qu'on note aussi $\int_A f d\mu$). Alors ν est une mesure.

Démonstration Exercice. □

III Intégrale dans le cas général

Toute fonction f est décomposable en partie positive et partie négative en posant $f^+ = \sup(0, f)$ et $f^- = \sup(0, -f)$ de telle sorte que $f = f^+ - f^-$. Grâce à ceci, on peut étendre la définition de l'intégrale aux fonctions mesurables de signe quelconque.

Définition 2.17 (Fonctions intégrables)

Soit (S, \mathcal{A}, μ) un espace mesuré et $f : S \rightarrow \mathbb{R}$ une fonction mesurable. On dit que f est μ -intégrable si

$$\int |f| d\mu < \infty.$$

Dans ce cas, on pose

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

On note $L^1(S, \mathcal{A}, \mu)$ l'ensemble des fonctions μ -intégrables.

On remarquera que toute fonction positive n'est pas forcément intégrable mais que son intégrale est toujours bien définie.

Remarque 2.18

L'espace L^1 ainsi défini est un espace vectoriel, et l'application $f \mapsto \int |f| d\mu$ a toutes les propriétés

d'une norme sauf la suivante : $\int |f| d\mu = 0$ n'implique pas que $f = 0$ mais seulement que $f = 0$ μ -presque partout. Il faut en fait quotienter l'espace considéré ici par la relation d'équivalence $f \sim g \Leftrightarrow f = g$ μ -presque partout, pour obtenir un espace vectoriel normé. On peut alors prouver qu'il s'agit d'un espace de Banach.

On a les propriétés suivantes.

Proposition 2.19

- (i) Soit $f \in L^1(S, \mathcal{A}, \mu)$. On a $|\int f d\mu| \leq \int |f| d\mu$.
- (ii) Si $f, g \in L^1(S, \mathcal{A}, \mu)$ et $\alpha, \beta \in \mathbb{R}$ alors $\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$.
- (iii) Si $f, g \in L^1(S, \mathcal{A}, \mu)$ et $f \leq g$ alors $\int f d\mu \leq \int g d\mu$.
- (iv) Si $f, g \in L^1(S, \mathcal{A}, \mu)$ et $f = g$ μ -presque partout alors $\int f d\mu = \int g d\mu$.

Démonstration Exercice. □

On a ainsi étendu les propriétés qu'on avait établies pour l'intégrale dans le cas des fonctions étagées au cas des fonctions mesurables de signe quelconque. Cependant, la propriété de commutation entre limite et intégrale établie dans le théorème de convergence monotone est fortement liée au caractère positif des fonctions considérées. Dans le cas général, on pourra permuter limite et intégrale sous certaines hypothèses, ce sera le contenu du théorème de convergence dominée. Nous avons besoin d'abord du lemme suivant.

Lemme 2.20 (Lemme de Fatou)

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables positives. Alors on a

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu.$$

Démonstration On écrit $\liminf f_n = \lim_{m \rightarrow \infty} \inf_{n \geq m} f_n = \lim_{m \rightarrow \infty} g_m$. Or $(g_m)_{m \in \mathbb{N}}$ est une suite croissante de fonctions mesurables positives, donc par le théorème de convergence monotone, on a

$$\int \lim_{m \rightarrow \infty} g_m d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu.$$

Or $g_m \leq f_n$ pour tout $n \geq m$, donc $\int g_m d\mu \leq \int f_n d\mu$ pour tout $n \geq m$. On déduit que $\int g_m d\mu \leq \inf_{n \geq m} \int f_n d\mu$. En regroupant ce qu'on a obtenu, on a

$$\int \lim_{m \rightarrow \infty} g_m d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu \leq \lim_{m \rightarrow \infty} \inf_{n \geq m} \int f_n d\mu.$$

□

Remarque 2.21

On a déjà montré un cas particulier de ce lemme dans le chapitre 1. En effet si A_n sont des événements et $\mu = \mathbb{P}$ est une probabilité, on pose $f_n = \mathbf{1}_{A_n}$ et on remarque $\int f_n d\mu = \int \mathbf{1}_{A_n} d\mathbb{P} = \mathbb{P}(A_n)$. D'autre part, on a vu que $\liminf \mathbf{1}_{A_n} = \mathbf{1}_{\liminf A_n}$ et alors

$$\int \liminf f_n d\mu = \int \mathbf{1}_{\liminf A_n} d\mathbb{P} = \mathbb{P}(\liminf A_n).$$

Ce lemme nous donne donc que $\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n)$.

Théorème 2.22 (Théorème de convergence dominée)

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions dans $L^1(S, \mathcal{A}, \mu)$ telle que $\lim_n f_n(x) = f(x)$ μ -presque partout. Supposons qu'il existe une fonction g mesurable, positive et intégrable telle que pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ μ -presque partout. Alors on a $f \in L^1(S, \mathcal{A}, \mu)$ et

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu, \quad \text{et} \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

Démonstration Soit $A = \{x \in S : \lim_n f_n(x) = f(x) \text{ et } |f_n(x)| \leq g(x) \forall n \in \mathbb{N}\}$. On a par hypothèse que $\mu(A^c) = 0$. Posons

$$\tilde{f}_n(x) = f_n(x)\mathbf{1}_A(x) \quad \text{et} \quad \tilde{f}(x) = f(x)\mathbf{1}_A(x).$$

On a clairement que $\tilde{f}_n = f_n$ et $\tilde{f} = f$ μ -presque partout. On a donc que $\int \tilde{f}_n d\mu = \int f_n d\mu$, $\int \tilde{f} d\mu = \int f d\mu$ et $\int |\tilde{f}_n - \tilde{f}| d\mu = \int |f_n - f| d\mu$. On peut ainsi travailler avec \tilde{f}_n et \tilde{f} à la place de f_n et f .

Puisque $|\tilde{f}| \leq g$ et $\int g d\mu < \infty$ alors $\int |f| d\mu = \int |\tilde{f}| d\mu < \infty$ et $f \in L^1(S, \mathcal{A}, \mu)$.

On a $2g - |\tilde{f} - \tilde{f}_n| \geq 0$ donc par le lemme de Fatou, on a

$$\liminf \int (2g - |\tilde{f} - \tilde{f}_n|) d\mu \geq \int \liminf (2g - |\tilde{f} - \tilde{f}_n|) d\mu = 2 \int g d\mu.$$

Or

$$\liminf \int (2g - |\tilde{f} - \tilde{f}_n|) d\mu = \int 2g d\mu - \limsup \int |\tilde{f} - \tilde{f}_n| d\mu \leq 2 \int g d\mu.$$

On déduit donc que $\limsup \int |\tilde{f} - \tilde{f}_n| d\mu = 0$ donc en particulier que $\lim \int |f - f_n| d\mu = 0$. Finalement, par linéarité de l'intégrale, on a

$$\left| \int f d\mu - \int f_n d\mu \right| = \left| \int (f - f_n) d\mu \right| \leq \int |f - f_n| d\mu \xrightarrow{n \rightarrow \infty} 0.$$

□

On termine cette section en énonçant un résultat très utile pour les intégrales à paramètres.

Théorème 2.23

Soit (S, Σ, μ) un espace mesuré et $I \subset \mathbb{R}$ un intervalle ouvert. Soit $f : I \times S \rightarrow \mathbb{R}$ une application telle que :

- (i) pour tout $u \in I$, l'application $x \mapsto f(u, x)$ est dans $L^1(S, \Sigma, \mu)$,
- (ii) pour μ -presque tout $x \in S$, l'application $u \mapsto f(u, x)$ est dérivable sur I . On notera sa dérivée $\partial_u f(u, x)$.
- (iii) il existe une fonction mesurable positive $g \in L^1(S, \Sigma, \mu)$ telle que pour tout $u \in I$ et pour μ -presque tout $x \in S$ on a l'hypothèse de domination

$$|\partial_u f(u, x)| \leq g(x).$$

Alors l'application $F : u \mapsto \int_S f(u, x)\mu(dx)$ est dérivable et l'on a

$$F'(u) = \int_S \partial_u f(u, x)\mu(dx) .$$

La preuve repose sur le théorème de convergence dominée, nous renvoyons au cours de L3 ou au livre [BP00].

IV Tribu et mesure produit

On se donne deux espaces mesurables (S_1, \mathcal{A}_1) et (S_2, \mathcal{A}_2) , et on aimerait munir l'espace produit $S_1 \times S_2$ d'une tribu "naturelle".

Définition 2.24 (σ -algèbre produit)

Soient (S_1, \mathcal{A}_1) et (S_2, \mathcal{A}_2) deux espaces mesurables. La σ -algèbre produit sur $S_1 \times S_2$ est définie comme la plus petite σ -algèbre contenant les produits d'éléments de \mathcal{A}_1 et \mathcal{A}_2 :

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

Remarque 2.25

1. Notons ρ_1 et ρ_2 les projections canoniques i.e. $\rho_1 : S_1 \times S_2 \rightarrow S_1$ est définie par $\rho_1(s_1, s_2) = s_1$, et $\rho_2 : S_1 \times S_2 \rightarrow S_2$ est définie par $\rho_2(s_1, s_2) = s_2$. Alors on peut vérifier que $\mathcal{A}_1 \otimes \mathcal{A}_2$ est la plus petite σ -algèbre qui rend ρ_1 et ρ_2 mesurables. Plus précisément, on a que $\mathcal{A}_1 \otimes \mathcal{A}_2$ est la σ -algèbre engendrée par les ensembles de la forme $\rho_1^{-1}(A_1)$, $A_1 \in \mathcal{A}_1$, et $\rho_2^{-1}(A_2)$, $A_2 \in \mathcal{A}_2$.
2. On peut évidemment étendre cette définition à un produit fini d'espaces mesurables.

On peut se demander si la tribu Borélienne sur \mathbb{R}^2 coïncide avec le produit des tribus Boréliennes sur \mathbb{R} .

Proposition 2.26

On a $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Plus généralement $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$ pour tout entier $n \geq 1$.

Démonstration Les applications ρ_1 et ρ_2 de la remarque précédente sont continues de \mathbb{R}^2 dans \mathbb{R} . Ainsi, elles sont mesurables pour les tribus Boréliennes associées. Comme la tribu produit est la plus petite tribu rendant mesurable ρ_1 et ρ_2 , on en déduit que la tribu $\mathcal{B}(\mathbb{R}^2)$ contient la tribu produit.

Réciproquement tout ouvert de \mathbb{R}^2 peut s'écrire comme union dénombrable de produits d'ouverts. En effet, on peut considérer les ensembles $(a, b) \times (c, d)$ pour $a, b, c, d \in \mathbb{Q}$ qui sont en quantité dénombrable. Soit $O \subset \mathbb{R}^2$ ouvert. Pour tout $x \in O$, par densité des rationnels, il existe $(a, b) \times (c, d)$ tel que $x \in (a, b) \times (c, d) \subset O$. On peut alors écrire O comme réunion dénombrable de tels ensembles. Or tous ces ensembles appartiennent à la tribu produit par définition. On en déduit donc que la tribu produit contient tous les ouverts de \mathbb{R}^2 , et donc la plus petite tribu engendrée par ceux-ci. \square

La proposition suivante montre que les "fibres" sont mesurables.

Proposition 2.27

Soit $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Alors pour tout $s_1 \in S_1$, $A_{s_1} := \{s_2 \in S_2 : (s_1, s_2) \in A\}$ est \mathcal{A}_2 -mesurable. De même, pour tout $s_2 \in S_2$, $A^{s_2} := \{s_1 \in S_1 : (s_1, s_2) \in A\}$ est \mathcal{A}_1 -mesurable.

Démonstration Soit $s_1 \in S_1$. Alors l'ensemble $\{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : A_{s_1} \in \mathcal{A}_2\}$ est une tribu, qui contient les ensembles produits $A_1 \times A_2$ pour tous $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. Elle contient donc la tribu produit. Le raisonnement est identique pour A^{s_2} . \square

Notre but désormais est de définir une mesure sur $\mathcal{A}_1 \otimes \mathcal{A}_2$ étant données deux mesures sur \mathcal{A}_1 et \mathcal{A}_2 .

Proposition 2.28 (Mesure produit)

Soient $(S_1, \mathcal{A}_1, \mu_1)$ et $(S_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés avec μ_1 et μ_2 σ -finies. Il existe une unique mesure μ sur $(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ satisfaisant

$$\mu(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2) ,$$

pour tout $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$. Cette mesure sera notée $\mu = \mu_1 \otimes \mu_2$. Elle vérifie

$$\mu(A) = \int_{S_1} \left(\int_{S_2} \mathbf{1}_A(s_1, s_2) d\mu_2 \right) d\mu_1 = \int_{S_2} \left(\int_{S_1} \mathbf{1}_A(s_1, s_2) d\mu_1 \right) d\mu_2,$$

pour tout $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

Démonstration L'ensemble des produits $A_1 \times A_2$ est stable par intersection finie. On déduit donc du Corollaire 1.17 l'unicité de la mesure μ .

