

Cours de probabilités – MA 101

L. Decreusefond

2020

Table des matières

1	Mesures et intégration	5
1.1	Introduction	5
1.2	Tribus	6
1.3	Fonctions mesurables	8
1.4	Mesures	12
1.5	Intégration	19
2	Fonctions de répartition	37
2.1	Variables aléatoires réelles	37
2.2	Construction de variables aléatoires	41
3	Moments	43
3.1	Espérance	43
3.2	Variance et autres moments	44
3.3	Inégalités	45
3.4	Fonction caractéristique	46
4	Vecteurs aléatoires	49
4.1	Changement de variables	50
4.2	Formulaire	55
5	Fonction caractéristique	57
5.1	Définition et propriétés	57
5.2	Fonctions caractéristiques de variable aléatoire usuelles	59
5.3	Caractérisation de la loi	61
5.4	Caractérisation de l'indépendance	62
5.5	Calcul de moments	62
6	Vecteurs gaussiens	67
6.1	Définition et propriétés	67
6.2	Caractérisation de l'indépendance	70
6.3	Stabilité par transformation affine	71
6.4	Somme de vecteurs gaussiens indépendants	72
6.5	La loi d'un vecteur gaussien admet-elle une densité ?	73
7	Convergences	75
7.1	Loi des grands nombres	75
7.2	Limite centrée	75

Mesures et intégration

Ce chapitre a un statut particulier. D'un côté, il introduit les concepts et les notations au cœur de ce cours. De l'autre, son contenu en tant que tel n'est pas évalué dans le contrôle final. En bref, il n'y aura pas de questions sur la théorie de la mesure mais toutes les notations et définitions sont d'un usage constant.

1.1 Introduction

Ce chapitre a pour but l'introduction des outils nécessaires à la construction de probabilités sur des ensembles plus généraux que les seuls ensembles discrets. Une probabilité sur l'univers Ω est une application $\mathbf{P}$ qui à un événement A associe une valeur $\mathbf{P}(A)$ comprise entre 0 et 1. Formellement,

$$\begin{aligned}\mathbf{P} : \mathcal{F} &\rightarrow [0, 1] \\ A &\mapsto \mathbf{P}(A)\end{aligned}$$

où $\mathcal{F}$ est le domaine de définition de $\mathbf{P}$. Lorsque l'univers Ω est au plus dénombrable, nous pouvons choisir $\mathcal{F}$ comme l'ensemble des parties. En revanche, dans des espaces plus complexes tels que $\Omega = \mathbf{R}$, sauf à renoncer à l'axiome du choix [?], il est impossible de définir $\mathbf{P}$ sur l'ensemble des parties. Il nous faut donc restreindre le domaine de définition de $\mathbf{P}$.

Les conditions qui pèsent sur $\mathcal{F}$ sont liées aux considérations suivantes :

- a) On veut pouvoir définir les probabilités de l'événement impossible (zéro) et de l'événement certain (un). Donc $\mathcal{F}$ doit contenir $\emptyset$ et Ω .
- b) Si on sait évaluer la probabilité qu'un événement A se réalise, on doit logiquement pouvoir parler de la probabilité qu'il ne se réalise pas. Autrement dit, si $\mathcal{F}$ est stable par passage au complémentaire.
- c) Si A et B sont des événements dont on sait évaluer les probabilités, on doit pouvoir donner un sens à $\mathbf{P}(A \cup B)$, donc $\mathcal{F}$ est stable pour l'union. Pour des raisons techniques qui apparaîtront clairement plus loin, nous supposons en outre que $\mathcal{F}$ est stable par union dénombrable.

Les axiomes ci-dessus sont ceux qui définissent une *tribu*. Le paragraphe 1.2 sera consacré à des rappels sur les tribus. Le paragraphe 1.4 est consacré à la définition des mesures, qui sont des applications sur $\mathcal{F}$ à valeurs positives. Les mesures de probabilité en sont un cas particulier. Le paragraphe 1.3 introduit la notion d'applications mesurables qui donnera le cadre formel nécessaire à la définition des variables aléatoires à valeurs dans des espaces plus généraux que les espaces discrets.

1.2 Tribus

Définition 1.1 (Tribu). Une famille $\mathcal{F}$ de sous-ensembles de Ω est appelée une *tribu* sur Ω si elle vérifie les propriétés suivantes :

- 1.1.i) $\Omega \in \mathcal{F}$;
- 1.1.ii) $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$;
- 1.1.iii) $\forall A_1, A_2, \dots \in \mathcal{F}, \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Autrement dit, une tribu est stable par passage au complémentaire et stable par union dénombrable.

Citons quelques exemples de tribus :

- la tribu grossière : $\mathcal{F} = \{\emptyset, \Omega\}$;
- la tribu des parties : $\mathcal{F} =$ l'ensemble des sous-ensembles de Ω , noté $\mathcal{P}(\Omega)$ ou 2^Ω ;

Un *espace mesurable* est un couple $(\Omega, \mathcal{F})$ où Ω est un ensemble et $\mathcal{F}$ est une tribu sur Ω . On parle parfois d'*espace probabilisable*.

Proposition 1.1. Toute tribu satisfait les propriétés suivantes.

- a) $\emptyset \in \mathcal{F}$;
- b) $\forall A_1, A_2, \dots \in \mathcal{F}, \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$;
- c) $\forall A, B \in \mathcal{F}, A \cup B \in \mathcal{F}$ et $A \cap B \in \mathcal{F}$.

Démonstration. — **a)** $\Omega \in \mathcal{F}$ donc $\Omega^c = \emptyset \in \mathcal{F}$ par l'axiome ii).

- **b)** Les complémentaires $A_1^c, A_2^c, \dots$ sont dans $\mathcal{F}$ par l'axiome ii). Donc $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$ par l'axiome 1.1.iii). En invoquant à nouveau l'axiome 1.1.ii), le complémentaire de cet ensemble est également dans $\mathcal{F}$ par l'axiome 1.1.ii). Or d'après les lois de De Morgan, le complémentaire coïncide avec $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
- **c)** Il suffit de poser $A_1 = A, A_2 = B$ et $A_i = \emptyset$ pour tout $i \geq 3$. Les deux résultats découlent de l'axiome 1.1.iii) et de **b)** respectivement.

□

En appliquant les axiomes définissant une tribu, il n'est pas difficile de montrer le résultat suivant.

Lemme 1.2. L'intersection de deux tribus est une tribu.

On peut alors envisager la notion fondamentale de *tribu engendrée*.

Théorème 1.3 (Tribu engendrée). Soit $\mathcal{C}$ une collection d'ensembles sur Ω . L'intersection de toutes les tribus sur Ω contenant $\mathcal{C}$ est une tribu sur Ω . On la note $\sigma(\mathcal{C})$ et on l'appelle la tribu engendrée par $\mathcal{C}$ sur Ω .

Démonstration. Soit S l'ensemble des tribus contenant $\mathcal{C}$. S est non-vide puisqu'il contient la tribu des parties. On vérifie que $\sigma(\mathcal{C}) = \bigcap_{\tau \in S} \tau$ vérifie les trois axiomes d'une tribu.

- *a)* Pour tout $\tau \in S$, $\emptyset \in \tau$ puisque τ est une tribu. Donc $\emptyset \in \bigcap_{\tau \in S} \tau$.
- *b)* Soit $A \in \sigma(\mathcal{C})$. Pour tout $\tau \in S$, on a $A \in \tau$ par définition de $\sigma(\mathcal{C})$. Donc $A^c \in \tau$ car une tribu est stable par passage au complémentaire. Ainsi $A^c \in \bigcap_{\tau \in S} \tau$.
- *c)* Soient $A_1, A_2, \dots \in \sigma(\mathcal{C})$. Pour tout $\tau \in S$, on a $A_1, A_2, \dots \in \tau$. Donc, l'union des A_i est dans τ quelque soit $\tau \in S$. Finalement, $\bigcup_{i=1}^{\infty} A_i \in \bigcap_{\tau \in S} \tau$.

□

EXEMPLE 1.— Soit $A \subset \Omega$, calculons $\sigma(A)$. Nécessairement, $\emptyset$, A et Ω appartiennent à $\sigma(A)$. D'après l'axiome *b)*, $A^c \in \sigma(A)$. En envisageant alors tous les cas possibles, on se convainc aisément que $\sigma(A) = \{\emptyset, \Omega, A, A^c\}$.

EXEMPLE 2.— Rappelons nous qu'une tribu est la famille des ensembles que l'on peut « mesurer ». Si $\Omega = \mathbf{N}$, la moindre des choses est de pouvoir mesurer les singletons. Dans ce cas, comme toute partie de Ω s'écrit la réunion des singletons qui la composent et que cette réunion contient un nombre au plus dénombrable d'éléments, il ressort de l'axiome *1.1.iii)* que toutes les parties de Ω appartiennent à la tribu engendrée par les singletons.

EXEMPLE 3.— Le troisième exemple est le plus fondamental pour toutes les applications qui vont suivre. Il va nous permettre de préciser enfin ce que sont les parties de $\mathbf{R}$ (ou de $\mathbf{R}^d$) susceptible d'être mesurée sans renoncer à l'axiome du choix.

Définition 1.2 (Tribu borélienne). La *tribu de Borel sur $\mathbf{R}^d$* , notée $\mathcal{B}(\mathbf{R}^d)$, est la tribu engendrée par les « pavés » :

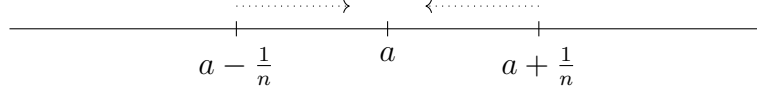
$$\mathcal{B}(\mathbf{R}^d) := \sigma \left(\left\{ \prod_{i=1}^d]a_i, b_i[: \forall i = 1, \dots, d, (a_i, b_i) \in \mathbf{R}^2, a_i < b_i \right\} \right).$$

En particulier sur $\mathbf{R}$, $\mathcal{B}(\mathbf{R})$ est la plus petite tribu qui contient les intervalles de la forme $]a, b[$.

Un élément de la tribu de Borel est appelé un *borélien*.

Théorème 1.4 (Eléments caractéristiques de $\mathcal{B}(\mathbf{R})$). Les ensembles suivants sont dans $\mathcal{B}(\mathbf{R})$: le singleton $\{a\}$, les intervalles de la forme $[a, b]$, $] - \infty, b]$, $[a, +\infty[$, l'ensemble des rationnels, l'ensemble des irrationnels.

Démonstration. Le singleton $\{a\}$ est un borélien car il s'écrit comme une intersection dénombrable de boréliens $\{a\} = \bigcap_{n \geq 1}]a - \frac{1}{n}, a + \frac{1}{n}[$.



L'intervalle $[a, b]$ est un borélien car il s'écrit comme l'union de trois boréliens $[a, b] = \{a\} \cup]a, b[\cup \{b\}$. L'intervalle $] - \infty, b]$ est un borélien car il s'écrit comme une union dénombrable de boréliens $] - \infty, b] = \bigcup_{n \geq 1}]b - n, b]$. La preuve est similaire pour $[a, +\infty[$. L'ensemble des rationnels $\mathbf{Q}$ s'écrit comme l'union de ses singletons $\mathbf{Q} = \bigcup_{x \in \mathbf{Q}} \{x\}$. Puisque $\mathbf{Q}$ est dénombrable et que les singletons sont des boréliens, $\mathbf{Q}$ est un borélien en vertu de 1.1.iii). L'ensemble des irrationnels est le complémentaire de $\mathbf{Q}$, c'est donc aussi un borélien puisque $\mathcal{B}(\mathbf{R})$ est stable par passage au complémentaire. $\square$

Proposition 1.5. $\mathcal{B}(\mathbf{R})$ est la tribu engendrée par les intervalles de la forme $] - \infty, b]$. Plus généralement, $\mathcal{B}(\mathbf{R}^d)$ est engendrée par les ensembles de la forme $\prod_{i=1}^d] - \infty, b_i]$.

Démonstration. D'après la propriété précédente, la classe $\mathcal{C}$ des ensembles $] - \infty, b]$ sont dans $\mathcal{B}(\mathbf{R})$, donc $\sigma(\mathcal{C}) \subset \mathcal{B}(\mathbf{R})$ car $\sigma(\mathcal{C})$ est la plus petite tribu contenant $\mathcal{C}$. Inversement, tout intervalle $]a, b[$ s'écrit $] - \infty, a]^c \cap (\bigcup_n] - \infty, b - \frac{1}{n}]$, donc $]a, b[$ est inclus dans $\sigma(\mathcal{C})$. Donc $\mathcal{B}(\mathbf{R}) \subset \sigma(\mathcal{C})$. $\square$

REMARQUE.— Dans ce cours, nous utiliserons assez fréquemment la tribu de Borel $\mathcal{B}(\overline{\mathbf{R}})$ sur $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$. On admettra que les éléments de $\mathcal{B}(\overline{\mathbf{R}})$ sont exactement les ensembles de la forme H , $H \cup \{+\infty\}$, $H \cup \{-\infty\}$, $H \cup \{-\infty, +\infty\}$ où $H \in \mathcal{B}(\mathbf{R})$. On en déduit que $\mathcal{B}(\overline{\mathbf{R}})$ est la tribu engendrée par les intervalles de la forme $[-\infty, b]$ pour b décrivant $\overline{\mathbf{R}}$.

REMARQUE.— A l'inverse, pour n'importe quel ensemble $E \subset \mathbf{R}^d$, la tribu borélienne sur E est la *trace* de la tribu borélienne de $\mathbf{R}^d$ sur E : en termes moins sophistiqués, cela signifie que

$$\mathcal{B}(E) = \mathcal{B}(\mathbf{R}^d) \cap E.$$

1.3 Fonctions mesurables

On rappelle la notion d'*image réciproque* d'un ensemble par une application. On se donne deux ensembles Ω et E et une application X de Ω dans E . Pour une partie de E , notée

H , l'image réciproque de H par X , notée $X^{-1}(H)$ est l'ensemble des éléments de Ω dont l'image par X est dans H :

$$X^{-1}(H) = \{\omega \in \Omega, X(\omega) \in H\}.$$

Cette notion est définie que X soit inversible ou pas.

EXEMPLE 4.— Considérons l'application qui à un entier n associe la partie entière de sa moitié : $f(n) = [n/2]$. Dans ces conditions, pour tout entier k , $f^{-1}(\{k\}) = \{2k, 2k+1\}$.

La notion d'image réciproque (au contraire de la notion d'image directe) se marie bien avec les intersections et les unions d'ensemble.

Proposition 1.6. Pour toute famille dénombrable d'ensemble $(A_n, n \geq 1)$

$$\begin{aligned} X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \bigcup_{n=1}^{\infty} X^{-1}(A_n), \\ X^{-1}\left(\bigcap_{n=1}^{\infty} A_n\right) &= \bigcap_{n=1}^{\infty} X^{-1}(A_n). \end{aligned}$$

Nantis de ces propriétés de base, nous sommes prêts pour donner la définition des fonctions mesurables.

Définition 1.3. Une application $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ est dite *mesurable* si :

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

En langage probabiliste, une application mesurable s'appelle une *variable aléatoire*.

EXEMPLE 5.— Si Ω est munie de la tribu grossière, c'est-à-dire que $\mathcal{F} = \{\emptyset, \Omega\}$ alors une fonction mesurable est nécessairement constante (dès que les singletons sont des mesurables de E).

En effet, soit $\omega \in \Omega$ et considérons $A = X^{-1}(\{X(\omega)\})$, l'ensemble des éléments de Ω qui ont la même image par X que le ω considéré. Puisque A est mesurable, A vaut $\emptyset$ ou Ω . Comme $\omega \in A$, A ne peut être vide donc $A = \Omega$. Autrement dit, X est constante.

EXEMPLE 6.— De la même manière, on montre que si $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ alors les fonctions mesurables ne prennent que deux valeurs : elles sont de la forme

$$X(\omega) = \alpha \mathbf{1}_A(\omega) + \beta \mathbf{1}_{A^c}(\omega),$$

avec α et β , deux éléments de E .

REMARQUE.— Si l'espace Ω est dénombrable (ou fini), on a vu que la seule tribu sensée était la tribu des parties de Ω . En d'autres termes, toute fonction définie sur un tel ensemble est nécessairement mesurable puisque l'image inverse d'un ensemble quelconque de E est par définition une partie de Ω .

La propriété de composition des fonctions mesurables bien que totalement basique est d'un usage constant.

Théorème 1.7. Soit $(E', \mathcal{E}')$ un espace mesurable. Soit $X : \Omega \rightarrow E$ et $f : E \rightarrow E'$ deux applications mesurables. La composée $f \circ X : \Omega \rightarrow E'$ est une application mesurable.

Démonstration. Soit $H' \in \mathcal{E}'$. L'image réciproque de H' par $f \circ X$ est égale à $X^{-1}(f^{-1}(H'))$. Comme f est mesurable, $f^{-1}(H') \in \mathcal{E}$. Comme X est mesurable, l'image réciproque d'un élément de $\mathcal{E}$ est dans $\mathcal{F}$, d'où $X^{-1}(f^{-1}(H')) \in \mathcal{F}$. $\square$

On admettra les deux résultats suivants. Le premier stipule que pour vérifier la mesurabilité d'une application X , il suffit de vérifier la propriété « $X^{-1}(H) \in \mathcal{F}$ » non pas pour *tout* $H \in \mathcal{E}$, mais seulement pour H dans une classe plus réduite, qui engendre la tribu d'arrivée.

Théorème 1.8. Supposons que $\mathcal{E} = \sigma(\mathcal{C})$ pour une certaine classe $\mathcal{C}$. Soit une fonction $X : \Omega \rightarrow E$ telle que pour tout $C \in \mathcal{C}$, $X^{-1}(C) \in \mathcal{F}$. Alors X est mesurable.

Le corollaire d'usage pratique est le suivant.

Corollaire 1.9. Soit $d \in \mathbf{N}^*$ et soient $X_1, \dots, X_d$ une collection de fonctions de Ω dans $\mathbf{R}$. On définit $X : \Omega \rightarrow \mathbf{R}^d$ par $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$. Les propositions suivantes sont équivalentes :

- a) $X_1, \dots, X_d$ sont des fonctions mesurables ;
- b) X est une fonction mesurable.

Démonstration. a) $\Rightarrow$ b). Puisque $\mathcal{B}(\mathbf{R}^d)$ est la tribu engendrée par les pavés, il suffit de vérifier que $X^{-1}(H) \in \mathcal{F}$ pour $H = \prod_{k=1}^d H_k$, où H_k est un intervalle réel. Or $X^{-1}(\prod_k H_k) = \bigcap_k X_k^{-1}(H_k)$ est bien un élément de $\mathcal{F}$ comme intersection d'éléments de $\mathcal{F}$.

b) $\Rightarrow$ a) Donnons la preuve pour X_1 . Pour tout intervalle $]a, b[$, l'ensemble $X_1^{-1}(]a, b[)$ est l'image réciproque par X de $]a, b[\times \mathbf{R} \times \dots \times \mathbf{R}$ qui appartient à $\mathcal{B}(\mathbf{R}^d)$. Donc $X_1^{-1}(]a, b[) \in \mathcal{F}$, ce qui montre que X_1 est mesurable. $\square$

Le deuxième qui nous simplifiera considérablement la vie fait le lien entre fonctions continues et fonctions mesurables.

Théorème 1.10. Une application $f : \mathbf{R}^d \rightarrow \mathbf{R}^n$ continue est mesurable. On dit alors qu'elle est *borélienne*.

REMARQUE.— On prendra soin de noter que la réciproque est fausse : il existe des fonctions mesurables qui ne sont pas continues. Par exemple, $\mathbf{1}_{\mathbf{Q}}$ la fonction indicatrice de l'ensemble des rationnels puisque $\mathbf{Q}$ est mesurable (union dénombrable de singletons) mais manifestement pas continue.

A partir de là, on peut démontrer les propriétés élémentaires des fonctions mesurables.

Corollaire 1.11. Si X, Y sont des fonctions mesurable sur $\mathbf{R}$, alors $X + Y, XY, X \vee Y := \max(X, Y), X \wedge Y := \min(X, Y)$ sont des fonctions mesurables.