Passons à l'existence. Supposons que μ_1 et μ_2 sont finies, et définissons la mesure μ par

$$\mu(A) := \int_{S_1} \left(\int_{S_2} \mathbf{1}_A(s_1, s_2) d\mu_2 \right) d\mu_1 , \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2 .$$

Examinons cette définition. On observe que

$$\int_{S_2} \mathbf{1}_A(s_1, s_2) d\mu_2 = \int_{S_2} \mathbf{1}_{A_{s_1}}(s_2) d\mu_2 = \mu_2(A_{s_1}) ,$$

car $A_{s_1} \in \mathcal{A}_2$ par la proposition précédente. Pour que la définition ait bien un sens, il faut que $s_1 \mapsto \mu_2(A_{s_1})$ soit mesurable. Pour vérifier cela, on introduit

$$\mathcal{G} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : s_1 \mapsto \mu_2(A_{s_1}) \text{ est } \mathcal{A}_1\text{-mesurable}\} .$$

Montrons que \mathcal{G} est une classe monotone. Il est facile de voir qu'elle contient $S_1 \times S_2$. Par ailleurs, si $A \subset A'$, on a $(A' \setminus A)_{s_1} = (A'_{s_1}) \setminus (A_{s_1})$ et donc $\mu_2((A' \setminus A)_{s_1}) = \mu_2((A'_{s_1})) - \mu_2(A_{s_1})$ (on rappelle que μ_2 est finie). Ainsi \mathcal{G} est stable par différence. Enfin pour toute suite A_n croissante, on voit que $(\cup_n A_n)_{s_1} = \cup_n (A_n)_{s_1}$ qui est croissante, de sorte que le Lemme 1.13 donne $\mu_2((\cup_n A_n)_{s_1}) = \mu_2(\cup_n (A_n)_{s_1}) = \lim_n \mu_2((A_n)_{s_1})$. Comme la limite d'une suite de fonctions mesurables est mesurable, on en déduit que \mathcal{G} est stable par limite croissante. On notera que ça n'est pas une tribu en général car la mesure d'une union n'est pas forcément la somme des mesures.

Par ailleurs pour tout produit $A_1 \times A_2$ d'éléments de \mathcal{A}_1 et \mathcal{A}_2 , on a $\mu_2((A_1 \times A_2)_{s_1}) = \mu_2(A_2) \mathbf{1}_{A_1}(s_1)$ qui est bien \mathcal{A}_1 -mesurable. Ainsi \mathcal{G} est une classe monotone qui contient les ensembles produits. Par le Lemme de classe monotone, on en déduit qu'elle contient la tribu produit. Ceci assure que

$s_1 \mapsto \mu_2(A_{s_1})$ est bien mesurable et montre que la définition de μ a bien un sens. Les arguments précédents montrent que

$$\mu(A_1 \times A_2) = \int_{S_1} \mu_2(A_2) \mathbf{1}_{A_1}(s_1) d\mu_1 = \mu_1(A_1) \mu_2(A_2).$$

Il nous reste à vérifier que μ est une mesure. Clairement $\mu(\emptyset) = 0$. Par ailleurs pour toute suite de parties disjointes $A_n \in \mathcal{A}_1 \otimes \mathcal{A}_2$, les parties $(A_n)_{s_1}$ sont également disjointes de sorte que

$$\mu_2((\cup_n A_n)_{s_1}) = \mu_2((\cup_n (A_n)_{s_1})) = \sum_n \mu_2((A_n)_{s_1}).$$

Cette fonction est \mathcal{A}_1 -mesurable comme limite de fonctions \mathcal{A}_1 -mesurables. Par le Théorème de Convergence Monotone, on voit alors que

$$\mu(\cup_n A_n) = \int_{S_1} \sum_n \mu_2((A_n)_{s_1}) d\mu_1 = \sum_n \int_{S_1} \mu_2((A_n)_{s_1}) d\mu_1 = \sum_n \mu(A_n).$$

Les mêmes arguments assurent que

$$\mu'(A) := \int_{S_2} \left(\int_{S_1} \mathbf{1}_A(s_1, s_2) d\mu_1 \right) d\mu_2, \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

est également une mesure qui vérifie $\mu(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$. Par l'unicité déjà établie, on déduit que $\mu = \mu'$.

L'extension au cas σ -fini est laissée en exercice. □

Remarque 2.29

1. Pour identifier l'intégrant, on notera parfois $\int_{S_2} \mathbf{1}_A(s_1, s_2) d\mu_2 = \int_{S_2} \mathbf{1}_A(s_1, s_2) \mu_2(ds_2)$.
2. Notons que ces hypothèses de finitude sont essentielles. En effet, si $(S_1, \mathcal{A}_1) = (S_2, \mathcal{A}_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\mu_1 = \lambda$ et $\mu_2 = \nu$ la mesure de comptage et si $A = \{(x, x) : x \in \mathbb{R}\}$, alors clairement

$$\int_{\mathbb{R}} \mathbf{1}_A(s_1, s_2) d\lambda(s_1) = \lambda(\{s_2\}) = 0$$

et

$$\int_{\mathbb{R}} \mathbf{1}_A(s_1, s_2) d\nu(s_2) = \nu(\{s_1\}) = 1.$$

Ainsi

$$0 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(s_1, s_2) d\lambda \right) d\nu \neq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_A(s_1, s_2) d\nu \right) d\lambda = \infty.$$

On passe à présent à l'intégration contre la mesure produit.

Théorème 2.30 (Fubini)

Soient $(S_1, \mathcal{A}_1, \mu_1)$ et $(S_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés et μ_1, μ_2 σ -finies.

- (i) Soit $f : (S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow \overline{\mathbb{R}}$ mesurable. Alors pour tout $s_1 \in S_1$, l'application $f_{s_1} : s_2 \rightarrow f(s_1, s_2)$ est \mathcal{A}_2 -mesurable. De même, pour tout $s_2 \in S_2$, l'application $f^{s_2} : s_1 \rightarrow f(s_1, s_2)$ est

\mathcal{A}_1 -mesurable.

(ii) Soit $f : (S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow [0, \infty]$ mesurable positive. Alors

$$\int f d\mu_1 \otimes \mu_2 = \int_{S_1} \left(\int_{S_2} f(s_1, s_2) d\mu_2 \right) d\mu_1 = \int_{S_2} \left(\int_{S_1} f(s_1, s_2) d\mu_1 \right) d\mu_2.$$

(iii) Si $f \in L^1(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, alors

$$\int f d\mu_1 \otimes \mu_2 = \int_{S_1} \left(\int_{S_2} f(s_1, s_2) d\mu_2 \right) d\mu_1 = \int_{S_2} \left(\int_{S_1} f(s_1, s_2) d\mu_1 \right) d\mu_2.$$

Démonstration

(i) On remarque que pour tout $B \in \mathcal{B}(\overline{\mathbb{R}})$,

$$(f_{s_1})^{-1}(B) = \{s_2 \in S_2 : f(s_1, s_2) \in B\} = \{s_2 \in S_2 : (s_1, s_2) \in f^{-1}(B)\} = (f^{-1}(B))_{s_1}$$

qui est \mathcal{A}_2 -mesurable par un résultat précédent.

(ii) On se concentre sur la première égalité. La proposition qui définit la mesure produit assure que ceci est vrai pour toute fonction indicatrice (en particulier, la preuve assure que $s_1 \mapsto \int_{S_2} f(s_1, s_2) d\mu_2$ est bien mesurable). Par linéarité de l'intégrale, ceci reste vrai pour toute fonction étagée positive. Soit f une fonction mesurable positive et soit f_n une suite croissante de fonctions étagées positives convergeant vers f . Pour tout $s_1 \in S_1$, on peut appliquer le Théorème de Convergence Monotone et obtenir que

$$\lim_n \int_{S_2} f_n(s_1, s_2) d\mu_2(s_2) = \int_{S_2} f(s_1, s_2) d\mu_2(s_2),$$

ainsi $s_1 \mapsto \int_{S_2} f(s_1, s_2) d\mu_2(s_2)$ est \mathcal{A}_1 -mesurable comme limite de fonctions mesurables. Par ailleurs, la suite de fonctions $s_1 \mapsto \int_{S_2} f_n(s_1, s_2) d\mu_2(s_2)$ étant croissante et positive, une nouvelle application du TCM assure que

$$\begin{aligned} \lim_n \int_{S_1} \left(\int_{S_2} f_n(s_1, s_2) d\mu_2 \right) d\mu_1 &= \int_{S_1} \lim_n \left(\int_{S_2} f_n(s_1, s_2) d\mu_2 \right) d\mu_1 \\ &= \int_{S_1} \left(\int_{S_2} \lim_n f_n(s_1, s_2) d\mu_2 \right) d\mu_1 \\ &= \int_{S_1} \left(\int_{S_2} f(s_1, s_2) d\mu_2 \right) d\mu_1. \end{aligned}$$

Par ailleurs, le TCM appliqué l'intégrale contre la mesure produit donne

$$\int f d\mu_1 \otimes \mu_2 = \lim_n \int f_n d\mu_1 \otimes \mu_2.$$

Il suffit alors d'appliquer le résultat dans le cas des fonctions étagées.

(iii) Il suffit de décomposer f en partie positive et partie négative qui sont toutes deux intégrables et vérifient l'égalité en question par le (ii). La linéarité termine alors la preuve. □

Remarque 2.31

1. Le fait que $f \in L^1(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$ est essentiel. En effet si $f(x, y) = 2e^{-2xy} - e^{-xy}$ définie sur $(0, \infty) \times (0, 1]$. Alors

$$\int_{(0, \infty)} f(x, y) dx = 0 \quad \text{et} \quad \int_{(0, 1]} f(x, y) dy = \frac{e^{-x} - e^{-2x}}{x}.$$

Si l'on applique le théorème sans vérifier l'hypothèse d'intégrabilité on aurait

$$0 = \int_{(0, 1]} \left(\int_{(0, \infty)} f(x, y) dx \right) dy \neq \int_{(0, \infty)} \left(\int_{(0, 1]} f(x, y) dy \right) dx > 0.$$

En fait la fonction f n'est pas intégrable.

2. Un cas particulier qu'on va utiliser dans la suite. Prenons μ_1 la mesure de Lebesgue (qui est σ -finie) et μ_2 une mesure de probabilité. Si $f(x_1, x_2)$ est mesurable positive et $x_1 \rightarrow f(x_1, x_2)$ est intégrable par rapport à μ_1 et $x_2 \rightarrow \int f(x_1, x_2) \mu_1(dx_1)$ est bornée, alors $f \in L^1$.

En effet, puisque f est positive, on peut appliquer Fubini pour avoir

$$\int f d(\mu_1 \otimes \mu_2) = \int \left(\int f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2)$$

Comme $x_2 \rightarrow \int f(x_1, x_2) \mu_1(dx_1)$ est bornée et μ_2 est une probabilité, alors la quantité ci-dessus est finie et $f \in L^1$.

Ainsi, si on dispose d'une fonction quelconque g telle que $|g| \leq f$ et f vérifie les conditions ci-dessus, on pourra affirmer que $g \in L^1$ et utiliser Fubini pour intervertir les intégrales sur g .

Chapitre 3

Variable aléatoire

I Définition d'une variable aléatoire

En probabilités, il est fréquent d'appeler variable aléatoire toute application mesurable.

Définition 3.1 (*Variable aléatoire*)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et (F, \mathcal{F}) un espace mesurable. Une application $(\mathcal{A}, \mathcal{F})$ -mesurable $X : \Omega \rightarrow F$ est appelée variable aléatoire. Lorsque $F = \mathbb{R}$ et $\mathcal{F} = \mathcal{B}(\mathbb{R})$ on parle de variable aléatoire réelle.

Remarque 3.2

1. Il ne faut pas oublier qu'une variable aléatoire est avant tout une application.
2. La plupart des variables aléatoires que l'on manipulera seront réelles. Il sera fréquent d'omettre ce terme dans la suite.

Exemple 3.3

1. Durant une année, le PSG et Lyon se rencontrent 6 fois. Il y a 3^6 résultats possibles (à chaque match, soit Lyon gagne, soit le PSG, soit match nul). On s'intéresse au nombre de matchs gagnés par Lyon et on définit donc X comme étant la variable aléatoire égale au nombre de matchs gagnés par Lyon. Soit $X : \Omega \rightarrow \{0, \dots, 6\}$, on peut choisir Ω comme étant l'ensemble de tous les résultats possibles de l'expérience, c'est-à-dire, l'ensemble des 6-uplets formés des mots "Lyon", "PSG" ou "nul". Par exemple $\{\text{Lyon, Lyon, Lyon, Lyon, Lyon, nul}\}$ est un élément de Ω , il correspond à la situation dans laquelle Lyon gagne les cinq premiers matchs, et le sixième match est un nul. La tribu associée est alors l'ensemble des parties de Ω .
2. On jette un dé 2 fois et on s'intéresse à la somme des faces obtenues. On note Ω l'ensemble de toutes les issues possibles de l'expérience de jeter 2 dés i.e. $\Omega = \{(i, j) : i, j \in \{1, \dots, 6\}\}$. Ainsi $X : \Omega \rightarrow \{2, \dots, 12\}$ est définie par $X(i, j) = i + j$. Si l'on munit Ω et $\{2, \dots, 12\}$ des σ -algèbres $\mathcal{P}(\Omega)$ et $\mathcal{P}(\{2, \dots, 12\})$, alors X est une variable aléatoire. En revanche, si l'on munit Ω de la σ -algèbre $\mathcal{A} := \{\emptyset, \Omega, \{(1, 1)\}, \Omega \setminus \{(1, 1)\}\}$ alors X n'est pas une variable aléatoire puisque $X^{-1}(\{3\}) = \{(1, 2), (2, 1)\} \notin \mathcal{A}$.
3. On s'intéresse au nombre de "pile"s obtenues lorsqu'on jette une pièce de monnaie n fois. Soit alors $\Omega = \{P, F\}^n$ et soit $\mathcal{A} = \mathcal{P}(\Omega)$. Pour tout $i \in \{1, \dots, n\}$, on définit $X_i : \Omega \rightarrow \{0, 1\}$ par $X_i(\omega) = 1$ si $\omega_i = P$ et 0 sinon, pour tout $\omega = (\omega_1, \dots, \omega_n) \in \Omega$. Ainsi X_i nous dit si le i -ième

jet de pièce a produit pile ou non. Clairement X_i est une variable aléatoire réelle. Pour compter le nombre de piles dans n expériences, on définit $X = X_1 + X_2 + \dots + X_n$. Comme chaque X_i est une variable aléatoire réelle et que X en est la somme alors X est bien une variable aléatoire réelle.