Démonstration. La fonction

$$\begin{aligned} \mathbf{R} \times \mathbf{R} &\longrightarrow \mathbf{R} \\ (x, y) &\longmapsto x + y \end{aligned}$$

est continue donc borélienne de $\mathbf{R}^2$ dans $\mathbf{R}$ en vertu du théorème 1.10. Le résultat se déduit alors du théorème 1.7 par composition de fonctions mesurables. En utilisant les formules

$$x \vee y = \frac{x + y + |x - y|}{2} \text{ et } x \wedge y = \frac{x + y - |x - y|}{2},$$

on obtient la mesurabilité de $x \wedge y$ et $x \vee y$. □

Jusqu'à présent, toutes les propriétés des fonctions mesurables existent à l'identique pour les fonctions continues. Ce qui fait la richesse du concept de mesurabilité tient aux théorèmes limites. En effet, rien ne garantit qu'une limite simple de fonctions continues soit une fonction continue alors que la limite d'une suite de fonctions mesurables est mesurable !

Nous allons maintenant montrer que si $(X_n)_n$ est une suite de fonctions mesurables, alors le sup et l'inf de la suite sont aussi des variables aléatoires. Bien sûr, le sup d'une suite réelle n'est pas forcément fini. Pour donner un sens à $\sup_n X_n$ et $\inf_n X_n$ en tant que variables aléatoires, nous devons dorénavant nous placer sur $\overline{\mathbf{R}}$.

Soit $(X_n, n \in \mathbf{N})$ une suite de variables aléatoires sur $\overline{\mathbf{R}}$. On désigne par respectivement par $\sup_n X_n, \inf_n X_n, \lim_n X_n$ les fonctions définies sur $\overline{\mathbf{R}}$ par $\omega \mapsto \sup_n X_n(\omega), \omega \mapsto \inf_n X_n(\omega)$ et, lorsqu'une telle fonction existe, $\omega \mapsto \lim_n X_n(\omega)$.

Théorème 1.12. Soit $(X_n, n \in \mathbf{N})$ une suite de fonctions mesurables sur $\overline{\mathbf{R}}$.

- 1.12.a) $\sup_n X_n, \inf_n X_n$ sont des fonctions mesurables sur $\overline{\mathbf{R}}$;
- 1.12.b) Les fonctions $\liminf_n X_n$ et $\limsup_n X_n$ sont mesurables ;
- 1.12.c) Si $\lim_n X_n$ existe, c'est une fonction mesurable sur $\overline{\mathbf{R}}$.

Démonstration.

- 1.12.a)** Posons $X := \sup_n X_n$. $\mathcal{B}(\overline{\mathbf{R}})$ est la tribu engendrée par les intervalles $[-\infty, b]$ où $b \in \overline{\mathbf{R}}$. On laisse au lecteur le soin de vérifier que $X^{-1}([-\infty, b]) = \bigcap_n X_n^{-1}([-\infty, b])$. Ainsi, $X^{-1}([-\infty, b])$ est dans $\mathcal{F}$ comme intersection dénombrable d'éléments de $\mathcal{F}$. D'après le critère de mesurabilité du paragraphe précédent, X est mesurable. Pour montrer que $\inf_n X_n$ est mesurable, il suffit d'écrire $\inf_n X_n = -\sup_n (-X_n)$.
- 1.12.b)** D'après le point précédent, les fonctions $Y_k = \inf_{n \geq k} X_n$ et $Z_k = \sup_{n \geq k} X_n$ sont mesurables. Toujours d'après ce point, les variables $\sup_k Y_k$ et $\inf_k Z_k$ sont aussi mesurables. Or par définition, elles sont respectivement égales à $\liminf_n X_n$ et $\limsup_n X_n$.
- 1.12.c)** Si X_n converge, alors $\lim_n X_n = \liminf_n X_n = \limsup_n X_n$ donc d'après le point précédent, c'est une variable aléatoire.

□

1.4 Mesures

1.4.1 Définition et exemples

Définition 1.4. Une mesure sur $(E, \mathcal{E})$ est une fonction d'ensemble μ telle que

1.4.i) $\mu : \mathcal{E} \rightarrow [0, +\infty]$;

1.4.ii) $\mu(\emptyset) = 0$;

1.4.iii) Pour toute famille $(A_n, n \in \mathbf{N})$ d'événements deux à deux disjoints,

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

On prêtera attention à l'intervalle fermé à droite dans l'axiome **1.4.i)** : la mesure d'un ensemble A est une quantité positive, possiblement infinie.

Lorsque $\mu(E) < \infty$, on dit que la mesure est *finie*. Si en outre $\mu(E) = 1$, la mesure μ est une *mesure de probabilité*.

Dans ce dernier cas, on utilise alors les notations probabilistes.

- $(E, \mathcal{E})$ devient $(\Omega, \mathcal{A})$.
- Les fonctions mesurables notées f ou g deviennent des variables aléatoires notées par des capitales X, Y , etc.
- La variable $x \in E$ devient $\omega \in \Omega$.
- Les mesures μ, ν deviennent des probabilités notées $\mathbf{P}, \mathbf{Q}$.

Pour autant les concepts restent rigoureusement les mêmes !

Un triplet $(E, \mathcal{E}, \mu)$ où μ est une mesure sur $(E, \mathcal{E})$, est appelé un *espace mesuré*. C'est un espace de probabilité lorsque μ est une mesure de probabilité.

L'axiome *1.4.i*) traduit le fait que l'événement impossible a une probabilité nulle d'être réalisé, alors que l'événement certain est réalisé avec une probabilité égale à un. L'axiome *1.4.ii*) est connu sous le nom de propriété de σ -additivité.

EXEMPLE 7 (Mesures de Dirac).— Soit $(E, \mathcal{E})$ un espace mesurable quelconque et soit $a \in E$. La mesure de Dirac au point a est l'application δ_a définie sur $\mathcal{E}$ par $\delta_a(A) = \mathbf{1}_A(a)$ c'est à dire :

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

EXEMPLE 8 (Mesure de comptage et mesures discrète).— Soit I un ensemble au plus dénombrable et $(a_i)_{i \in I}$ une collection de points de E . La fonction définie pour tout $A \in \mathcal{E}$ par :

$$\mu(A) := \sum_{i \in I} \delta_{a_i}(A)$$

définit une mesure sur $\mathcal{E}$, appelée la mesure de comptage de $(a_i)_{i \in I}$. La quantité $\mu(A)$ est le nombre de points a_i contenus dans l'ensemble A : $\mu(A) = \text{cardinal}\{i : a_i \in A\}$.

Plus généralement, si $(\alpha_i)_{i \in I}$ est une suite de coefficients réels positifs, alors :

$$\mu(A) := \sum_{i \in I} \alpha_i \delta_{a_i}(A)$$

définit une mesure sur $\mathcal{E}$. La quantité $\mu(A)$ est égale à la somme des α_i pour tous les i tels que $a_i \in A$. Une telle mesure est appelée une mesure *discrète*.

Une conséquence immédiate de l'axiome *1.1.iii*) est la très utile formule des probabilités totales.

Théorème 1.13 (Formule des probabilités totales). Pour toute partition $(B_n, n \geq 1)$ constituée d'ensembles mesurables de E , pour tout mesurable A ,

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap B_n).$$

Démonstration. Les ensembles $(A \cap B_n, n \geq 1)$ forment une partition de A , le résultat découle de l'axiome 1.1.iii). $\square$

Les propriétés suivantes sont d'un usage constant.

Théorème 1.14. Soient $A, B, (A_n, n \in \mathbf{N}^*)$ des éléments de $\mathcal{E}$.

1.14.a) Si $\mu(A \cap B) < \infty$, alors $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

1.14.b) Si $A \subset B$, alors $\mu(A) \leq \mu(B)$.

1.14.c) Si $A_n \uparrow A$, alors $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.

1.14.d) Si $A_n \downarrow A$ et si $\mu(A_1) < \infty$, alors $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$.

1.14.e) Pour une famille quelconque $(A_n, n \in \mathbf{N}^*)$ dans $\mathcal{E}$, on a la *borne de l'union* :

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

Démonstration.

1.14.a) On remarque que

$$A = A \setminus (A \cap B) \cup (A \cap B).$$

Comme ces deux ensembles sont disjoints, d'après l'axiome 1.1.iii),

$$\mu(A) = \mu(A \setminus (A \cap B)) + \mu(A \cap B).$$

De même,

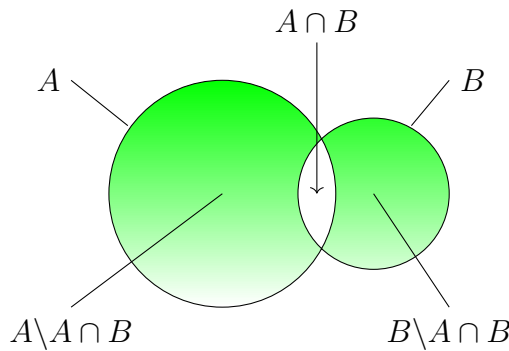
$$\mu(B) = \mu(B \setminus (A \cap B)) + \mu(A \cap B).$$

Enfin,

$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B).$$

Par conséquent,

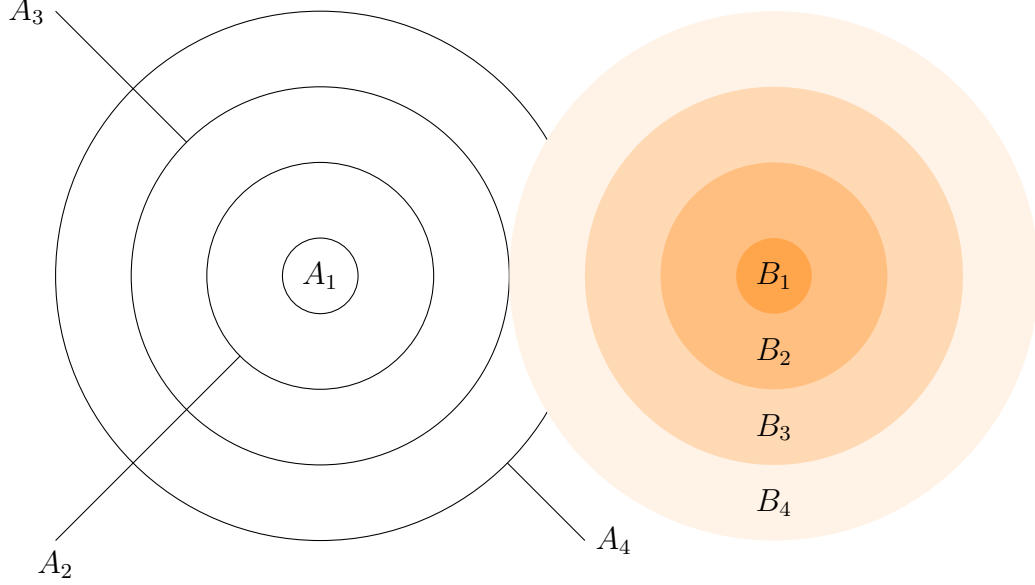
$$\begin{aligned} \mu(A \cup B) &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A) - \mu(A \cap B) + \mu(B) - \mu(A \cap B) + \mu(A \cap B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B). \end{aligned}$$



1.14.b) Si $A \subset B$, $B = A \cup (B \setminus A)$, donc $\mu(B) = \mu(A) + \mu(B \setminus A)$. Comme μ est à valeurs positives, on en déduit que $\mu(B \setminus A) \geq 0$ donc $\mu(B) \geq \mu(A)$.

1.14.c) Comme la famille des A_n est croissante, on peut considérer la famille B_n définie par

$$B_1 = A_1, \quad B_n = A_n \setminus A_{n-1}.$$



On de façon évidente que

$$\bigcup_{k=1}^n B_k = A_n \text{ et } \bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

Comme les B_n sont disjoints,

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(B_k) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

1.14.d) Comme tout se passe dans l'ensemble A_1 , on ne change rien en supposant que $E = A_1$ et $\mu(E) < \infty$. La famille des A_n^c est croissante donc on peut appliquer le résultat précédent **1.14.c)**,

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

Or,

$$\bigcup_{n=1}^{\infty} A_n^c = \left(\bigcap_{n=1}^{\infty} A_n\right)^c$$

et d'après **1.14.a)**, on obtient

$$\mu(E) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(E) - \mu(A_n).$$

En simplifiant par $\mu(E)$, on obtient le résultat voulu.

1.14.e) Soit la propriété au rang k

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

D'après 1.14.a), elle est vraie pour $k = 2$. Par récurrence, en utilisant 1.14.a), elle est vraie pour tout k . Maintenant, la famille $B_k = \bigcup_{n=1}^k A_n$ est croissante donc d'après 1.14.c)

$$\begin{aligned} \mu \left(\bigcup_{n=1}^{\infty} A_n \right) &= \mu \left(\bigcup_{k=1}^{\infty} B_k \right) = \lim_{k \rightarrow \infty} \mu(B_k) = \lim_{k \rightarrow \infty} \mu \left(\bigcup_{n=1}^k A_n \right) \\ &\leq \lim_{k \rightarrow \infty} \sum_{n=1}^K \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

Le résultat s'en suit. □

Le lemme qui suit est fondamental dans nombre d'applications.

Lemme 1.15 (Lemme de Borel-Cantelli). Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et $(A_n, n \geq 1)$ une famille d'éléments de $\mathcal{E}$.

$$\sum_{n \geq 1} \mathbf{P}(A_n) < +\infty \implies \mathbf{P}(\limsup_n A_n) = 0, \quad (1.1)$$

où la limite supérieure d'ensembles est définie par

$$\limsup_n A_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n.$$

Démonstration. Puisque la limite supérieure d'ensembles est une famille monotone d'ensembles,

$$\mathbf{P}(\limsup_n A_n) = \mathbf{P} \left(\bigcap_{k \geq 1} \bigcup_{n \geq k} A_n \right) = \lim_k \uparrow \mathbf{P} \left(\bigcup_{n \geq k} A_n \right) \leq \lim_k \sum_{n \geq k} \mathbf{P}(A_n).$$

La limite à droite est nulle comme limite du reste d'une série convergente $\sum_n \mathbf{P}(A_n) < +\infty$. □

1.4.2 Caractérisation

Dès que Ω n'est pas dénombrable, il est impossible de décrire une mesure en donnant sa valeur pour tous les ensembles mesurables. Afin d'éviter des pathologies désagréables, on se restreindra aux mesures *localement finies* :

Définition 1.5. Une mesure μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est *localement finie* si pour tout $x \in \mathbf{R}^d$, il existe un pavé borné $I \subset \mathcal{I}_d$ contenant x tel que $\mu(I) < \infty$.

Définition 1.6 (Pavés de $\mathbf{R}^d$). Les pavés de $\mathbf{R}^d$ sont les ensembles de la forme

$$]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_d, b_d[,$$

avec a_i, b_i des réels quelconques. On note $\mathcal{I}_d$ cet ensemble.

Nous sommes alors en mesure de réduire l'ensemble des parties sur lesquelles l'on doit vérifier l'égalité de deux mesures pour certifier qu'elles sont égales.

Théorème 1.16 (Kolmogorov). Deux mesures localement finies sur $\mathbf{R}^d$ qui coïncident sur $\mathcal{I}_d$ sont égales :

$$\mu(A) = \nu(A), \forall A \in \mathcal{I}_d \implies \mu(A) = \nu(A), \forall A \in \mathcal{B}(\mathbf{R}^d).$$

On montre facilement que

$$\sigma(\mathcal{I}_d) = \sigma([-\infty, a_1] \times \dots \times [-\infty, a_d]),$$

ce qui implique le théorème suivant :

Corollaire 1.17. Pour identifier une probabilité sur $\mathbf{R}^d$, il faut et il suffit que l'on connaisse $\mathbf{P}([-\infty, a_1] \times \dots \times [-\infty, a_d] - \infty, x])$ pour tout d -uplet $(a_1, \dots, a_d)$.

1.4.3 Construction

Un autre théorème fondamental de la théorie de la mesure est le suivant :

Théorème 1.18 (Kolmogorov). Il existe une unique mesure, notée λ , sur $\mathbf{R}^d$ muni de la tribu des boréliens, qui coïncide avec la mesure de longueur/surface/volume sur les pavés, c'est-à-dire telle que

$$\lambda([a_1, b_1[\times \dots \times]a_d, b_d]) = (b_1 - a_1) \dots (b_d - a_d).$$

Cette mesure s'appelle la *mesure de Lebesgue*.

Corollaire 1.19. La mesure de Lebesgue est invariante par translation : c'est-à-dire que pour tout $x \in \mathbf{R}^d$, pour tout borélien A ,

$$\lambda(A + x) = \lambda(A),$$

où

$$A + x = \{y + x, y \in A\}.$$

Démonstration. Pour $x \in \mathbf{R}^d$ fixé, pour un pavé $A =]a_1, b_1[\times \dots \times]a_d, b_d[$, son translaté est donné par

$$A + x =]a_1 + x_1, b_1 + x_1[\times \dots \times]a_d + x_d, b_d + x_d[$$

donc

$$\lambda(A + x) = \lambda(A).$$

Considérons la mesure λ_x définie par $\lambda_x(A) = \lambda(A + x)$. La mesure λ_x et la mesure de Lebesgue coïncident sur les pavés donc d'après l'unicité résultant du théorème 1.18, elles sont égales d'où le résultat. $\square$

Cet exemple de transformation de mesure est un cas particulier du concept très important de *mesure image*. Étant donné un ensemble mesuré $(E, \mathcal{E}, \mathbf{P})$ et d'une variable aléatoire X à valeurs dans $(F, \mathcal{F})$, comment construire une probabilité sur $\mathcal{F}$? La réponse somme toute assez naturelle, dans le sens où il n'y en a pas tellement d'autre possible, est donnée par la définition suivante.

Définition 1.7 (Mesure image). Soit un ensemble mesuré $(E, \mathcal{E}, \mu)$ et une fonction mesurable f à valeurs dans $(F, \mathcal{F})$. La mesure image de μ par f notée μ_f (ou $f^*\mu$ ou $f^{-1}\mu$) est définie par :

$$\forall A \in \mathcal{F}, \mu_f(A) = \mu(f^{-1}(A)),$$

où, on le rappelle, $f^{-1}(A)$ est l'image réciproque de A par f , c'est-à-dire $f^{-1}(A) = \{x \in E, f(x) \in A\}$. En probabilités, comme on omet systématiquement les x , on privilégiera donc la notation $(f^{-1}(A)) = (f \in A)$, d'où au final,

$$\forall A \in \mathcal{F}, \mu_f(A) = \mu(f \in A).$$

EXEMPLE 9.— Si l'on revient à l'exemple de mesure de Lebesgue « translatée », il faut considérer l'application

$$\begin{aligned} f_x : \mathbf{R}^d &\longrightarrow \mathbf{R}^d \\ y &\longmapsto y - x. \end{aligned}$$

En effet,

$$f_x^{-1}(A) = (y \in \mathbf{R}^d, f(y) \in A) = (y \in \mathbf{R}^d, y - x \in A) = (y \in \mathbf{R}^d, y \in A + x) = A + x.$$

La mesure λ_x introduite ci-dessus est donc la mesure image de la mesure de Lebesgue λ par l'application f_x qui translate tous les éléments de $\mathbf{R}^d$ d'une quantité x fixe. Le corollaire 1.19 implique donc que cette mesure est la même que la mesure de Lebesgue. Ceci traduit le fait que du point de vue de la mesure de Lebesgue, le point 0 ne joue pas de rôle particulier puisque la mesure est uniforme dans tout l'espace.