La classe des variables aléatoires qui prennent un nombre fini de valeurs nous sera particulièrement utile.

Définition 3.4 (Variable aléatoire étagée)

Une variable aléatoire $X : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ est dite étagée s'il existe des nombres $a_1, \dots, a_n \in \mathbb{R}$ et des événements $A_1, \dots, A_n \in \mathcal{A}$ tels que $X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$.

Il s'avère que toute variable aléatoire réelle est la limite simple d'une suite de variables étagées. C'est ce qu'on a vu et montré au chapitre précédent.

Lemme 3.5

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

- (i) Si X est positive, alors X est limite simple d'une suite croissante de variables aléatoires étagées i.e. il existe une suite de variables aléatoires étagées $X_1 \leq X_2 \dots$ telles que pour tout $\omega \in \Omega$, on a $X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)$.
- (ii) Si X est quelconque, alors X est limite simple d'une suite de variables aléatoires étagées.

Démonstration

- (i) Déjà fait.
- (ii) Il suffit d'écrire $X = X^+ - X^-$ où X^+ et X^- sont définies par $X^+(\omega) = \max(0, X(\omega))$ et $X^-(\omega) = -\min(0, X(\omega))$. Il suffit alors d'appliquer le (i) à X^+ et X^- pour déduire le résultat. □

Comme nous l'avons remarqué dans l'exemple 3.3, choisir l'ensemble des parties de Ω comme σ -algèbre garantit toujours la mesurabilité de la variable aléatoire. Bien entendu, dans la plupart des cas on peut se restreindre à une σ -algèbre plus petite.

Définition 3.6 (σ -algèbre engendrée par une application)

Soit $X : \Omega \rightarrow F$ une application et \mathcal{F} une tribu sur F . On définit $\sigma(X)$ comme la plus petite σ -algèbre sur Ω pour laquelle X est une variable aléatoire.

Il s'agit de l'intersection de toutes les tribus sur Ω rendant X mesurable. On a mentionné au chapitre précédent qu'une intersection quelconque de tribus est une tribu.

Remarque 3.7

Si $X : (\Omega, \mathcal{A}) \rightarrow (F, \mathcal{F})$ est une variable aléatoire, alors $\sigma(X) \subseteq \mathcal{A}$.

Lemme 3.8

On a $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{F}\}$.

Démonstration Tout d'abord, les propriétés de l'image réciproque impliquent que $\{X^{-1}(A) : A \in \mathcal{F}\}$ est bien une tribu. Par ailleurs, il est clair que X est $\{X^{-1}(A) : A \in \mathcal{F}\}$ -mesurable ce qui

implique que $\sigma(X) \subseteq \{X^{-1}(A) : A \in \mathcal{F}\}$. D'autre part, comme X est $\sigma(X)$ -mesurable alors pour tout $A \in \mathcal{F}$ on a $X^{-1}(A) \in \sigma(X)$ ce qui implique que $\{X^{-1}(A) : A \in \mathcal{F}\} \subseteq \sigma(X)$. \square

Exemple 3.9

1. Soit $c \in \mathbb{R}$ et soit $X : \mathbb{R} \rightarrow \mathbb{R}$ définie par $X(a) = c$ pour tout $a \in \mathbb{R}$. Alors $\sigma(X) = \{\emptyset, \mathbb{R}\}$.
2. Soit Ω un ensemble et $A \subset \Omega$ un sous-ensemble de Ω . Soit $X = \mathbf{1}_A$ l'indicatrice de A . Alors $\sigma(X) = \{\emptyset, A, A^c, \Omega\}$.
3. Dans les trois exemples de l'Exemple 3.3, on avait $\sigma(X) = \mathcal{P}(\Omega)$.

Proposition 3.10

Soient $X, Y : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ deux variables aléatoires réelles. Alors les deux assertions suivantes sont équivalentes :

- (i) Y est $\sigma(X)$ -mesurable.
- (ii) Il existe une fonction $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ -mesurable $f : \mathbb{R} \rightarrow \mathbb{R}$ telle que $Y = f(X)$.

Démonstration On a déjà vu que la composition d'applications mesurables est mesurable, ce qui implique que si $Y = f(X)$ alors Y est $\sigma(X)$ -mesurable. Il reste à montrer que (i) implique (ii). Supposons d'abord que Y est étagée et écrivons $Y = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ où $A_i \in \sigma(X)$. Pour tout $i \leq n$, prenons $B_i \in \mathcal{B}(\mathbb{R})$ tel que $A_i = X^{-1}(B_i)$. Ainsi

$$Y = \sum_{i=1}^n a_i \mathbf{1}_{X^{-1}(B_i)} = \sum_{i=1}^n a_i \mathbf{1}_{B_i} \circ X = f \circ X,$$

avec $f = \sum_{i=1}^n a_i \mathbf{1}_{B_i}$ qui est mesurable.

Dans le cas général, on a que Y est limite simple de variables aléatoires étagées qui sont toutes de la forme $Y_n = f_n(X)$ avec f_n mesurable. Pour tout $x \in \mathbb{R}$, on pose alors $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ si la limite existe et 0 sinon. Par les propriétés vues au chapitre précédent f est mesurable. Pour tout ω , $Y_n(\omega) = f_n(X(\omega))$ converge vers $Y(\omega)$, ce qui implique que $x = X(\omega)$ est un point tel que $f_n(x) \rightarrow f(x)$. Ainsi $f(X(\omega)) = \lim_n f_n(X(\omega)) = Y(\omega)$. \square

II Loi d'une variable aléatoire

On introduit à présent un concept fondamental en probabilité.

Définition 3.11

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, (F, \mathcal{F}) un espace mesurable et $X : \Omega \rightarrow F$ une variable aléatoire. On définit la loi de X comme l'application $\mathcal{L}_X : \mathcal{F} \rightarrow [0, 1]$ définie par $\mathcal{L}_X(A) = \mathbb{P}(X^{-1}(A))$ pour tout $A \in \mathcal{F}$.

Remarque 3.12

1. La loi de X est une mesure de probabilité sur \mathcal{F} . (Exercice)
2. On transporte sur \mathcal{F} , à l'aide de X , la mesure de probabilité \mathbb{P} qu'on a à disposition sur \mathcal{A} . Il s'agit en fait de la mesure image de \mathbb{P} par l'application mesurable X : la notion de mesure image fait sens dans le cas général où \mathbb{P} n'est pas une mesure de probabilité.

Exemple 3.13

On reprend les trois exemples de l'Exemple 3.3 dans l'ordre :

1. On définit une probabilité sur Ω en stipulant d'une part que $\mathbb{P}(\{\omega\}) = 0$ si ω contient le mot "nul". En d'autres termes, la probabilité qu'il y ait au moins un match nul parmi les 6 est nulle. On stipule par ailleurs que les autres issues sont toutes équi-probables : tout 6-uplets consistué de "PSG" et "Lyon" a une probabilité égale à 2^{-6} . Ainsi, \mathcal{L}_X vérifie

$$\mathcal{L}_X(\{k\}) = \mathbb{P}(X^{-1}(\{k\})) = \frac{\binom{6}{k}}{2^6},$$

pour tout $k = 0, \dots, 6$. Notons que les ensembles $\{k\}$, $k = 0, \dots, 6$, forment (en ajoutant l'ensemble vide) un π -système générateur de $\sigma(\{0, \dots, 6\})$: par le Théorème 1.19, cela caractérise entièrement la mesure \mathcal{L} .

2. On suppose que toutes les issues ont la même probabilité. Ainsi la loi de X est donnée par

$$\mathcal{L}_X(\{2\}) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}, \quad \mathcal{L}_X(\{3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{1}{18}, \quad \mathcal{L}_X(\{4\}) = \frac{1}{12},$$

et ainsi de suite. Comme précédemment, cela caractérise \mathcal{L}_X .

3. On suppose également que toutes les issues sont équiprobables, ainsi $\mathbb{P}(\{\omega\}) = 2^{-n}$ pour tout $\omega \in \{P, F\}^n$. Pour obtenir $X = k$ pour un $k \in \{0, \dots, n\}$, il faudrait avoir un élément ω contenant k fois P . On note Ω_k l'ensemble de tous les ω contenant k fois P . Clairement, on a $|\Omega_k| = \binom{n}{k}$. Ainsi,

$$\mathcal{L}_X(\{k\}) = \mathbb{P}(X^{-1}(\{k\})) = \mathbb{P}(\Omega_k) = \sum_{\omega \in \Omega_k} \mathbb{P}(\{\omega\}) = 2^{-n} |\Omega_k| = 2^{-n} \binom{n}{k}.$$

Définition 3.14 (Variable aléatoire discrète)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle sur cet espace. Si l'ensemble des valeurs prises par X est dénombrable alors on dit que X est une variable aléatoire discrète. La loi de X est alors donnée par $\mathcal{L}_X = \sum_{a \in \text{Im } X} \mathbb{P}(X = a) \delta_a$ où δ_a est la mesure de Dirac en a , qui est définie par $\delta_a(A) = \mathbf{1}_A(a)$ pour tout $A \in \mathcal{P}(\Omega)$.

On peut vérifier aisément ce que l'on vient d'affirmer : pour tout $A \in \mathcal{P}(\Omega)$ on a

$$\mathcal{L}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\left(\bigcup_{a \in A} \{X = a\}\right) = \sum_{a \in A} \mathbb{P}(X = a) = \sum_{a \in \text{Im } X} \mathbb{P}(X = a) \delta_a(A).$$

Remarque 3.15

1. Remarquons que l'écriture $\mathbb{P}(X = a)$ est une abréviation pour $\mathbb{P}(\{X = a\})$ qui est la probabilité de l'événement $\{\omega \in \Omega : X(\omega) = a\} = X^{-1}(\{a\})$.
2. Une variable étagée est discrète.
3. Si on revisite le calcul des loi de l'exemple 3.13, alors pour (1) on peut écrire que $\mathcal{L}_X = \sum_{k=0}^6 2^{-6} \binom{6}{k} \delta_k$. Pour le (2), on peut écrire $\mathcal{L}_X = \frac{1}{36} \delta_2 + \frac{1}{18} \delta_3 + \frac{1}{12} \delta_4 + \dots$

On rappelle que dx désigne implicitement la mesure de Lebesgue sur \mathbb{R} .

Définition 3.16 (Variable aléatoire à densité)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle sur cet espace. On dit que X est à densité s'il existe une fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}_+$ telle que $\mathcal{L}_X(A) = \int_A f(x)dx$ pour tout $A \in \mathcal{B}(\mathbb{R})$.

Nécessairement une telle fonction f est d'intégrale égale à 1. La définition ci-dessus ainsi que celle qui la précède s'étendent aux variables aléatoires à valeurs dans \mathbb{R}^n .

Exemple 3.17 (Lois classiques discrètes)

1. *Loi uniforme* : soit S un ensemble fini de cardinal n . On dit qu'une variable aléatoire X suit une loi uniforme sur S si $\mathbb{P}(X = a) = \frac{1}{n}$ pour tout $a \in S$.
2. *Loi de Bernoulli de paramètre $p \in [0, 1]$* . C'est la loi d'une variable aléatoire X à valeurs dans $\{0, 1\}$ telle que $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$.
Notons que ceci décrit seulement le côté "loi" de la variable aléatoire et ne dit rien sur son côté "application" à part l'ensemble des valeurs prises par cette variable. Notons que toute variable de Bernoulli peut être représentée comme l'indicatrice d'un évènement de probabilité p .
3. *Loi binomiale $\mathcal{B}(n, p)$* . C'est la loi d'une variable aléatoire X à valeurs dans $\{0, \dots, n\}$ telle que $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ pour tout $k \in \{0, \dots, n\}$.
4. *Loi géométrique de paramètre p* . C'est la loi d'une variable aléatoire X à valeurs dans \mathbb{N} telle que $\mathbb{P}(X = k) = (1 - p)p^k$ pour tout $k \in \mathbb{N}$.
5. *Loi de Poisson de paramètre $\alpha > 0$* . C'est la loi d'une variable aléatoire X à valeurs dans \mathbb{N} telle que $\mathbb{P}(X = k) = \frac{\alpha^k}{k!} e^{-\alpha}$ pour tout $k \in \mathbb{N}$.

Exemple 3.18 (Lois classiques à densité)

1. *Loi uniforme sur $[a, b]$* . C'est la loi de la variable aléatoire à valeurs dans \mathbb{R} et de densité $f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$.
2. *Loi exponentielle de paramètre $\alpha > 0$* . C'est la loi de la variable aléatoire à valeurs dans \mathbb{R} et de densité $f(x) = \alpha e^{-\alpha x} \mathbf{1}_{\mathbb{R}_+}(x)$.
3. *Loi Gaussienne $\mathcal{N}(m, \sigma^2)$* . C'est la loi de la variable aléatoire à valeurs dans \mathbb{R} et de densité $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$.