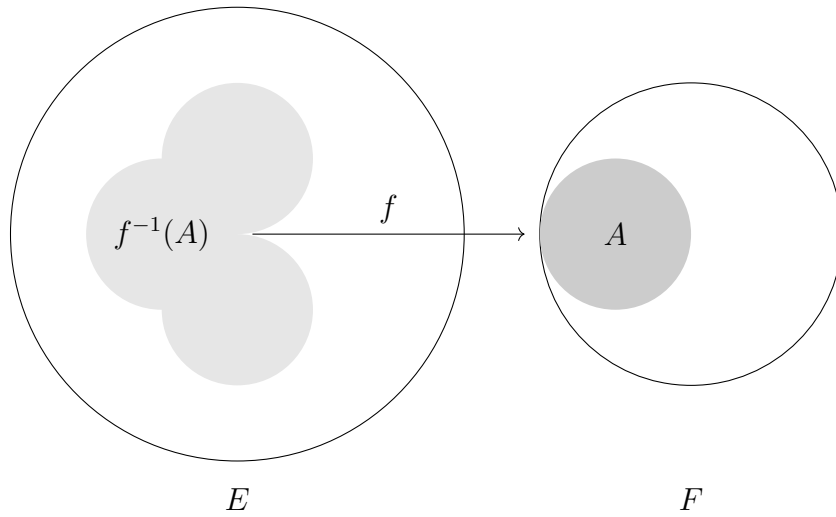


FIGURE 1.1 – Principe de construction d’une mesure image.

Si on dispose de deux espaces mesurés $(E_1, \mathcal{E}_1, \mu)$ et $(E_2, \mathcal{E}_2, \nu)$, on veut construire une mesure sur le produit cartésien $E_1 \times E_2$. La première difficulté à surmonter est la définition de la tribu sur $\mathcal{E}_1 \times \mathcal{E}_2$. Les éléments de $\mathcal{E}_1 \times \mathcal{E}_2$ sont les produits cartésiens d’un élément de $\mathcal{E}_1$ et d’un élément de $\mathcal{E}_2$. Comme la réunion de deux rectangles n’est pas un rectangle, $\mathcal{E}_1 \times \mathcal{E}_2$ n’est pas une tribu. Qu’à cela ne tienne, on note $\mathcal{E}_1 \otimes \mathcal{E}_2$ la plus petite tribu qui contient $\mathcal{E}_1 \times \mathcal{E}_2$ et le tour est joué ! Le même procédé de preuve que celui qui est utilisé pour le théorème 1.18 permet de montrer le théorème suivant.

Théorème 1.20. Il existe une unique mesure, notée $\mu \otimes \nu$, dite mesure produit de μ et ν , sur $(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2)$ qui soit telle que

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B), \text{ pour tout } A \in \mathcal{E}_1, B \in \mathcal{E}_2. \quad (1.2)$$

1.5 Intégration

1.5.1 Construction de l’intégrale

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

Définition 1.8. Une fonction f mesurable de $(E, \mathcal{E})$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ est dite étagée si elle prend un nombre fini de valeurs.

Une telle fonction f est alors de la forme

$$f(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x) \quad (1.3)$$

où $(\alpha_i, i = 1, \dots, n)$ sont les valeurs distinctes prises par la fonction f et où pour tout i , $A_i = f^{-1}(\{\alpha_i\})$.

REMARQUE.— Les fonctions en *escalier* sont des cas particuliers des fonctions étagées où les A_i sont des intervalles.

Définition 1.9. La μ -intégrale d'une fonction étagée positive f est la valeur (éventuellement égale à $+\infty$) :

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

avec la convention $0 \cdot \infty = 0$, où les valeurs $(\alpha_i, i = 1, \dots, n)$ et les ensembles $(A_i, i = 1, \dots, n)$ sont donnés par (1.3).

La convention $0 \cdot \infty = 0$ est indispensable pour l'éventualité où, pour un certain i , $\alpha_i = 0$ et $\mu(A_i) = \infty$. Grâce à cette convention, l'intégrale de la fonction nulle est égale à zéro.

Par définition de l'intégrale, on a la conséquence immédiate que pour tout $A \in \mathcal{E}$,

$$\int \mathbf{1}_A \, d\mu = \mu(A).$$

En ce sens, la notion d'intégrale généralise la notion de mesure.

Lemme 1.21. Soient $a, b \geq 0$ et f, g deux fonctions étagées positives. Alors

$$\int (af + bg) \, d\mu = a \int f \, d\mu + b \int g \, d\mu.$$

Démonstration. Pour $f = \mathbf{1}_A$ et $g = \mathbf{1}_B$,

$$af + bg = a\mathbf{1}_{A \setminus A \cap B} + b\mathbf{1}_{B \setminus A \cap B} + (a + b)\mathbf{1}_{A \cap B}.$$

Par conséquent,

$$\begin{aligned} \int (af + bg) \, d\mu &= a\mu(A \setminus A \cap B) + b\mu(B \setminus A \cap B) + (a + b)\mu(A \cap B) \\ &= a\mu(A) + b\mu(B) = a \int f \, d\mu + b \int g \, d\mu. \end{aligned}$$

Le résultat général s'en déduit par récurrence double sur le nombre de valeurs de f et g . $\square$

On définit maintenant l'intégrale de toute fonction f mesurable *positive*.

Définition 1.10. Soit $f : \mathbf{E} \rightarrow \overline{\mathbf{R}}$ une fonction mesurable positive ($f \geq 0$). On

définit :

$$\int f d\mu = \sup \left\{ \int g d\mu : 0 \leq g \leq f, g \text{ étagée} \right\}.$$

On vérifie aisément que si f est étagée, cette définition coïncide avec la précédente.

Il reste à désormais à définir l'intégrale dans le cas général de fonctions non-nécessairement positives. Pour toute fonction mesurable f à valeurs réelles, on introduit les notations

$$\begin{aligned} f^+ &= \max(f, 0) \\ f^- &= \max(-f, 0). \end{aligned}$$

Définition 1.11. Soit $f : E \rightarrow \overline{\mathbf{R}}$ une fonction mesurable, telle que $\int f^+ d\mu < \infty$ ou $\int f^- d\mu < \infty$. On définit

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

La fonction f est dite μ -intégrable si $\int f^+ d\mu < \infty$ et $\int f^- d\mu < \infty$.

L'intégrale de f est bien définie si $\int f^+ d\mu < \infty$ ou $\int f^- d\mu < \infty$: cette précaution est nécessaire pour éviter la soustraction $\infty - \infty$ qui n'a pas de sens. La fonction f est intégrable si les deux conditions sont satisfaites simultanément. En ce sens, une fonction peut ne pas être intégrable, bien que son intégrale soit définie. L'ensemble des fonctions μ -intégrables est noté $\mathcal{L}^1(\mu)$.

Pour désigner l'intégrale de f , on utilise indifféremment les notations $\int f d\mu$, $\int f(x) d\mu(x)$, $\int f(x) \mu(dx)$ ou, de manière à première vue impropre, $\mu(f)$.

1.5.2 Premières propriétés

Lemme 1.22. Soient f, g deux fonctions mesurables de $E \rightarrow \overline{\mathbf{R}}$, dont les intégrales sont définies. Alors

1. $f \leq g$ implique $\int f d\mu \leq \int g d\mu$.
2. $|\int f d\mu| \leq \int |f| d\mu$.

Démonstration. 1. Supposons $0 \leq f \leq g$. Toute fonction φ étagée plus petite que f est aussi plus petite que g . On a donc l'inclusion

$$\left\{ \int \varphi d\mu : 0 \leq \varphi \leq f, \varphi \text{ étagée} \right\} \subset \left\{ \int \varphi d\mu : 0 \leq \varphi \leq g, \varphi \text{ étagée} \right\}.$$

Il en découle que le supremum de l'ensemble de gauche est plus petit que le supremum de l'ensemble de droite, ce qui se lit $\int f d\mu \leq \int g d\mu$.

Traisons le cas où $f \leq g$ sont quelconques. On a $f^+ \leq g^+$ et $f^- \geq g^-$, donc, d'après ci-dessus, $\int f^+ d\mu \leq \int g^+ d\mu$ et $\int f^- d\mu \geq \int g^- d\mu$. En soustrayant, $\int f d\mu \leq \int g d\mu$.

2. On a $f \leq |f|$ donc $\int f d\mu \leq \int |f| d\mu$. De même $-f \leq |f|$, donc $\int (-f) d\mu \leq \int |f| d\mu$. Or $\int (-f) d\mu = \int (-f)^+ d\mu - \int (-f)^- d\mu = \int f^- d\mu - \int f^+ d\mu = -\int f d\mu$. On a donc $-\int f d\mu \leq \int |f| d\mu$, ce qui conclut la preuve. $\square$

Dans la suite, la notation $0 \leq f_n \uparrow f$ est utilisée pour désigner une suite de fonctions f_n positives, convergeant ponctuellement vers f .

Lemme 1.23. Soient f, f_n ($n \in \mathbf{N}$) des fonctions mesurables de $E \rightarrow \overline{\mathbf{R}}^+$ telles que $0 \leq f_n \uparrow f$. Alors

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

Démonstration. Pour tout n , $f_n \leq f_{n+1} \leq f$, donc $\int f_n d\mu$ est une suite croissante majorée par $\int f d\mu$. Sa limite existe, et $\lim \int f_n d\mu \leq \int f d\mu$. Il reste à démontrer l'autre inégalité.

Soit une fonction étagée g telle que $0 \leq g \leq f$. Soit $\varepsilon > 0$ et $B_n = \{f_n \geq (1 - \varepsilon)g\}$. Décrivons g sous la forme $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$ (choisir par exemple, les α_i comme les valeurs distinctes de g et $A_i = g^{-1}(\{\alpha_i\})$). On a :

$$\begin{aligned} \int f_n d\mu &\geq \int f_n \mathbf{1}_{B_n} d\mu \geq \int (1 - \varepsilon)g \mathbf{1}_{B_n} d\mu \\ &= \int \left(\sum_{i=1}^p (1 - \varepsilon)\alpha_i \mathbf{1}_{A_i \cap B_n} \right) d\mu \\ &= \sum_{i=1}^p (1 - \varepsilon)\alpha_i \mu(A_i \cap B_n), \end{aligned}$$

où, dans la dernière égalité, on a utilisé la linéarité de l'intégrale pour les fonctions étagées (Lemme 1.21). Comme $B_n \uparrow E$, $\mu(A_i \cap B_n)$ tend vers $\mu(A_i)$ pour tout i , lorsque $n \rightarrow \infty$. Par conséquent,

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq (1 - \varepsilon) \sum_{i=1}^p \alpha_i \mu(A_i) = (1 - \varepsilon) \int g d\mu.$$

Cela étant vrai pour tout $\varepsilon > 0$, nous avons $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int g d\mu$. En prenant le supremum sur l'ensemble des fonctions g étagées positives minorant f , on conclut que $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int f d\mu$. $\square$

Le lemme qui suit établit que toute fonction mesurable (positive) peut être approchée par une limite *simple* de fonctions mesurables. Il jouera un rôle essentiel dans la suite : pour démontrer telle ou telle propriété de l'intégrale, il suffira dans un premier temps de la démontrer pour des fonctions étagées positives, et dans un deuxième temps de choisir une suite de fonctions étagées approchant les fonctions de départ, en passant à la limite grâce au Lemme 1.23.

Lemme 1.24. Toute fonction mesurable f à valeurs dans $\overline{\mathbf{R}}^+$ est limite croissante simple de fonctions étagées positives.

Démonstration. Soit

$$I = \{x : f(x) \geq n\} \text{ et } E_{n,k} = f^{-1}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right]\right).$$

Posons

$$t_n(x) = \sum_{k=0}^{2^n} \frac{k}{n} \mathbf{1}_{E_{n,k}}(x) + n \mathbf{1}_I(x).$$

Il est clair que $t_n(x) \leq f(x)$ et que t_n converge simplement vers f . En posant,

$$f_1 = t_1, f_2 = f_1 \vee t_2, f_n = f_{n-1} \vee t_n, \dots$$

on construit une suite de fonctions qui prennent chacune un nombre fini de valeurs donc sont étagées, croissante et qui converge simplement vers f . $\square$

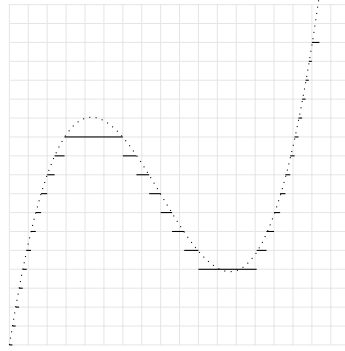


FIGURE 1.2 – Une fonction (en pointillés) et son approximation par une fonction étagée.

1.5.3 Linéarité

Théorème 1.25. Soient f, g des fonctions mesurables de $E \rightarrow \overline{\mathbf{R}}$, et $(a, b) \in \mathbf{R}^2$. On suppose que l'une au moins des deux conditions suivantes est satisfaite :

1. $f, g, a, b \geq 0$

ou

2. f et g sont intégrables.

Alors $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.

Démonstration. 1. On se place sous la première condition. Considérons deux suites de fonctions étagées f_n et g_n telles que $0 \leq f_n \uparrow f$ et $0 \leq g_n \uparrow g$. D'après le lemm 1.23,

$a \int f_n d\mu + b \int g_n d\mu \rightarrow a \int f d\mu + b \int g d\mu$. Par ailleurs, $a \int f_n d\mu + b \int g_n d\mu = \int (af_n + bg_n) d\mu$. Il est facile de vérifier que $0 \leq af_n + bg_n \uparrow af + bg$ et donc $\int (af_n + bg_n) d\mu \rightarrow \int (af + bg) d\mu$. Le résultat découle de l'identification des limites.

2. On se place sous la seconde condition. Il est facile de montrer que $\int (af) d\mu = a \int f d\mu$, il suffit donc dans la suite de considérer le cas où $a = b = 1$. En écrivant $f + g$ de deux façons différentes,

$$\begin{aligned} f + g &= (f + g)^+ - (f + g)^- \\ &= f^+ - f^- + g^+ - g^-. \end{aligned}$$

Ainsi,

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

On laisse au lecteur le soin de vérifier l'intégrabilité de $f + g$. Finalement, par passage à l'intégrale de l'équation précédente et utilisation de la linéarité dans le cas positif,

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

ce qui correspond à l'égalité souhaitée. $\square$

Deux corollaires découlent de la propriété précédente et du lemme 1.23.

Corollaire 1.26. f est μ -intégrable si et seulement si $\int |f| d\mu < \infty$.

Démonstration. Par définition $f \in \mathcal{L}^1(\mu)$ précisément si $\int f^+ d\mu + \int f^- d\mu < \infty$, soit par linéarité de l'intégrale $\int (f^+ + f^-) d\mu < \infty$. La conclusion provient du fait que $f^+ + f^- = |f|$. $\square$

Corollaire 1.27. Soit $(f_n : n \in \mathbb{N})$ une suite de fonctions mesurables à valeurs positives. Alors

$$\sum_{n \in \mathbb{N}} \int f_n d\mu = \int \left(\sum_{n \in \mathbb{N}} f_n \right) d\mu.$$

Démonstration. Posons $g_n = \sum_{k \leq n} f_k$. On a $0 \leq g_n \uparrow \sum_k f_k$ donc par le lemme 1.23, $\int g_n d\mu \rightarrow \int \sum_k f_k d\mu$. Par ailleurs, la linéarité de l'intégrale implique que $\int g_n d\mu = \sum_{k \leq n} \int f_k d\mu$ qui converge vers $\sum_k \int f_k d\mu$. Il suffit d'identifier les limites. $\square$

1.5.4 Ensembles μ -négligeables

Un ensemble μ -négligeable (ou simplement “négligeable” s'il n'y a pas d'ambiguïté) est un ensemble $N \in \mathcal{E}$ tel que $\mu(N) = 0$.

Proposition 1.28. Soit $f : E \rightarrow \overline{\mathbf{R}}$ une fonction mesurable et N un ensemble μ -négligeable. Alors

$$\int f \mathbf{1}_N d\mu = 0.$$

Démonstration. Traitons d'abord le cas où $f \geq 0$. Soit g une fonction étagée positive telle que $g \leq f \mathbf{1}_N$, disons de la forme $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$ avec $A_i = g^{-1}(\{\alpha_i\})$. On a $\int g d\mu = \sum_i \alpha_i \mu(A_i)$. Il est facile de voir que chaque terme de la somme est nulle. En effet, pour i fixé tel que $\alpha_i \neq 0$, on $g(x) = \alpha_i > 0$ et par construction de $g \leq f \mathbf{1}_N$, on a $\mathbf{1}_N(x) \neq 0$. Donc $x \in N$, ce qui montre que $A_i \subset N$, et donc $\mu(A_i) = 0$. Toute fonction étagée g minorant $f \mathbf{1}_N$ est d'intégrale nulle. Donc $f \mathbf{1}_N$ est d'intégrale nulle. On déduit la propriété dans le cas général en utilisant la décomposition $\int f \mathbf{1}_N d\mu = \int f^+ \mathbf{1}_N d\mu - \int f^- \mathbf{1}_N d\mu = 0$. $\square$

Les ensembles mesurables de mesure nulle jouent un rôle particulier parce qu'ils ne sont pas « visibles » par la mesure même s'ils sont non vides.

Définition 1.12 (Presque-partout, presque-sûre). On dit d'une propriété qu'elle est vraie μ -presque-partout ou μ -presque-sûrement, en abrégé μ -p.p. ou μ -p.s., lorsque son complémentaire est de mesure nulle.

Donnons un exemple. L'écriture " $0 \leq f_n \uparrow f$ μ -pp" signifie qu'il existe un ensemble $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et tel que pour tout $x \in E \setminus N$, $f_n(x)$ est une suite positive croissante convergeant vers $f(x)$.

Soit $(E, \mathcal{E}, \mu)$ un espace probabilisé. Soit f une fonction mesurable positive de μ -intégrale nulle. Pour tout $n > 0$ et $A_n = \{x : f(x) > 1/n\}$,

$$0 = \int f d\mu \geq \int_{A_n} f d\mu \geq \frac{1}{n} \mu(A_n),$$

donc $\mu(A_n) = 0$ pour tout $n \geq 1$. Comme $\bigcap_{n \geq 1} A_n = A_0$, par monotonie, on obtient $\mu(A_0) = 0$. On a donc démontré le théorème suivant :

Proposition 1.29. Soit f une fonction mesurable positive, de μ -intégrale nulle alors f est nulle μ -presque partout.

Mais une fonction nulle « presque partout » n'est pas une fonction nulle. Par exemple, la fonction indicatrice de $\mathbf{Q}$, l'ensemble des rationnels, est presque-partout nulle pour la mesure de Lebesgue mais elle est non nulle sur un ensemble dense !

REMARQUE.— Ainsi, l'application $f \mapsto \int |f| d\mu$ ne définit pas une norme sur l'espace vectoriel $\mathcal{L}^1(\mu)$, car la nullité de l'intégrale n'implique pas la nullité de la fonction. Afin de construire une norme, il faut que deux fonctions qui sont égales μ -pp soient identifiées comme un seul et même élément. Formellement, cela revient à définir la relation d'équivalence suivante.

f est en relation d'égalité p.p. avec g lorsque $\mu(f \neq g) = 0$. On note $f \mathcal{R} g$.

On définit alors $L^1(\mu)$ comme l'ensemble des classes d'équivalence de la relation $\mathcal{R}$, et l'application qui à toute classe d'équivalence c associe la valeur $\int |f| d\mu$ pour une fonction f quelconque de la classe c (le théorème 1.30 énoncé plus bas établit qu'elles ont toutes la même intégrale), définit bien une norme sur $L^1(\mu)$. L'espace $L^1(\mu)$ est un espace vectoriel normé *complet*, c'est à dire que toute suite de Cauchy est convergente dans cet espace. La complétude est une propriété essentielle en analyse fonctionnelle.

Dans la pratique, travailler avec des classes d'équivalence est fastidieux, on peut continuer de penser les fonctions mesurables comme des fonctions ordinaires en prenant garde qu'elles ne sont définies qu'à un ensemble de mesure nulle près.

1.5.5 Résultats fondamentaux

Théorème 1.30. Soient $f, g : E \rightarrow \overline{\mathbf{R}}$ des fonctions dont les intégrales sont définies.

1. $f \leq g$ μ -pp implique que $\int f d\mu \leq \int g d\mu$.
2. $f = g$ μ -pp implique que $\int f d\mu = \int g d\mu$.