III Fonction de répartition

Définition 3.19 (Fonction de répartition d'une variable aléatoire réelle)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. On définit $F_X : \mathbb{R} \rightarrow [0, 1]$, la fonction de répartition de X , par

$$F_X(t) = \mathcal{L}_X(-\infty, t] = \mathbb{P}(X \leq t),$$

pour tout $t \in \mathbb{R}$.

Remarque 3.20

1. Clairement, F_X est croissante puisque si $t \leq t'$ alors $(-\infty, t] \subseteq (-\infty, t']$ et donc $\mathbb{P}((-\infty, t]) \leq \mathbb{P}((-\infty, t'])$.
2. On a également $\lim_{t \rightarrow -\infty} F_X(t) = \lim_{t \rightarrow -\infty} \mathbb{P}(X \in (-\infty, t]) = \mathbb{P}(X \in \bigcap_{t \in \mathbb{R}} (-\infty, t]) = \mathbb{P}(X \in \emptyset) = 0$, par une application du Lemme 1.13.
3. De même, on a $\lim_{t \rightarrow \infty} F_X(t) = \lim_{t \rightarrow \infty} \mathbb{P}(X \in (-\infty, t]) = \mathbb{P}(X \in \bigcup_{t \in \mathbb{R}} (-\infty, t]) = \mathbb{P}(X \in \mathbb{R}) = 1$.
4. Finalement, F_X est continue à droite. En effet, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(X \leq t + n^{-1}) = \mathbb{P}(X \in \bigcap_{n \rightarrow \infty} (-\infty, t + n^{-1}]) = \mathbb{P}(X \leq t).$$

Dans la remarque précédente, on a regroupé les propriétés satisfaites par une fonction de répartition d'une variable aléatoire. Il s'avère que toute fonction satisfaisant ces propriétés est la fonction de répartition d'une variable aléatoire.

Proposition 3.21

Soit $F : \mathbb{R} \rightarrow [0, 1]$ une fonction qui vérifie les propriétés suivantes :

1. F est croissante.
2. $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow \infty} F(t) = 1$.
3. F est continue à droite.

Alors il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire X tels que $F = F_X$. Par la suite, toute fonction qui satisfait les propriétés ci-dessus sera appelée fonction de répartition.

Démonstration Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, définie sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Notons qu'il est facile de construire une telle variable aléatoire : on peut considérer $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ et $\mathbb{P} = \lambda$, la mesure de Lebesgue ; puis prendre $U : \Omega \rightarrow \Omega$ définie par $U(\omega) := \omega$. Cette variable aléatoire est bien de loi uniforme car $\mathbb{P}(U \leq t) = \lambda(\{\omega \in [0, 1] : \omega \leq t\}) = t$ pour tout $t \in [0, 1]$.

L'idée de la preuve de la proposition est simple : si l'on pose $X := F^{-1}(U)$ alors pour tout $x \in \mathbb{R}$, on a

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(F \circ F^{-1}(U) \leq F(x)) = \mathbb{P}(U \leq F(x)) = F(x),$$

et ainsi X a pour fonction de répartition F . Malheureusement, ce raisonnement est défaillant : l'inverse de F , ici notée F^{-1} , n'est pas toujours bien définie. En effet, si F n'est pas continue et strictement croissante, cette inverse n'existe pas.

Commençons donc par introduire l'inverse généralisée de F

$$F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in]0, 1],$$

avec pour convention $\inf \emptyset = +\infty$. Montrons que pour tout $p \in]0, 1]$ et tout $x \in \mathbb{R}$

$$F(x) \geq p \Leftrightarrow x \geq F^{-1}(p). \tag{3.1}$$

Si $F(x) \geq p$ alors par définition de F^{-1} , on trouve $F^{-1}(p) \leq x$. Réciproquement supposons que $F^{-1}(p) \leq x$. Alors pour tout $\epsilon > 0$ on a $F^{-1}(p) < x + \epsilon$ de sorte que $F(x + \epsilon) \geq p$ par définition de F^{-1} et par croissance de F . En utilisant la continuité à droite de F , on en déduit que $F(x) \geq p$, ce

qui prouve l'équivalence.

On définit alors $X := F^{-1}(U)$. On calcule pour tout $x \in \mathbb{R}$

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

□

Remarque 3.22

1. La donnée de la fonction de répartition rend le calcul des probabilités simple en pratique. En effet on a

$$\mathbb{P}(]a, b]) = F(b) - F(a), \mathbb{P}([a, b]) = F(b) - F(a_-), \mathbb{P}([a, b[) = F(b_-) - F(a_-).$$

2. Si X admet f comme densité alors $\mathbb{P}(]a, b]) = \int_a^b f(x)dx$. Plus généralement, $F(t) = \int_{-\infty}^t f(x)dx$.

3. On a $\mathbb{P}(\{a\}) = F(a) - F(a_-)$ pour tout $a \in \mathbb{R}$. Ainsi, si X est à densité, alors on a $\mathbb{P}(\{a\}) = 0$ pour tout $a \in \mathbb{R}$.

La fonction de répartition caractérise la loi, comme le montre le résultat (plus général) suivant.

Proposition 3.23

Soient X et Y deux variables aléatoires réelles. Les assertions suivantes sont équivalentes :

- (i) $\mathcal{L}_X = \mathcal{L}_Y$,
- (ii) $F_X = F_Y$,
- (iii) $\mathcal{L}_X(O) = \mathcal{L}_Y(O)$ pour tout intervalle ouvert borné $O \subset \mathbb{R}$,
- (iv) $\int \varphi d\mathcal{L}_X = \int \varphi d\mathcal{L}_Y$ pour toute fonction continue bornée $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

Démonstration Le fait que (i) implique les trois autres assertions est immédiat. Montrons que (ii) (resp. (iii)) implique (i). Supposons que les deux mesures coïncident sur la classe $\mathcal{C} = \{]-\infty, t] : t \in \mathbb{R}\}$, resp. $\mathcal{C} = \{O \text{ intervalle ouvert borné}\}$, qui est stable par intersection finie, et engendre la tribu borélienne. Ainsi par le Corollaire 1.17 on en déduit (i).

Nous allons à présent montrer que (iv) implique (ii) ce qui conclura la preuve. Fixons $O =]a, b[$ un intervalle ouvert borné non vide, et introduisons l'approximation suivante de l'indicatrice de O

$$\varphi_\epsilon(x) := \begin{cases} 1 & \text{si } x \in]a + \epsilon, b - \epsilon[\\ 0 & \text{si } x \in \mathbb{R} \setminus]a, b[\\ \frac{x-a}{\epsilon} & \text{si } x \in]a, a + \epsilon] \\ \frac{b-x}{\epsilon} & \text{si } x \in [b - \epsilon, b[\end{cases}$$

Cette fonction peut aussi s'écrire comme $\varphi_\epsilon(x) = 1 \wedge \frac{\text{dist}(x, O^c)}{\epsilon}$, où $u \wedge v = \min(u, v)$. Il s'agit d'une fonction continue bornée. On voit aisément que $\varphi_\epsilon \leq \mathbf{1}_O$, qui est intégrable contre \mathcal{L}_X et \mathcal{L}_Y , et que $\varphi_\epsilon \rightarrow \mathbf{1}_O$ quand $\epsilon \downarrow 0$. Ainsi par le théorème de convergence dominée, on obtient

$$\int \varphi_\epsilon d\mathcal{L}_X \rightarrow \mathcal{L}_X(O), \quad \epsilon \downarrow 0,$$

et de même pour \mathcal{L}_Y .

□

IV Espérance et loi

Dans le chapitre précédent, on a défini l'intégrale par rapport à une mesure quelconque d'une fonction mesurable. En probabilité, on parle d'espérance.

Définition 3.24

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. On définit l'espérance de X par

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

Cette espérance existe si $X \geq 0$ (dans ce cas $\mathbb{E} X \in [0, \infty]$) ou si $\mathbb{E}|X| < \infty$. Lorsque $\mathbb{E}|X| < \infty$, on dit que $X \in L^1$.

Remarque 3.25

1. Si $X = (X_1, \dots, X_n)$ est une variable aléatoire à valeurs dans \mathbb{R}^n alors on peut définir $\mathbb{E} X = (\mathbb{E} X_1, \dots, \mathbb{E} X_n) \in \mathbb{R}^n$.
2. Un cas particulier à garder en tête est le cas où $X = \mathbf{1}_A$ pour $A \in \mathcal{A}$. Dans ce cas, on a $\mathbb{E} \mathbf{1}_A = \mathbb{P}(A)$.

Traduisons les résultats du chapitre précédent dans le langage probabiliste. On rappelle qu'une propriété est vraie presque sûrement si l'ensemble des ω la vérifiant a une probabilité égale à 1.

Proposition 3.26

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

- (i) (Convergence monotone) Si X_n est une suite croissante de variables aléatoires positives et que $X = \lim_{n \rightarrow \infty} X_n$. Alors $\lim_{n \rightarrow \infty} \mathbb{E} X_n = \mathbb{E} X$.
- (ii) (Fatou) Si $X_n \geq 0$ alors $\mathbb{E} \liminf_n X_n \leq \liminf_n \mathbb{E} X_n$.
- (iii) (Convergence dominée) Si $X_n \in L^1$ et que $X_n \rightarrow X$ presque sûrement et $|X_n| \leq Y$ presque sûrement avec $Y \in L^1$ alors

$$\mathbb{E} X_n \xrightarrow{n \rightarrow \infty} \mathbb{E} X \quad \text{et} \quad \mathbb{E} |X_n - X| \xrightarrow{n \rightarrow \infty} 0.$$

Dans le cas particulier d'une variable aléatoire discrète $X \in L^1$, l'espérance admet une expression simple :

$$\mathbb{E} X = \sum_{a \in \text{Im } X} a \mathbb{P}(X = a).$$

De même, si X est une variable aléatoire intégrable de densité f , on a

$$\mathbb{E} X = \int x f(x) dx.$$

Ces identités sont en fait des cas particuliers du résultat général suivant.

Proposition 3.27

Soit X une variable aléatoire à valeurs dans un espace (F, \mathcal{F}) et $h : F \rightarrow \mathbb{R}$ une fonction $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ -

mesurable. Si $h(X)$ est intégrable alors

$$\mathbb{E}[h(X)] = \int_F h d\mathcal{L}_X .$$

Rappelons que \mathcal{L}_X est la loi de X , c'est-à-dire, la mesure image de \mathbb{P} par X : il s'agit d'une mesure de probabilité sur (F, \mathcal{F}) . On retrouve les identités précédentes en prenant $F = \mathbb{R}$ et $h(x) = x$. Notons aussi que l'espérance d'une variable aléatoire réelle $X \in L^1$ peut donc s'écrire

$$\mathbb{E} X = \int X(\omega) d\mathbb{P}(\omega) = \int x d\mathcal{L}_X(x) .$$

Cette identité permet de déduire que l'espérance d'une variable aléatoire ne dépend que de sa loi \mathcal{L}_X . Ainsi deux variables aléatoires ayant même loi ont même espérance.

Démonstration Il suffit de traiter le cas où h est positive, le cas général s'obtenant en décomposant h en ses parties positives et négatives. Comme toute fonction mesurable positive h est limite croissante de fonctions étagées positives h_n , le théorème de convergence monotone affirme que l'égalité de l'énoncé est vraie pour peu qu'elle soit vraie sur h_n . Il suffit donc de considérer le cas où h est étagée positive. Une telle fonction h peut s'écrire $h = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ avec $a_i > 0$ et $A_i \in \mathcal{F}$. Ainsi par linéarité

$$\mathbb{E}[h(X)] = \mathbb{E}\left[\sum_{i=1}^n a_i \mathbf{1}_{A_i}(X)\right] = \sum_{i=1}^n a_i \mathbb{E}[\mathbf{1}_{A_i}(X)] .$$

Or $\mathbf{1}_{A_i}(X)$ est une variable aléatoire de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ qui vérifie pour tout $\omega \in \Omega$

$$\mathbf{1}_{A_i}(X)(\omega) = \mathbf{1}_{B_i}(\omega) ,$$

avec $B_i = X^{-1}(A_i) \in \mathcal{A}$. Par la définition de l'intégrale vue au chapitre précédent on en déduit donc que

$$\mathbb{E}[\mathbf{1}_{A_i}(X)] = \mathbb{P}(B_i) = \mathbb{P}(X \in A_i) ,$$

et ainsi

$$\mathbb{E}[h(X)] = \sum_{i=1}^n a_i \mathbb{P}(X \in A_i) .$$

Par ailleurs, par définitions de l'intégrale des fonctions étagées et de la loi de X on a

$$\int_F h d\mathcal{L}_X = \sum_{i=1}^n a_i \mathcal{L}_X(A_i) = \sum_{i=1}^n a_i \mathbb{P}(X \in A_i) ,$$

ce qui achève la preuve. □

V Moments et inégalités classiques

Définition 3.28

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. On dit que $X \in L^p$, pour $p \geq 1$, si $\mathbb{E}|X|^p < \infty$. On définit dans ce cas $\|X\|_{L^p} = (\mathbb{E}|X|^p)^{\frac{1}{p}}$.

Remarque 3.29

On peut vérifier que $\|\cdot\|_{L^p}$ définit bien une norme à condition que l'on ait quotienté l'espace des variables aléatoires par la relation d'équivalence "égalité \mathbb{P} -presque partout". L'espace L^p est alors constitué de classes d'équivalence, et l'écriture $X \in L^p$ signifie que la classe d'équivalence de X est dans L^p .