Démonstration. 1. Soit $N = \{f > g\}$. On a $\mu(N) = 0$. Comme $f\mathbf{1}_N \leq g\mathbf{1}_N$, on a $\int f\mathbf{1}_N d\mu \leq \int g\mathbf{1}_N d\mu$. Par ailleurs $\int f\mathbf{1}_{\overline{N}} d\mu = \int g\mathbf{1}_{\overline{N}} d\mu = 0$. Donc, par linéarité,

$$\int (f\mathbf{1}_N + f\mathbf{1}_{\overline{N}}) d\mu \leq \int (g\mathbf{1}_N + g\mathbf{1}_{\overline{N}}) d\mu,$$

soit $\int f d\mu \leq \int g d\mu$.

2. Immédiat à partir du résultat 1. □

Théorème 1.31 (Théorème de convergence monotone). Soit $(f_n, n \geq 1)$ une suite de fonctions (mesurables) positives qui converge en croissant, μ -p.p. vers f alors

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

La preuve découle directement du lemme 1.23, en traitant les ensembles négligeables de la même façon que dans la preuve du théorème 1.30.

Théorème 1.32 (Lemme de Fatou). Soit $(f_n, n \geq 1)$ une suite de fonctions (mesurables) positives,

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

Démonstration. Par définition,

$$\liminf_n f_n(x) = \sup_k \inf_{n \geq k} f_n(x) = \lim_k \inf_{n \geq k} f_n(x).$$

La suite $(g_k = \inf_{n \geq k} f_n, k \in \mathbb{N})$ est croissante positive donc le théorème de convergence monotone assure que

$$\int \lim_k g_k d\mu = \lim_k \int g_k d\mu.$$

Par conséquent,

$$\int \liminf_n f_n d\mu = \lim_k \int \inf_{n \geq k} f_n d\mu \leq \lim_k \inf_{n \geq k} \int f_n d\mu = \liminf_n \int f_n d\mu.$$

La preuve est terminée. $\square$

Si la mesure est finie, on peut évidemment remplacer l'hypothèse de positivité par l'hypothèse que les f_n sont inférieurement bornées. Dans le cas où toutes les fonctions ne sont pas positives, il faut une contrainte de domination.

Théorème 1.33 (Convergence dominée). Soit $(f_n, n \geq 1)$ une suite de fonctions (mesurables) qui converge μ p.p. vers f . Si de plus, il existe g telle que

$$|f_n(x)| \leq g(x), \text{ pour presque tout } x \text{ et } \int g d\mu < \infty$$

alors

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

Démonstration. Comme $|f_n| \leq g$ et que l'intégrale de g est finie, celle de $|f_n|$ l'est aussi pour tout n entier. D'autre part, $g - f_n$ et $g + f_n$ sont mesurables positives et $g \pm f_n$ tend vers $g \pm f$ donc

$$\liminf_n g \pm f_n = g \pm f.$$

Appliquons le lemme de Fatou aux deux suite $((g \pm f_n), n \geq 1)$, on obtient

$$\begin{aligned} \int (g - f) d\mu &\leq \liminf_n \int (g - f_n) d\mu \\ \int (g + f) d\mu &\leq \liminf_n \int (g + f_n) d\mu. \end{aligned}$$

De plus, le lemme de Fatou implique aussi que

$$\int |f| d\mu = \int \liminf_n |f_n| d\mu \leq \liminf_n \int |f_n| d\mu \leq \int g d\mu < +\infty,$$

donc f est intégrable. On peut donc écrire

$$\begin{aligned} \int g d\mu - \int f d\mu &\leq \int g d\mu + \liminf_n \int (-f_n) d\mu \\ &= \int g d\mu - \limsup_n \int f_n d\mu \\ \int g d\mu + \int f d\mu &\leq \int g d\mu + \liminf_n \int f_n d\mu. \end{aligned}$$

On tire de ces inégalités que

$$\int f d\mu \leq \liminf_n \int f_n d\mu \leq \limsup_n \int f_n d\mu \leq \int f d\mu,$$

d'où l'on conclut que la suite $\int f_n d\mu$ converge et que la limite est $\int f d\mu$. $\square$

1.5.6 Exemples

EXEMPLE 10 (Mesure de comptage).— Si $\mu = \sum_{n \in \mathbf{N}} \delta_n$, alors

$$\mu(A) = \text{card}(A) = \sum_{n \in A} 1 = \sum_{n \in \mathbf{N}} \mathbf{1}_A(n) = \int \mathbf{1}_A d\mu.$$

Par conséquent, pour $f : \mathbf{N} \rightarrow \mathbf{R}$, intégrable,

$$\int f d\mu = \sum_{n \in \mathbf{N}} f(n).$$

Il faut donc garder à l'esprit que l'intégrale de Lebesgue peut se réaliser en ce qui est communément appelé une série.

EXEMPLE 11 (Mesure de Lebesgue).— Par construction, la mesure de Lebesgue est telle que

$$\lambda([a, b[) = b - a.$$

Réécrit dans le langage de la TdM, cela revient à dire

$$\int \mathbf{1}_{[a, b[} d\lambda = b - a = \int_a^b dx, \quad (1.4)$$

où $\int f(x)dx$ représente l'intégrale de Riemann.

REMARQUE.— Pour les fonctions en escalier, qui sont des cas particuliers de fonctions étagées avec des A_i qui sont nécessairement des intervalles, les intégrales de Riemann et Lebesgue coïncident. Plus généralement, on admettra que toutes les fonctions Riemann intégrables (fonctions continues sur un compact ou d'intégrale absolument convergente) sont Lebesgue intégrables (voir [?] pour la démonstration). C'est pour cette raison que l'on notera

$$\int f d\lambda = \int f(x) dx,$$

même si maintenant le terme de droite désigne l'intégrale de Lebesgue de f .

REMARQUE.— L'intégration « à la Lebesgue » permet d'intégrer plus de fonctions qu'avec l'intégrale de Riemann : pour être Riemann intégrable, une fonction (positive) se doit d'être continue sauf en un nombre dénombrable de points. Pour être Lebesgue intégrable, une fonction positive peut se contenter d'être mesurable. La notion de mesurabilité est nettement moins contraignante que la continuité.

Il est par exemple clair que $\mathbf{Q}$ est mesurable (en tant que réunion dénombrable de singletons), de mesure de Lebesgue nulle (puisque les singletons sont de mesure nulle

et en vertu de l'axiome 1.1.iii) de la définition 1.4). Par conséquent, $\mathbf{1}_{\mathbb{Q}}$ est mesurable mais discontinue en tout point irrationnel : soit $r \in \mathbb{Q}$, il existe une suite $(q_n, n \geq 1)$ de rationnels qui converge vers r et

$$\lim_{n \rightarrow +\infty} \mathbf{1}_{\mathbb{Q}}(q_n) = 1 \neq 0 = \mathbf{1}_{\mathbb{Q}}(r).$$

Cette fonction est donc Lebesgue intégrable mais pas Riemann intégrable.

Plus généralement, l'intérêt de l'intégrale de Lebesgue se trouve aussi dans la complétude de l'espace des fonctions intégrables, dont il a été question plus haut.

1.5.7 Théorème de Fubini-Tonelli

Le théorème de Fubini est celui qui permet de calculer des intégrales multiples en choisissant l'ordre des intégrations. La mesure produit a été définie dans la définition 1.20.

Théorème 1.34 (Fubini-Tonelli). Soit $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$ un espace mesuré produit et f une fonction mesurable de $E \times F$ dans $\mathbf{R}^+$. On a

- $x \mapsto \int f(x, y) d\nu(y)$ est mesurable,
- $y \mapsto \int f(x, y) d\mu(x)$ est mesurable,
- on peut choisir l'ordre d'intégration :

$$\int f(x, y) d(\mu \otimes \nu)(x, y) = \int \left(\int f(x, y) d\nu(y) \right) d\mu(x) = \int \left(\int f(x, y) d\mu(x) \right) d\nu(y). \quad (1.5)$$

Démonstration. Pour $f = \mathbf{1}_{A \times B}$ avec $A \in \mathcal{E}$ et $B \in \mathcal{F}$,

$$\int f(x, y) d\nu(y) = \nu(B) \mathbf{1}_A(x),$$

qui est une fonction mesurable puisque $A \in \mathcal{E}$. Par un argument de classe monotone, on en déduit que ce résultat reste vrai pour $f = \mathbf{1}_C$ avec $C \in \mathcal{E} \otimes \mathcal{F}$ dont on rappelle qu'il s'agit de la tribu engendrée par les ensembles de la forme $A \times B$ avec $A \in \mathcal{E}$ et $B \in \mathcal{F}$. Par linéarité, la mesurabilité est toujours vérifiée pour les fonctions étagées (cf. définition 1.8) de $E \times F$ dans $\mathbf{R}^+$. Pour une fonction mesurable positive, par construction

$$\int f(x, y) d\nu(y) = \lim_{n \rightarrow \infty} \int f_n(x, y) d\nu(y)$$

où $(f_n, n \geq 1)$ est une suite de fonctions étagées qui converge vers f . La fonction

$$x \mapsto \int f(x, y) d\nu(y)$$

apparaît donc comme la limite d'une suite de fonctions mesurables, elle est donc mesurable (cf. 1.12.c). Le même raisonnement s'applique évidemment pour

$$y \mapsto \int f(x, y) d\mu(x).$$

La deuxième partie de la preuve suit rigoureusement le même schéma. Pour $f = \mathbf{1}_{A \times B}$ avec $A \in \mathcal{E}$ et $B \in \mathcal{F}$,

$$\int f(x, y) d(\mu \otimes \nu)(x, y) = \mu \otimes \nu(A \times B) = \mu(A)\nu(B),$$

et

$$\int \left(\int f(x, y) d\nu(y) \right) d\mu(x) = \int \nu(B) \mathbf{1}_A(x) d\mu(x) = \nu(B) \int \mathbf{1}_A(x) d\mu(x) = \nu(B)\mu(A),$$

$$\int \left(\int f(x, y) d\mu(x) \right) d\nu(y) = \int \mu(A) \mathbf{1}_B(y) d\nu(y) = \mu(A) \int \mathbf{1}_B(y) d\nu(y) = \mu(A)\nu(B).$$

La formule (1.5) est donc vraie pour ces fonctions f particulières. Par un argument de classe monotone, il reste vrai pour les indicatrices de $C \in \mathcal{E} \otimes \mathcal{F}$. Par linéarité, il est vrai pour les fonctions étagées. Par passage à la limite, il est vrai pour les fonctions mesurables positives. $\square$

Dans le cas où n'est pas de signe constant, on vérifie d'abord que l'intégrale double de $|f|$ est finie (en utilisant le théorème précédent pour la calculer) puis on applique le théorème de Fubini proprement dit.