Proposition 3.30 (Inégalité de Jensen)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, $X \in L^1$ une variable aléatoire et $g : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction convexe. Alors

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

La preuve repose sur le fait que toute fonction convexe s'écrit comme un supremum de fonctions affines, et sur le fait que l'inégalité est vraie pour les fonctions affines.

Corollaire 3.31 (Croissance des normes L^p)

Soit $1 \leq p \leq r < \infty$ et $X \in L^r$. Alors $X \in L^p$ et $\|X\|_{L^p} \leq \|X\|_{L^r}$.

Démonstration On applique l'inégalité de Jensen à $|X|^{p\mathbf{1}_{|X|<n}}$ (qui est intégrable) avec la fonction convexe

$$h(x) = \begin{cases} x^{r/p} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

pour $1 \leq p \leq r < \infty$. On obtient alors

$$\mathbb{E}[|X|^{p\mathbf{1}_{|X|<n}}]^{r/p} \leq \mathbb{E}[|X|^r \mathbf{1}_{|X|<n}].$$

Par le théorème de convergence monotone on en déduit que

$$\mathbb{E}[|X|^p]^{r/p} \leq \mathbb{E}[|X|^r],$$

ce qui assure que $X \in L^p$ ainsi que l'inégalité voulue. □

Dans le cas $p = 2$ on obtient l'inégalité suivante.

Proposition 3.32 (Inégalité de Cauchy-Schwarz)

Soit $X, Y \in L^2$. Alors $XY \in L^1$ et $|\mathbb{E}[XY]| \leq \mathbb{E}|XY| \leq \|X\|_{L^2}\|Y\|_{L^2}$.

Démonstration La première affirmation résulte de l'inégalité

$$2|XY| \leq X^2 + Y^2.$$

Alors, en développant la relation $\mathbb{E}[(a|X| + |Y|)^2] \geq 0$ pour $a \in \mathbb{R}$, on obtient que le trinôme $a^2\mathbb{E}[X^2] + 2a\mathbb{E}[|XY|] + \mathbb{E}[Y^2] \geq 0$ pour tout a . Ainsi le discriminant est négatif, ce qui donne

$$\mathbb{E}[|XY|] \leq \|X\|_{L^2}\|Y\|_{L^2}.$$

Or par les propriétés de l'intégrale on sait que $|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|]$, ce qui conclut la preuve. □

Une inégalité très utile est l'inégalité de Markov qui permet de contrôler les queues de distribution à l'aide de l'espérance.

Proposition 3.33 (Inégalité de Markov)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, X une variable aléatoire réelle intégrable. Alors pour tout $t > 0$, on a

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E} X}{t}.$$

Démonstration On a

$$\mathbb{P}(X \geq t) = \mathbb{E} \mathbf{1}_{\{X \geq t\}}.$$

Or

$$t \mathbf{1}_{\{X \geq t\}} \leq X,$$

de sorte que

$$\mathbb{E} [t \mathbf{1}_{\{X \geq t\}}] = t \mathbb{P}(X \geq t) \leq \mathbb{E} [X].$$

□

Définition 3.34 (Variance et covariance)

Soit $X \in L^2$. Alors on définit la variance de X par

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

Si de plus $Y \in L^2$, alors on définit la covariance de X et Y par

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

Remarque 3.35

En particulier on a $\text{Var}(X) = \text{Cov}(X, X)$. La proposition précédente justifie qu'on peut écrire $\mathbb{E} [XY]$ puisque $XY \in L^1$ lorsque $X, Y \in L^2$.

La transformée de Laplace est un outil pratique pour calculer les moments d'une variable aléatoire.

Définition 3.36 (Série génératrice, transformée de Laplace)

— Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle. La transformée de Laplace de X est définie par

$$M_X(\lambda) = \mathbb{E} e^{\lambda X}, \quad \lambda \in \mathbb{R}.$$

— Lorsque X est à valeurs entières, on définit plutôt sa série génératrice,

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

Remarque 3.37

1. La quantité $\mathbb{E} e^{\lambda X}$ a toujours du sens car il s'agit de l'espérance d'une variable aléatoire positive. L'ensemble des $\lambda \in \mathbb{R}$ pour lesquels cette quantité est finie est un intervalle qui contient 0 (exercice). Si cet intervalle contient un voisinage de 0, alors par les théorèmes de dérivation sous l'intégrale on peut montrer que $M'_X(\lambda) = \mathbb{E} [X e^{\lambda X}]$ de telle sorte que

$\mathbb{E} X = M'_X(0)$. De la même manière, on a $\mathbb{E} X^p = M_X^{(p)}(0)$.

2. Si $X = (X_1, \dots, X_n)$ est à valeurs dans \mathbb{R}^n , alors on peut définir $M_X : \mathbb{R}^n \rightarrow \mathbb{R}$ par

$$M_X(\lambda) = \mathbb{E}[e^{\langle \lambda, X \rangle}] = \mathbb{E} \prod_{i=1}^n e^{\lambda_i X_i},$$

pour tout $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$.

3. Dans le cas où X est à valeurs entières,

$$G_X(z) = \sum_{k=0}^{\infty} P(X = k) z^k$$

est une série entière de rayon de convergence $R \geq 1$ et à coefficients positifs.

Exemple 3.38

1. Si $X \sim \mathcal{N}(0, 1)$ alors $M_X(\lambda) = e^{\lambda^2/2}$ pour tout $\lambda \in \mathbb{R}$. En effet

$$\mathbb{E} e^{\lambda X} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\lambda x} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{\lambda^2}{2}} \int_{\mathbb{R}} e^{-\frac{(x-\lambda)^2}{2}} dx = e^{\frac{\lambda^2}{2}}.$$

2. Si $X \sim \text{Ber}(1/2)$ alors pour tout $\lambda \in \mathbb{R}$

$$M_X(\lambda) = \frac{e^{\lambda} + e^{-\lambda}}{2} = \cosh(\lambda).$$

VI Fonction caractéristique

Nous avons déjà vu une caractérisation de la loi d'une variable aléatoire réelle au travers de sa fonction de répartition. Une autre caractérisation est donnée par sa transformée de Fourier, aussi appelée fonction caractéristique.

Définition 3.39 (Fonction caractéristique)

Soit X une variable aléatoire réelle. La fonction caractéristique $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ de X au point $t \in \mathbb{R}$ est définie par

$$\phi_X(t) = \mathbb{E} e^{itX} = \mathbb{E}[\cos(tX)] + i \mathbb{E}[\sin(tX)].$$

Plus généralement, soit X une variable aléatoire à valeurs dans \mathbb{R}^d . La fonction caractéristique de X est définie par $\phi_X(t) = \mathbb{E}[\exp(i\langle X, t \rangle)]$ pour tout $t \in \mathbb{R}^d$, où $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans \mathbb{R}^d .

Remarque 3.40

1. Nous n'avons pas introduit formellement l'intégrale des fonctions à valeurs dans \mathbb{C} . Soit $f : S \rightarrow \mathbb{C}$. On dit que f est mesurable (resp. intégrable) si ses parties réelles et imaginaires le sont, et l'on pose alors $\int f d\mu = \int \Re(f) d\mu + i \int \Im(f) d\mu$.

2. Si $Z = X + iY$ avec $X, Y \in L^1$. Alors $|\mathbb{E} Z| \leq \mathbb{E} |Z|$ où $|Z|$ ici veut dire le module de Z . En effet, on écrit

$$|\mathbb{E} Z| = \mathbb{E} Z \cdot \frac{\overline{\mathbb{E} Z}}{|\mathbb{E} Z|} = \operatorname{Re}(\mathbb{E} Z \cdot \frac{\overline{\mathbb{E} Z}}{|\mathbb{E} Z|}) = \mathbb{E} \operatorname{Re}(Z \cdot \frac{\overline{\mathbb{E} Z}}{|\mathbb{E} Z|}),$$

or $\operatorname{Re}(Z \cdot \frac{\overline{\mathbb{E} Z}}{|\mathbb{E} Z|}) \leq |Z|$ ce qui termine la preuve.

Ainsi, on a toujours $|\phi_X(t)| \leq 1$.

3. On a $\phi_X(0) = 1$. On peut vérifier que $\overline{\phi_X} = \phi_{-X}$ et que pour tout $a, b \in \mathbb{R}$, on a $\phi_{aX+b}(t) = e^{itb} \phi_X(at)$.

4. ϕ_X est uniformément continue. En effet, on a

$$\sup_{t \in \mathbb{R}} |\phi_X(t + \varepsilon) - \phi_X(t)| \leq \mathbb{E} |e^{i\varepsilon X} - 1|.$$

Comme $|e^{i\varepsilon X} - 1| \xrightarrow{\varepsilon \rightarrow 0} 0$ alors par le théorème de convergence dominée on obtient que $\mathbb{E} |e^{i\varepsilon X} - 1| \xrightarrow{\varepsilon \rightarrow 0} 0$.

Un cas important est le cas gaussien présenté dans l'exemple suivant.

Exemple 3.41

Soit $X \sim \mathcal{N}(0, \sigma^2)$. Alors $\phi_X(t) = e^{-\frac{\sigma^2 t^2}{2}}$ pour tout $t \in \mathbb{R}$.

En effet, on a par définition que

$$\phi_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} e^{itx} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} \cos(tx) dx + \frac{i}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} \sin(tx) dx$$

Comme $e^{-\frac{x^2}{2\sigma^2}} \sin(tx)$ est impaire, alors la deuxième intégrale ci-dessus est nulle. Ainsi, on a

$$\phi_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} \cos(tx) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} g(t, x) dx$$

Or, pour tout $t \in \mathbb{R}$, on a $g(t, x)$ est intégrable par rapport à x (car majorée par $e^{-\frac{x^2}{2\sigma^2}}$ qui l'est). De plus $g(t, x)$ est de classe C^1 par rapport à t puisque $\frac{\partial g}{\partial t} = -xe^{-\frac{x^2}{2\sigma^2}} \sin(tx)$ est continue. Finalement, on a pour tout t que $\frac{\partial g}{\partial t} \leq |x|e^{-\frac{x^2}{2\sigma^2}}$ qui est intégrable. Ainsi, on peut dériver sous l'intégrale et écrire

$$\phi'_X(t) = -\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} xe^{-\frac{x^2}{2\sigma^2}} \sin(tx) dx$$

On intègre par parties pour avoir

$$\phi'_X(t) = -\frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma^2}} t \cos(tx) dx = -\sigma^2 t \phi_X(t).$$

En résolvant l'équation différentielle et en utilisant que $\phi_X(0) = 1$, on obtient $\phi_X(t) = e^{-\frac{\sigma^2 t^2}{2}}$ qui est égale à la densité d'une Gaussienne $\mathcal{N}(0, \sigma^2)$ correctement normalisée.

Nous avons alors le résultat suivant, qui justifie l'appellation "fonction caractéristique".

Théorème 3.42

La fonction caractéristique caractérise la loi. Plus précisément si X et Y sont deux variables aléatoires réelles alors $\phi_X = \phi_Y$ si et seulement si $\mathcal{L}_X = \mathcal{L}_Y$.

Remarque 3.43

Le résultat est encore valide pour des variables aléatoires à valeurs dans \mathbb{R}^d .

Démonstration L'implication réciproque est claire, il suffit donc de montrer que si $\phi_X = \phi_Y$ alors $\mathcal{L}_X = \mathcal{L}_Y$. Par la caractérisation des mesures de probabilités établie à la Proposition 3.23, il suffit de montrer que pour toute fonction continue bornée $\varphi : \mathbb{R} \rightarrow \mathbb{R}$

$$\int \varphi d\mathcal{L}_X = \int \varphi d\mathcal{L}_Y .$$

Pour ce faire, nous posons

$$g_\sigma(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} ,$$

et nous introduisons la fonction

$$f_{\sigma,X}(x) := \int g_\sigma(x-y)\mathcal{L}_X(dy) ,$$

ainsi que la mesure à densité

$$\mu_{\sigma,X}(dx) := f_\sigma(x)dx .$$

Nous considérons des définitions analogues pour $f_{\sigma,Y}$ et $\mu_{\sigma,Y}$. Commençons par montrer que $\mu_{\sigma,X} = \mu_{\sigma,Y}$. Le cas gaussien traité dans l'exemple précédent assure que

$$\int e^{iuz} g_{1/\sigma}(z)dz = \sigma\sqrt{2\pi}g_\sigma(u) ,$$

et ainsi, par une application du théorème de Fubini on obtient

$$\begin{aligned} f_{\sigma,X}(x) &= \int g_\sigma(x-y)\mathcal{L}_X(dy) = \frac{1}{\sigma\sqrt{2\pi}} \int \int e^{i(x-y)z} g_{1/\sigma}(z)dz \mathcal{L}_X(dy) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int e^{ixz} g_{1/\sigma}(z) \int e^{-iyz} \mathcal{L}_X(dy) dz \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int e^{ixz} g_{1/\sigma}(z) \Phi_X(-z) dz . \end{aligned}$$

L'égalité $\phi_X = \phi_Y$ assure donc que $f_{\sigma,X}(x) = f_{\sigma,Y}(x)$ et donc que $\mu_{\sigma,X} = \mu_{\sigma,Y}$.

Fixons à présent une fonction continue bornée φ . Par le théorème de Fubini il vient

$$\int \varphi d\mu_{\sigma,X} = \int \varphi_\sigma d\mathcal{L}_X ,$$

avec

$$\varphi_\sigma(y) = \int_x \varphi(x)g_\sigma(x-y)dx .$$

En utilisant les propriétés de la densité gaussienne et le caractère continu borné de φ , on peut montrer que $\varphi_\sigma(y) \rightarrow \varphi(y)$ pour tout $y \in \mathbb{R}$ quand $\sigma \downarrow 0$, et que $\sup_y |\varphi_\sigma(y)| \leq \sup_y |\varphi(y)|$. Une application du théorème de convergence dominée donne alors

$$\int \varphi_\sigma d\mathcal{L}_X \rightarrow \int \varphi d\mathcal{L}_X , \quad \sigma \downarrow 0 ,$$

ce qui permet de conclure. □

Remarque 3.44

En anticipant des notions introduites aux chapitres suivants, on remarquera que $\mu_{\sigma, X}$ est la loi de $X + Z_\sigma$, où Z_σ est une variable aléatoire $\mathcal{N}(0, \sigma^2)$ indépendante de X . Les deux étapes de la preuve sont : 1) la loi de $X + Z_\sigma$ est entièrement caractérisée par ϕ_X ; 2) la loi de $X + Z_\sigma$ converge vers celle de X quand $\sigma \downarrow 0$.

Chapitre 4

Indépendance

I Événements indépendants et tribus indépendantes

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

Définition 4.1 (*Événements indépendants*)

On dit que deux événements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

On dit que n événements A_1, \dots, A_n sont indépendants si $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_n})$ pour tout sous-ensemble $\{i_1, \dots, i_n\}$ de $\{1, \dots, n\}$.

Remarque 4.2

Il ne faut pas confondre la notion d'indépendance de n événements avec l'indépendance deux à deux. En effet, si l'on lance à deux reprises une pièce de monnaie et si l'on considère les événements $A = \{\text{Pile au premier lancer}\}$, $B = \{\text{Pile au second lancer}\}$, et $C = \{\text{même résultat aux deux lancers}\}$ alors on peut vérifier que ces événements sont deux à deux indépendants mais ne sont pas indépendants.

L'indépendance de n événements A_1, \dots, A_n implique l'indépendance de toute collection formée à partir des A_i ou de leurs complémentaires, et réciproquement :

Proposition 4.3

Les événements A_1, \dots, A_n sont indépendants si et seulement si $\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n)$ pour tous $B_i \in \sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$, $i = 1, \dots, n$.

Démonstration Le sens réciproque est facile. Le sens direct repose sur le fait suivant : si C_1, \dots, C_k sont indépendants alors $C_1^c, C_2^c, \dots, C_k^c$ le sont aussi, en effet pour tout $2 \leq i_1 < \dots < i_p \leq k$

$$\begin{aligned} \mathbb{P}(C_1^c \cap C_{i_2} \cap \dots \cap C_{i_p}) &= \mathbb{P}(C_{i_2} \cap \dots \cap C_{i_p}) - \mathbb{P}(C_1 \cap C_{i_2} \cap \dots \cap C_{i_p}) \\ &= \mathbb{P}(C_{i_2}) \dots \mathbb{P}(C_{i_p}) - \mathbb{P}(C_1) \dots \mathbb{P}(C_{i_p}) \\ &= \mathbb{P}(C_1^c) \mathbb{P}(C_{i_2}) \dots \mathbb{P}(C_{i_p}) . \end{aligned}$$

La fin de la preuve est laissée en exercice. □

La proposition précédente montre que l'indépendance de n événements peut être interprétée comme l'indépendance des n tribus associées, pour peu que l'on définisse l'indépendance de tribus de la façon suivante.

Définition 4.4 (Tribus indépendantes)

Soient $\mathcal{A}_1, \dots, \mathcal{A}_n$ des tribus contenues dans \mathcal{A} . On dit qu'elles sont indépendantes si

$$\forall A_1 \in \mathcal{A}_1, \dots, \forall A_n \in \mathcal{A}_n, \quad \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

En général, les tribus sont des objets très gros, et l'on pourrait souhaiter prouver leur indépendance en se restreignant à des collections d'ensembles plus petites :

Proposition 4.5

Soient $\mathcal{A}_1, \dots, \mathcal{A}_n \subset \mathcal{A}$ des σ -algèbres. Pour tout $i \in \{1, \dots, n\}$, soit \mathcal{C}_i une collection d'ensembles stable par intersection finie telle que $\sigma(\mathcal{C}_i) = \mathcal{A}_i$. Si pour tous $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ on a

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n),$$

alors les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont indépendantes.

Démonstration On fixe d'abord $A_2 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ et on pose

$$\mathcal{M}_1 = \{B_1 \in \mathcal{A}_1 : \mathbb{P}(B_1 \cap \dots \cap A_n) = \mathbb{P}(B_1) \dots \mathbb{P}(A_n)\}.$$

On peut vérifier que \mathcal{M}_1 est une classe monotone et par hypothèse $\mathcal{C}_1 \subset \mathcal{A}_1$. Ainsi par le lemme de classe monotone on a $\mathcal{M}_1 = \sigma(\mathcal{C}_1) = \mathcal{A}_1$.

On fixe maintenant $B_1 \in \mathcal{A}_1, A_3 \in \mathcal{C}_3, \dots, A_n \in \mathcal{C}_n$ et on pose

$$\mathcal{M}_2 = \{A_2 \in \mathcal{A}_2 : \mathbb{P}(B_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(B_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)\}.$$

On montre que \mathcal{M}_2 est une classe monotone et l'on utilise le lemme de classe monotone pour déduire que $\mathcal{M}_2 = \mathcal{A}_2$. On termine la preuve par récurrence. □

Une des conséquences de cette proposition est qu'on puisse affirmer l'indépendance de paquets de σ -algèbres si ces paquets sont formés à partir de σ -algèbres indépendantes.

Corollaire 4.6 (Regroupements par paquets)

Soient $\mathcal{A}_1, \dots, \mathcal{A}_n$ des σ -algèbres indépendantes. Soient $n_0 < n_1 < n_2 \dots < n_k = n$. Alors les σ -algèbres

$$\mathcal{D}_1 = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_{n_1}), \mathcal{D}_2 = \sigma(\mathcal{A}_{n_1+1}, \dots, \mathcal{A}_{n_2}), \dots, \mathcal{D}_k = \sigma(\mathcal{A}_{n_{k-1}+1}, \dots, \mathcal{A}_{n_k})$$

sont indépendantes.

Démonstration Pour $j \in \{1, \dots, k\}$, on définit \mathcal{C}_j comme l'ensemble des parties de la forme

$$A_{n_{j-1}+1} \cap \dots \cap A_{n_j},$$

avec $A_i \in \mathcal{A}_i$ pour tout $i \in \{n_{j-1} + 1, \dots, n_j\}$. Clairement, les \mathcal{C}_j sont stables par intersection finie et engendrent les \mathcal{D}_j . Le résultat découle de la proposition précédente. □

Définition 4.7 (Indépendance d'une collection infinie de tribus)

On dit qu'une collection infinie (dénombrable ou pas) de tribus sont indépendantes si l'on a l'indépendance de toute sous-collection finie de ces tribus.

II Variables aléatoires indépendantes

Définition 4.8 (*Variables aléatoires indépendantes*)

On dit que n variables aléatoires X_1, \dots, X_n à valeurs dans $(F_1, \mathcal{F}_1), \dots, (F_n, \mathcal{F}_n)$ sont indépendantes si les σ -algèbres $\sigma(X_1), \dots, \sigma(X_n)$ le sont. Plus précisément, si pour tout $A_i \in \mathcal{F}_i$, $1 \leq i \leq n$, on a

$$\mathbb{P}(\{X_1 \in A_1\} \cap \{X_2 \in A_2\} \cap \dots \cap \{X_n \in A_n\}) = \mathbb{P}(\{X_1 \in A_1\})\mathbb{P}(\{X_2 \in A_2\}) \dots \mathbb{P}(\{X_n \in A_n\}).$$

Remarque 4.9

1. Soient $\mathcal{A}_1, \dots, \mathcal{A}_n \subset \mathcal{A}$ des σ -algèbres indépendantes. Si pour tout $i \in \{1, \dots, n\}$, X_i est \mathcal{A}_i -mesurable alors X_1, \dots, X_n sont indépendantes. En effet pour tout $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, on a $X_1^{-1}(A_1) \in \mathcal{A}_1, \dots, X_n^{-1}(A_n) \in \mathcal{A}_n$ (car X_i est \mathcal{A}_i -mesurable) et ces événements sont donc indépendants; comme ceci est vrai pour tout choix de A_1, \dots, A_n alors les variables X_1, \dots, X_n sont indépendantes.
2. Soient X_1, \dots, X_n des variables aléatoires indépendantes et soient $n_0 < n_1 < n_2 \dots < n_k = n$. Alors, par le corollaire précédent, les variables aléatoires $Y_1 = (X_1, \dots, X_{n_1}), \dots, Y_k = (X_{n_{k-1}+1}, \dots, X_{n_k})$ sont indépendantes.

Définition 4.10 (*Loi jointe*)

Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $(F_1, \mathcal{F}_1), \dots, (F_n, \mathcal{F}_n)$ respectivement. La loi jointe de X_1, \dots, X_n est la loi du n -uplet (X_1, \dots, X_n) dans $F_1 \times \dots \times F_n$ muni de la σ -algèbre produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$. On notera $\mathcal{L}_{(X_1, \dots, X_n)}$ la loi jointe.

Exemple 4.11

Si X est une Bernoulli standard et $Y = 1 - X$. Alors la loi jointe de X, Y est donnée par

$$\mathcal{L}_{(X,Y)}(\{(1, 0)\}) = \mathcal{L}_{(X,Y)}(\{(0, 1)\}) = \frac{1}{2}.$$

On dit que les \mathcal{L}_{X_i} sont les lois marginales de $\mathcal{L}_{(X_1, \dots, X_n)}$. On a la caractérisation suivante de la l'indépendance de variables aléatoires.

Proposition 4.12

Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $(F_1, \mathcal{F}_1), \dots, (F_n, \mathcal{F}_n)$ respectivement. On a équivalence entre :

1. les X_1, \dots, X_n sont indépendantes
2. la loi jointe est égale au produit des lois marginales i.e. $\mathcal{L}_{(X_1, \dots, X_n)} = \mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}$.
3. pour toutes fonctions mesurables positives $f_i : F_i \rightarrow [0, \infty]$, on a

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

Démonstration 1. \Rightarrow 2. Supposons que X_1, \dots, X_n sont indépendantes et montrons que $\mathcal{L}_{(X_1, \dots, X_n)} =$

$\mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}$. Soient $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$ et écrivons

$$\mathcal{L}_{(X_1, \dots, X_n)}(A_1 \times \dots \times A_n) = \mathbb{P}((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) = \mathbb{P}(\{X_1 \in A_1\} \cap \dots \cap \{X_n \in A_n\})$$

et par indépendance des X_i , on obtient

$$\mathcal{L}_{(X_1, \dots, X_n)}(A_1 \times \dots \times A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

qui par définition n'est autre que $\mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}(A_1 \times \dots \times A_n)$. Ceci montre que $\mathcal{L}_{(X_1, \dots, X_n)}$ et $\mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}$ coïncident sur les pavés mesurables (qui sont stables par intersection finie et engendrent la tribu produit) donc par le corollaire 1.17, on a l'égalité voulue.

2. \Rightarrow 3. Supposons que $\mathcal{L}_{(X_1, \dots, X_n)} = \mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}$. On utilise le fait que la loi jointe est égale au produit des lois pour écrire que

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \int_{S_1 \times \dots \times S_n} \prod_{i=1}^n f_i(X_i) \mathcal{L}_{(X_1, \dots, X_n)}(dx_1, \dots, dx_n) = \int_{S_1 \times \dots \times S_n} \prod_{i=1}^n f_i(X_i) \mathcal{L}_{X_1}(dx_1) \otimes \dots \otimes \mathcal{L}_{X_n}(dx_n).$$

Par Fubini, on obtient que

$$\int_{S_1 \times \dots \times S_n} \prod_{i=1}^n f_i(X_i) \mathcal{L}_{X_1}(dx_1) \otimes \dots \otimes \mathcal{L}_{X_n}(dx_n) = \prod_{i=1}^n \int_{S_i} f_i(X_i) \mathcal{L}_{X_i}(dx_i) = \prod_{i=1}^n \mathbb{E} [f_i(X_i)],$$

ce qui termine la preuve.

3. \Rightarrow 1. Il suffit de prendre pour f_i l'indicatrice d'un événement A_i . □

Remarque 4.13

Dans le point 3., on aurait pu remplacer "positive" par intégrable (contre la loi de X_i).

Dans le cas particulier où les variables aléatoires sont réelles, on a le résultat suivant.

Proposition 4.14

Soient X_1, \dots, X_n des variables aléatoires réelles. On a équivalence entre :

1. les X_1, \dots, X_n sont indépendantes
2. pour toutes fonctions continues bornées $f_i : \mathbb{R} \rightarrow \mathbb{R}$, on a

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

3. les fonctions caractéristiques vérifient $\phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n \phi_{X_i}(t_i)$ pour tous $t_1, \dots, t_n \in \mathbb{R}$.

Démonstration Le fait que 1. implique 2. et 3. est une conséquence de la remarque précédente (dans le cas de la fonction caractéristique, on peut revenir à la définition de l'intégrale de fonction à valeurs complexes en distinguant parties positives et négatives). Montrons que 2. implique 1. Par le corollaire 1.17, il suffit de prouver que pour tous intervalles ouverts bornés I_1, \dots, I_n on a

$$\mathcal{L}_{(X_1, \dots, X_n)}(I_1 \times \dots \times I_n) = \mathcal{L}_{X_1}(I_1) \dots \mathcal{L}_{X_n}(I_n).$$

Or on peut approximer tout intervalle ouvert par une suite de fonctions croissantes et continues bornées (voir la preuve de la Proposition 3.23), de sorte que le Théorème de Convergence Monotone assure l'égalité voulue.

Pour montrer que 3. implique 1., il suffit d'observer que $\prod_{i=1}^n \phi_{X_i}(t_i)$ est la fonction caractéristique associée à la loi produit $\mathcal{L}_{X_1} \otimes \dots \otimes \mathcal{L}_{X_n}$ puis d'appliquer le Théorème 3.42, qui est valide en toute dimension. \square

Remarque 4.15

En combinant cette proposition avec le théorème de Fubini, on peut montrer que si X_1, \dots, X_n sont des variables aléatoires réelles à densité f_i alors (X_1, \dots, X_n) a une densité dans \mathbb{R}^n donnée par le produit des densités. La réciproque est également vraie : si (X_1, \dots, X_n) a une densité décomposable en un produit de densités marginales alors les X_i sont indépendantes.

Corollaire 4.16

Si X et Y sont des variables aléatoires réelles indépendantes alors $\mathcal{L}_{X+Y} = \mathcal{L}_X * \mathcal{L}_Y$ où $*$ est le produit de convolution i.e. pour tout $A \in \mathcal{B}(\mathbb{R})$ on a

$$\mathcal{L}_{X+Y}(A) = \int_{\mathbb{R}^2} \mathbf{1}_A(x+y) \mathcal{L}_X(dx) \mathcal{L}_Y(dy).$$

Démonstration La preuve est triviale puisque $\mathcal{L}_{X+Y}(A) = \mathbb{E} \mathbf{1}_A \circ h(X, Y)$ où $h(X, Y) = X + Y$. Donc

$$\mathcal{L}_{X+Y}(A) = \int \mathbf{1}_A(x+y) \mathcal{L}_{(X,Y)}(dx, dy).$$

Comme $\mathcal{L}_{(X,Y)} = \mathcal{L}_X \otimes \mathcal{L}_Y$, on termine la preuve. \square

III Deuxième lemme de Borel-Cantelli

Lemme 4.17 (Deuxième lemme de Borel-Cantelli)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements indépendants. Alors

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty \quad \Rightarrow \quad \mathbb{P}(\limsup A_n) = 1.$$

Démonstration On rappelle que $\limsup A_n = \bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} A_n$. Par indépendance des A_n , on a que pour tout $m \in \mathbb{N}$,

$$\mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right) = \prod_{n=m}^{\infty} (1 - \mathbb{P}(A_n)) \leq \prod_{n=m}^{\infty} \exp(-\mathbb{P}(A_n)) \leq \exp\left(-\sum_{n \geq m} \mathbb{P}(A_n)\right),$$

où on a utilisé que $1 - x \leq e^{-x}$ pour tout $x \geq 0$. Comme $\sum_n \mathbb{P}(A_n)$ est une série divergente, on déduit que pour tout $m \in \mathbb{N}$ on a $\mathbb{P}\left(\bigcap_{n \geq m} A_n^c\right) = 0$.

Ceci signifie que $\mathbb{P}\left(\bigcup_{n \geq m} A_n\right) = 1$ et que donc pour tout $m \in \mathbb{N}$ l'événement $\bigcup_{n \geq m} A_n$ est presque sûr. Comme l'intersection dénombrable d'événements presque sûrs est presque sûre on déduit que $\mathbb{P}\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} A_n\right) = 1$ ce qui achève la preuve. \square

Examinons quelques exemples d'applications de ce lemme.

Exemple 4.18

Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre 1 i.e. $\mathbb{P}(X_n \geq t) = e^{-t}$ pour tout $t \geq 0$. Alors

$$\limsup \frac{X_n}{\ln n} = 1 \quad p.s.$$

En effet, on a pour tout $\alpha > 0$

$$\mathbb{P}(X_n \geq \alpha \ln n) = n^{-\alpha},$$

qui forme une série convergente si et seulement si $\alpha > 1$. Ainsi par les deux lemmes de Borel-Cantelli, on a

$$\mathbb{P}(X_n > \alpha \ln n \text{ pour une infinité de } n) = \begin{cases} 0 & \text{si } \alpha > 1 \\ 1 & \text{si } \alpha \leq 1 \end{cases}$$

Clairement, si $X_n \geq \ln n$ pour une infinité de n alors $\limsup \frac{X_n}{\ln n} \geq 1$. Ainsi,

$$\mathbb{P}(\limsup \frac{X_n}{\ln n} \geq 1) \geq \mathbb{P}(X_n > \ln n \text{ pour une infinité de } n) = 1.$$

Donc on a $\limsup \frac{X_n}{\ln n} \geq 1$ p.s. et il reste à montrer que $\mathbb{P}(\limsup \frac{X_n}{\ln n} > 1) = 0$. On écrit

$$\left\{ \limsup \frac{X_n}{\ln n} > 1 \right\} = \bigcup_{k \in \mathbb{N}} \left\{ \limsup \frac{X_n}{\ln n} \geq 1 + \frac{2}{k} \right\}.$$

D'autre part, si $\limsup \frac{X_n}{\ln n} \geq 1 + \frac{2}{k}$ alors forcément $\frac{X_n}{\ln n} \geq 1 + \frac{1}{k}$ pour une infinité de n . Ainsi

$$\mathbb{P}(\limsup \frac{X_n}{\ln n} \geq 1 + \frac{2}{k}) \leq \mathbb{P}(X_n \geq (1 + \frac{1}{k}) \ln n \text{ pour une infinité de } n) = 0.$$

Comme $\left\{ \limsup \frac{X_n}{\ln n} > 1 \right\}$ est l'union dénombrable des ensembles précédents, on déduit que $\mathbb{P}(\limsup \frac{X_n}{\ln n} > 1) = 0$.

Exemple 4.19

Pour tout $n \in \mathbb{N}$, on note A_n l'ensemble des multiples de n . Il n'existe pas de probabilité \mathbb{P} sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ telle que $\mathbb{P}(A_n) = \frac{1}{n}$ pour tout $n \in \mathbb{N}$.

En effet, supposons qu'il en existe une. Soit \mathcal{P} l'ensemble des nombres premiers. Les $(A_p)_{p \in \mathcal{P}}$ sont indépendants puisque pour tous $p_1, \dots, p_k \in \mathcal{P}$ on a

$$\mathbb{P}(A_{p_1} \cap \dots \cap A_{p_k}) = \mathbb{P}(A_{p_1 \dots p_k}) = \frac{1}{p_1 \dots p_k} = \prod_{i=1}^k \mathbb{P}(A_{p_i}).$$

D'autre part, on a $\sum_{p \in \mathcal{P}} \frac{1}{p} = \infty$ car le k -ème nombre premier est d'ordre $k \ln k$ quand $k \rightarrow \infty$. Ainsi, par le deuxième lemme de Borel-Cantelli, on a que \mathbb{P} -presque tout entier est multiple d'une infinité de nombre premiers, ce qui est absurde.

IV Loi du tout ou rien

Théorème 4.20 (Loi du 0/1 de Kolmogorov)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes. Pour tout $n \in \mathbb{N}$, on pose $\mathcal{T}_n = \sigma(X_n, X_{n+1}, \dots)$ et $\mathcal{T}_\infty = \bigcap_{n \in \mathbb{N}} \mathcal{T}_n$. Alors on a

$$\forall A \in \mathcal{T}_\infty, \quad \mathbb{P}(A) \in \{0, 1\}.$$

Démonstration Pour tout $n \in \mathbb{N}$, on pose $\mathcal{B}_n = \sigma(X_1, \dots, X_n)$. Par l'indépendance par regroupements par paquets, on a que \mathcal{B}_n et \mathcal{T}_{n+1} sont indépendantes. Ainsi, on a que \mathcal{B}_n et \mathcal{T}_∞ sont indépendantes. Ceci étant vrai pour tout $n \in \mathbb{N}$, on peut donc écrire que

$$\forall A \in \mathcal{T}_\infty, \forall B \in \bigcup_{n=1}^{\infty} \mathcal{B}_n, \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Or $\bigcup_{n=1}^{\infty} \mathcal{B}_n$ est stable par intersection finie. On a vu que ceci implique donc que \mathcal{T}_∞ et $\sigma(\bigcup_{n=1}^{\infty} \mathcal{B}_n)$ sont indépendantes. Or $\sigma(\bigcup_{n=1}^{\infty} \mathcal{B}_n) = \sigma(X_n : n \geq 1)$. Ainsi \mathcal{T}_∞ est indépendante d'elle-même ce qui implique que pour tout $A \in \mathcal{T}_\infty$ on a $\mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$. Ceci veut dire que $\mathbb{P}(A) \in \{0, 1\}$ et termine la preuve. \square

Remarque 4.21

On a déjà vu un résultat dans le même esprit. En effet, les deux lemmes de Borel-Cantelli stipulent que si A_1, A_2, \dots sont des événements indépendants alors $\mathbb{P}(\limsup A_n) \in \{0, 1\}$. Si on pose $X_n = \mathbf{1}_{A_n}$ pour tout $n \in \mathbb{N}$ alors $\sigma(X_n) = \sigma(A_n)$ et la loi du 0/1 implique en particulier que $\mathbb{P}(\limsup A_n) \in \{0, 1\}$ puisque $\limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \in \bigcap_{n \in \mathbb{N}} \sigma(X_m : m \geq n)$.

Exemple 4.22

Soient X_1, X_2, \dots des variables aléatoires réelles indépendantes. Alors $\mathbb{P}(\sum_{n=1}^{\infty} X_n \text{ converge}) \in \{0, 1\}$. En d'autres termes, une série formée de termes aléatoires indépendants soit converge presque sûrement soit diverge presque sûrement.

En effet, on a pour tout $m \in \mathbb{N}$ que $\{\sum_{n=1}^{\infty} X_n \text{ converge}\} = \{\sum_{n=1}^{\infty} X_{n+m} \text{ converge}\} \in \sigma(X_{m+1}, \dots)$. Ainsi $\{\sum_{n=1}^{\infty} X_n \text{ converge}\} \in \bigcap_{n \in \mathbb{N}} \sigma(X_m : m \geq n)$ et par la loi du 0/1 on a le résultat.

Exemple 4.23

Un homme ivre marche sur la droite réelle en partant de l'origine. A chaque étape, il fait un pas vers la droite avec probabilité p et un pas vers la gauche avec probabilité $1 - p$. On suppose que les pas effectués à chaque étape sont indépendants. Alors l'événement que l'homme ivre s'échappe à l'infini est soit presque sûr soit de probabilité nulle.

On modélise le problème de la façon suivante. On pose $X_0 = 0$ et pour tout $n \geq 1$, on définit des variables aléatoires indépendantes $(X_n)_{n \geq 1}$ par $\mathbb{P}(X_n = -1) = 1 - p$ et $\mathbb{P}(X_n = 1) = p$. La position de l'homme ivre après n pas est donnée par $W_n = \sum_{k=1}^n X_k$. Il s'agit de montrer que

$$\mathbb{P}(\lim_n W_n = \pm\infty) \in \{0, 1\}.$$

Comme avant, ceci découle de la loi du 0/1 puisque l'événement $\{\lim_n W_n = \pm\infty\} \in \sigma(X_m : m \geq n)$ pour tout $n \in \mathbb{N}$.

V Processus de Poisson

De nombreux phénomènes aléatoires de la physique, des télécommunications, de l'informatique sont liés à des processus d'arrivées aléatoires dans le temps. Ainsi l'étude du congestionnement d'un central téléphonique ou d'un serveur doit commencer par celle des appels et des requêtes qui se produisent à des temps aléatoires et avec des durées de service aléatoires. En physique atomique les instants de désintégration des atomes d'une matière radioactive sont encore un autre exemple. Un *processus ponctuel* est la donnée d'une suite croissante $(T_i)_{i \geq 1}$ d'instants aléatoires

$$0 < T_1 < T_2 < \dots < T_n < \dots$$

tendant vers l'infini. On suppose donc qu'il n'y a pas d'arrivée au temps 0, qu'il n'y a pas d'arrivées simultanées, et qu'il n'y a pas d'accumulation d'arrivées. Les *délais inter-arrivées* sont les différences successives

$$\tau_1 = T_1, \quad \tau_2 = T_2 - T_1, \quad \tau_n = T_n - T_{n-1}, \quad \dots$$

De manière équivalente, on peut considérer la *fonction aléatoire de comptage* des arrivées jusqu'au temps t ,

$$N(t) = \sum_{i \geq 1} \mathbf{1}_{\{T_i \leq t\}}, \quad t > 0.$$

Comme $\lim_n T_n = \infty$, $N(t)$ est fini pour tout $t > 0$.

Le processus de Poisson est caractérisé par sa fonction de comptage, à accroissements indépendants stationnaires et de loi de Poisson.

Définition 4.24 (*Processus de Poisson*)

Un processus ponctuel est appelé processus de Poisson d'intensité $\theta > 0$ si sa fonction de comptage vérifie :

- pour tous $0 < s < t$, la v.a. $N(t) - N(s)$ suit la loi de Poisson de paramètre $\theta(t - s)$,
- pour tous $k \geq 2$ et $0 < t_1 < t_2 < \dots < t_k$, les v.a. $N(t_1), N(t_2) - N(t_1), \dots, N(t_k) - N(t_{k-1})$ sont indépendantes.

On rappelle que si Z suit la loi de Poisson de paramètre θ , alors

$$\mathbb{P}(Z = n) = \frac{\theta^n}{n!} e^{-\theta}, \quad n = 0, 1, \dots, \quad \text{et } \mathbb{E}[Z] = \theta.$$

Ainsi le nombre moyen d'arrivées $\mathbb{E}(N(t) - N(s))$ dans l'intervalle $]s, t]$ est proportionnel à la durée $t - s$.

Il n'est pas clair, a priori, que cette définition fait sens : il se pourrait qu'il n'existe pas de processus ponctuel vérifiant cette définition. La proposition suivante assure l'existence d'un tel processus, et montre que sa loi est complètement caractérisée par cette définition.

Proposition 4.25

Soit $(\tau_i)_{i \geq 1}$ une suite de v.a. i.i.d. de loi exponentielle de paramètre θ . Si l'on pose $T_n := \sum_{i=1}^n \tau_i$, alors $(T_n)_{n \geq 1}$ est un processus de Poisson d'intensité θ .

Réciproquement, si $(T_n)_{n \geq 1}$ est un processus de Poisson d'intensité θ alors $(\tau_i)_{i \geq 1}$ est une suite de v.a. i.i.d. de loi exponentielle de paramètre θ .

Démonstration L'observation fondamentale est que, pour tout $n \geq 1$, les lois des vecteurs $(\tau_i; 1 \leq i \leq n)$ et $(T_i; 1 \leq i \leq n)$ se déduisent l'une de l'autre par un changement de variables linéaire (de jacobien égal à 1) : le vecteur colonne $(T_i; 1 \leq i \leq n)$ est égal à A fois le vecteur colonne $(\tau_i; 1 \leq i \leq n)$ où A est la matrice

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & 0 & 0 \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}.$$

On peut alors vérifier par la formule de changement de variable que le premier vecteur a pour densité

$$\prod_{i=1}^n \theta \exp(-\theta s_i) \mathbf{1}_{s_i > 0},$$

si et seulement si le second vecteur a pour densité

$$\theta^n \exp(-\theta t_n) \mathbf{1}_{\Delta_n}(t_1, \dots, t_n)$$

en notant Δ_n le simplexe

$$\Delta_n = \{(t_i)_{i=1}^n : 0 < t_1 < t_2 < \dots < t_n\}.$$

Nous nous contentons à présent de prouver la réciproque de l'énoncé. Il suffit de vérifier que (T_1, \dots, T_{n+1}) a pour densité $\theta^{n+1} \exp(-\theta t_{n+1}) \mathbf{1}_{\Delta_{n+1}}$. On fixe $0 < u_1 < v_1 < u_2 < v_2 < \dots < u_n < v_n < t$, et on exprime

$$\begin{aligned} \{T_i \in]u_i, v_i], i = 1, \dots, n; t < T_{n+1}\} &= \{N(u_1) = 0, N(v_1) - N(u_1) = 1, \\ &N(u_2) - N(v_1) = 0, N(v_2) - N(u_2) = 1, \\ &\dots, \\ &N(u_n) - N(v_{n-1}) = 0, N(v_n) - N(u_n) = 1, \\ &N(t) - N(v_n) = 0\} \end{aligned}$$

pour calculer (avec la convention $v_0 = 0$)

$$\begin{aligned} \mathbb{P}(T_i \in]u_i, v_i], i = 1, \dots, n; t < T_{n+1}) &= e^{-\theta(t-v_n)} \prod_{i=1}^n e^{-\theta(u_i-v_{i-1})} \theta(v_i - u_i) e^{-\theta(v_i-u_i)} \\ &= e^{-\theta t} \prod_{i=1}^n \theta(v_i - u_i) \end{aligned}$$

Comme ceci est égal à l'intégrale de la fonction positive

$$p_{n+1}(t_1, \dots, t_{n+1}) = \theta^{n+1} e^{-\theta t_{n+1}} \mathbf{1}_{\Delta_{n+1}}$$

sur le pavé $\{t_i \in]u_i, v_i], i = 1, \dots, n; t < t_{n+1}\}$, on en déduit que p_{n+1} est la densité de probabilité de $(T_i)_{i=1, \dots, n+1}$. \square

Proposition 4.26

Soit $(T_n)_{n \geq 1}$ un processus de Poisson de paramètre $\theta > 0$. Soit $n \geq 1$ et $t > 0$. Conditionnellement à l'événement $\{N(t) = n\}$, la loi de $(T_i)_{1 \leq i \leq n}$ est uniforme sur le simplexe

$$\Delta_{n,t} = \{(t_i)_{1 \leq i \leq n} : 0 < t_1 < t_2 < \dots < t_n < t\}.$$

C'est la loi de la statistique d'ordre d'un n -échantillon de loi uniforme sur $[0, t]$, c'est à dire de n v.a. i.i.d. de loi uniforme sur $[0, t]$ réordonnées par ordre croissant.

Démonstration On calcule avec les notations précédentes

$$\begin{aligned} \mathbb{P}(T_i \in]u_i, v_i], i = 1 | N(t) = n) &= \frac{\mathbb{P}(T_i \in]u_i, v_i], i = 1, N(t) = n)}{\mathbb{P}(N(t) = n)} \\ &= \frac{e^{-\theta t} \prod_{i=1}^n \theta(v_i - u_i)}{e^{-\theta t} (\theta t)^n / n!} \\ &= n! t^{-n} \text{Vol}\left(\prod_{i=1}^n]u_i, v_i]\right) \end{aligned}$$

Cela montre que la loi de $(T_i)_{1 \leq i \leq n}$ conditionnellement à $N(t) = n$, qui est portée par le simplexe $\Delta_{n,t}$, est en fait uniforme sur cet ensemble.

Enfin, nous montrons que si U_1, \dots, U_n sont n v.a. i.i.d. de loi uniforme sur $[0, t]$ et si $Y_1 < Y_2 < \dots < Y_n$ est la suite de leurs valeurs classées par ordre croissant, alors $(Y_i; 1 \leq i \leq n)$ est de loi uniforme sur $\Delta_{n,t}$. Si $A \subset \Delta_{n,t}$ est un borélien, on a

$$\{(Y_i; 1 \leq i \leq n) \in A\} = \bigcup_{\sigma \in \mathfrak{S}_n} \{(U_{\sigma(i)}; 1 \leq i \leq n) \in A\},$$

et l'union est disjointe puisque $A \subset \Delta_n$. On en déduit la première égalité suivante

$$\begin{aligned} \mathbb{P}((Y_i; 1 \leq i \leq n) \in A) &= \sum_{\sigma \in \mathfrak{S}_n} \mathbb{P}((U_{\sigma(i)}; 1 \leq i \leq n) \in A) \\ &= n! \times \mathbb{P}((U_i; 1 \leq i \leq n) \in A) \end{aligned}$$

puisque $(U_{\sigma(i)}; 1 \leq i \leq n)$ a même loi que $(U_i; 1 \leq i \leq n)$. Cette dernière quantité s'écrivant comme

$$\int_A n! \times t^{-n} \mathbf{1}_{\Delta_{n,t}}(y_1, \dots, y_n) dy_1 \dots dy_n,$$

on voit que $(Y_i; 1 \leq i \leq n)$ est uniforme sur $\Delta_{n,t}$. □

Chapitre 5

Espérance conditionnelle

I Conditionnement par rapport à un événement

On se place sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Notre but dans un premier temps est de définir l'espérance d'une variable aléatoire conditionnée à la réalisation d'un événement de probabilité positive. Commençons par rappeler la définition de la probabilité conditionnelle.

Définition 5.1 (*Probabilité conditionnelle*)

Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. On définit la probabilité conditionnelle sachant B , qu'on note $\mathbb{P}(\cdot | B)$, par

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

Il est facile de vérifier que $\mathbb{P}(\cdot | B)$ définit bien une mesure de probabilité sur (Ω, \mathcal{A}) . En particulier, le facteur $\mathbb{P}(B)$ au dénominateur assure que la masse totale de cette mesure vaut bien 1. On notera que cette mesure de probabilité ne donne de la masse qu'aux événements qui intersecte B .

Cette notion de probabilité conditionnelle fournit une définition possible pour l'espérance conditionnelle d'une variable aléatoire étagée. En effet, on peut poser

$$\mathbb{E}[\mathbf{1}_A | B] = \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}[\mathbf{1}_A \mathbf{1}_B]}{\mathbb{P}(B)}.$$

Ainsi, pour toute fonction étagée positive, on aurait la définition $\mathbb{E}[X | B] = \frac{\mathbb{E}[X \mathbf{1}_B]}{\mathbb{P}(B)}$. C'est exactement la définition qu'on va introduire.

Définition 5.2 (*Espérance conditionnelle sachant un événement*)

Soit X une variable aléatoire et $B \in \mathcal{A}$ un événement. Si $X \geq 0$ ou $X \in L^1$, on définit l'espérance conditionnelle de X sachant B par

$$\mathbb{E}[X | B] := \frac{\mathbb{E}[X \mathbf{1}_B]}{\mathbb{P}(B)}.$$

Remarque 5.3

1. L'espérance conditionnelle de X sachant B peut être vue comme l'espérance de X contre la

mesure conditionnelle sachant B :

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

Enfin, cette égalité est immédiate pour $X = \mathbf{1}_A$, donc par linéarité pour toute variable étagée positive, puis par convergence monotone pour toute variable aléatoire positive. Le cas des variables L^1 s'en suit en décomposant en partie positive et négative. Cette interprétation nous permet de transférer toutes les propriétés établies pour l'espérance à l'espérance conditionnelle sachant un événement.

2. Si B et B^c ont des probabilités strictement positives,

$$\mathbb{E}[X] = \mathbb{P}(B)\mathbb{E}[X | B] + \mathbb{P}(B^c)\mathbb{E}[X | B^c].$$

Ceci découle de la définition de l'espérance conditionnelle, de la linéarité de l'espérance et du fait que $\mathbf{1}_B + \mathbf{1}_{B^c} = 1$.

3. Evidemment, on a aussi $\mathbb{E}[X] = \mathbb{E}[X | \Omega]$.

Exemple 5.4

Soit X une variable uniforme sur $\{-2, -1, 0, 1, 2\}$ et soit B l'événement que X est positive. Alors

$$\mathbb{E}[X | B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)} = \frac{0 \cdot \frac{1}{5} + 1 \cdot \frac{1}{5} + 2 \cdot \frac{1}{5}}{\frac{3}{5}} = 1.$$

En fait, X suit la loi uniforme sur $\{-2, -1, 0, 1, 2\}$, donc une fois conditionnée à être dans B , elle suit la loi uniforme sur B .

II Conditionnement par rapport à une variable discrète

Notre but à présent est de définir l'espérance conditionnelle d'une variable aléatoire X sachant une autre variable Y . Quel sens souhaite-t-on donner à cette quantité ?

De façon vague, il s'agirait de la "meilleure" approximation de X sachant Y . Un peu plus précisément, on pourrait définir l'espérance conditionnelle de X sachant Y comme la variable aléatoire qui, sur l'événement $\{Y = y\}$, vaut $\mathbb{E}[X | \{Y = y\}]$.

On voit immédiatement que cela n'a de sens que si les événements $\{Y = y\}$ sont de probabilité positives. On va donc commencer par traiter le cas où Y est une variable discrète.

Commençons par regarder le cas simple où $Y = \mathbf{1}_B$ pour un événement B . On sait que $B = \{\mathbf{1}_B = 1\}$ et ainsi $\mathbb{E}[X | B] = \mathbb{E}[X | \{\mathbf{1}_B = 1\}]$. Mais $\mathbf{1}_B$ donne lieu à un autre événement $\{\mathbf{1}_B = 0\} = B^c$. On aimerait donc définir $\mathbb{E}[X | \mathbf{1}_B]$ de telle façon à avoir accès aux deux quantités $\mathbb{E}[X | \{\mathbf{1}_B = 1\}]$ et $\mathbb{E}[X | \{\mathbf{1}_B = 0\}]$. Pour ce faire, on va définir $\mathbb{E}[X | \mathbf{1}_B]$ comme étant une variable aléatoire qui nous donne accès à ces deux quantités. Ainsi, pour tout $\omega \in \Omega$, si $\omega \in B$ on pose

$$\mathbb{E}[X | \mathbf{1}_B](\omega) = \mathbb{E}[X | \{\mathbf{1}_B = 1\}]$$

et si $\omega \notin B$ on pose

$$\mathbb{E}[X | \mathbf{1}_B](\omega) = \mathbb{E}[X | \{\mathbf{1}_B = 0\}].$$

En d'autres termes, on a posé pour tout $\omega \in \Omega$

$$\mathbb{E}[X | \mathbf{1}_B](\omega) = \mathbf{1}_B(\omega)\mathbb{E}[X | \{\mathbf{1}_B = 1\}] + \mathbf{1}_B^c(\omega)\mathbb{E}[X | \{\mathbf{1}_B = 0\}].$$

Bibliographie

- [BP00] M. BRIANE and G. PAGÈS. *Théorie de l'intégration : Licence de mathématiques cours et exercices*. Les grands cours Vuibert. Vuibert, Paris, 7e édition revue et augmentée ed., DL 2000.