Théorème 1.35 (Fubini). Soit $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$ un espace mesuré produit et f une fonction mesurable de $E \times F$ dans $\mathbf{R}$. Si

$$\iint |f(x, y)| d\mu(x) d\nu(y) < \infty,$$

alors les conclusions du théorème 1.34 restent inchangées.

Démonstration. D'après le théorème de Fubini-Tonnelli,

$$\int \left(\int |f(x, y)| d\nu(y) \right) d\mu(x) < \infty$$

donc la fonction

$$x \mapsto \int |f(x, y)| d\nu(y)$$

est finie μ -presque-sûrement : si l'on note

$$N_E = \{x \in E, \int |f(x, y)| d\nu(y) = \infty\}$$

alors $\mu(N_E) = 0$ et donc, par construction de la mesure produit, $(\mu \otimes \nu)(N_E \times F) = 0$. De même,

$$N_F = \{y \in F, \int |f(x, y)| d\mu(x) = \infty\}$$

est de ν mesure nulle donc $(\mu \otimes \nu)(E \times N_F) = 0$. On peut redéfinir f par 0 sur ces deux ensembles. En décomposant cette fonction (redéfinie mais toujours notée f) en $f = f^+ - f^-$, le résultat découle par linéarité de (1.5). $\square$

REMARQUE.— Il convient de noter que le théorème de Fubini permet de permuter des intégrales mais aussi des séries (puisque une série est au sens de Lebesgue, une intégrale par rapport à la mesure de comptage) ou des séries et des intégrales (ce que vous avez appelé jusqu'à présent le théorème d'intégration terme à terme).

REMARQUE.— Les plus attentifs remarqueront que le théorème suivant se distingue de celui que vous connaissez pour l'intégrale de Riemann en ce qu'il n'est pas nécessaire de vérifier que la fonction dominante et la fonction limite sont continues !

Théorème 1.36 (Continuité sous le signe somme). Soit I un ouvert de $\mathbf{R}^n$ et $\{f(x, t), t \in I\}$ une famille de fonctions mesurables telle que pour tout $t \in I$, $f(\cdot, t)$ soit μ -intégrable. Si $\mu(x)$ -p.p., $t \mapsto f(x, t)$ est continue sur I , s'il existe G une fonction mesurable telle que

- $\mu(x)$ p.p., $|f(x, t)| \leq G(x)$ pour tout $t \in I$,
- $\int G d\mu < \infty$,

alors l'application $t \mapsto \int f(x, t) d\mu(x)$ est continue.

Démonstration. Soit $(t_n, n \geq 1)$ une suite d'éléments de I qui tend vers t . Alors, on a évidemment

$$\left| \int f(x, t) d\mu(x) - \int f(x, t_n) d\mu(x) \right| \leq \int_E |f(x, t) - f(x, t_n)| d\mu(x).$$

Les assertions suivantes sont immédiates :

- $g_n : x \mapsto |f(x, t) - f(x, t_n)|$ est mesurable ;
- $g_n(x) = |f(x, t) - f(x, t_n)| \xrightarrow{n \rightarrow \infty} 0$, μ -p.p. ;
- $|g_n(x)| = |f(x, t) - f(x, t_n)| \leq 2G(x)$, μ -p.p. ;
- $\int G d\mu < \infty$;

donc on peut appliquer le théorème de convergence dominée (théorème 1.33) à g_n , ce qui induit que

$$\int_E |f(x, t) - f(x, t_n)| d\mu(x) \xrightarrow{n \rightarrow \infty} 0.$$

D'où le résultat. $\square$

Une preuve similaire permet d'obtenir le résultat de dérivation sous le signe « somme ».

Théorème 1.37. Soit I un ouvert de $\mathbf{R}^n$ et $\{f(x, t), t \in I\}$ une famille de fonctions mesurables telles que pour tout $t \in I$, $f(\cdot, t)$ soit μ -intégrable. Si $t \mapsto f(x, t)$ est dérivable sur I , $d\mu$ p.p., s'il existe $G(x)$ une fonction mesurable telle que

- μ p.p., $|\frac{d}{dt}f(x, t)| \leq G(x)$ pour tout $t \in I$,
- $\int G d\mu < \infty$,

alors l'application

$$\begin{aligned} I &\longrightarrow \mathbf{R} \\ t &\longmapsto \int f(x, t) d\mu(x) \end{aligned}$$

est dérivable sur I et

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

1.5.8 Fonctions test

Pour caractériser une mesure de probabilité sur $\mathbf{R}^d$, il est souvent utile de considérer l'intégrale d'une fonction « test », mesurable positive ou mesurable bornée.

Théorème 1.38. Deux mesures de probabilité $\mathbf{P}$ et $\mathbf{Q}$ sur $\mathbf{R}^d$ sont égales si et seulement si pour toute fonction f mesurable positive (respectivement, mesurable bornée),

$$\int f d\mathbf{P} = \int f d\mathbf{Q}. \quad (1.6)$$

Démonstration. Il suffit de remarquer que pour tout borélien B de $\mathbf{R}^d$, la fonction $f = \mathbf{1}_B$ est mesurable positive (respectivement, mesurable bornée). L'égalité (1.6) donne $\mathbf{P}(B) = \mathbf{Q}(B)$, si bien que $\mathbf{P} = \mathbf{Q}$. $\square$

1.5.9 Mesures à densité

Définition 1.13. Soit μ et ν deux mesures sur $(E, \mathcal{E})$. On dit que ν est absolument continue par rapport à μ si et seulement si

$$A \in \mathcal{E}, \mu(A) = 0 \implies \nu(A) = 0.$$

On note $\nu \ll \mu$.

Théorème 1.39 (Radon-Nikodym). Soit μ et ν deux mesures, finies sur tous compacts, sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ telles que $\nu \ll \mu$. Il existe une unique fonction mesurable $f \geq 0$ (définie μ presque-sûrement), telle que

$$\int_{\mathbf{R}^d} f d\mu < \infty \text{ et } \int_{\mathbf{R}^d} g d\nu = \int_{\mathbf{R}^d} g f d\mu,$$

pour toute fonction g mesurable bornée. La fonction f s'appelle la *densité* de ν par rapport à μ .

Lorsque l'on dit qu'une mesure ν est absolue continue sans autre précision, on sous-entend qu'elle l'est par rapport à la mesure de Lebesgue, c'est-à-dire qu'il existe f telle que

$$d\nu(x) = f(x)dx.$$

Définition 1.14. Le support d'une mesure μ est le complémentaire du plus grand ouvert de μ -mesure nulle.

EXEMPLE 12.— Par exemple, le support de la mesure δ_x est le singleton $\{x\}$. Pour une mesure à densité (sous-entendu par rapport à la mesure de Lebesgue) f supposée continue par morceaux, le support de la mesure $f d\lambda$ est l'adhérence de l'ensemble des valeurs où f ne s'annule pas.

EXEMPLE 13.— Soit $\mu = \delta_0$, les mesures absolument continues par rapport à μ sont de la forme $\nu = \alpha \delta_0$ avec $\alpha \geq 0$. Soit $\nu \ll \mu$, pour tout ouvert A ne contenant pas 0, $\mu(A) = 0$ donc $\nu(A) = 0$. Le support de ν est donc inclus dans $\{0\}$. Ce qui signifie que $\nu = \alpha \delta_0$. La fonction g est alors définie par $g(0) = \alpha$ et g quelconque sur $\mathbf{R} \setminus \{0\}$. En effet,

$$\int f d\nu = f(0)\alpha = \int f g d\mu.$$

REMARQUE.— Pour μ et ν , deux mesures de probabilité telles que $\nu \ll \mu$, on remarquera (en prenant $g = 1$) que nécessairement

$$\int_{\mathbf{R}^d} f d\mu = 1. \quad (1.7)$$

De nombreuses probabilités sont définies comme des mesures à densité par rapport à la mesure de Lebesgue. La table 1.1 fournit différents exemples de densités de probabilité f , sur lesquels nous aurons l'occasion de revenir pendant ce cours.

A titre d'exemple, la figure 1.3 représente la densité gaussienne de paramètres 0 et 1, notée $\mathcal{N}(0, 1)$ et appelée gaussienne *centrée réduite* :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (1.8)$$

La densité gaussienne porte aussi le nom de densité *normale*, ce qui justifie la notation $\mathcal{N}$.

Domaine	Densité	Expression de $f(x)$	Notation
$\mathbf{R}$	Densité uniforme sur $[a, b]$	$\frac{\mathbf{1}_{[a,b]}(x)}{b-a}$	$\mathcal{U}([a, b])$
$\mathbf{R}^d$	Densité uniforme sur une partie $A \subset \mathbf{R}^d$	$\frac{\mathbf{1}_A(x)}{\int \mathbf{1}_A}$	
$\mathbf{R}$	Densité exponentielle de paramètre $\alpha > 0$	$\alpha e^{-\alpha x} \mathbf{1}_{\mathbf{R}_+}(x)$	$\mathcal{E}(\alpha)$
$\mathbf{R}$	Densité gaussienne de paramètres m, σ^2 ($m \in \mathbf{R}, \sigma^2 > 0$)	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$	$\mathcal{N}(m, \sigma^2)$
$\mathbf{R}^d$	Gaussienne multivariée de paramètres m, Σ ($m \in \mathbf{R}^d, \Sigma \in \mathbf{R}^{d \times d}$ définie positive)	$\frac{e^{-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)}}{\sqrt{(2\pi)^d \det \Sigma}}$	$\mathcal{N}_d(m, \Sigma)$
$\mathbf{R}$	Densité de Cauchy de paramètres m, α ($m \in \mathbf{R}, \alpha > 0$)	$\frac{1}{\pi} \cdot \frac{\alpha}{(x-m)^2 + \alpha^2}$	
$\mathbf{R}$	Densité Gamma de paramètres a, b ($a >, b > 0$)	$\frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$	$\Gamma(a, b)$

TABLE 1.1 – Quelques exemples de densités de probabilité – (Rappel : $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$).

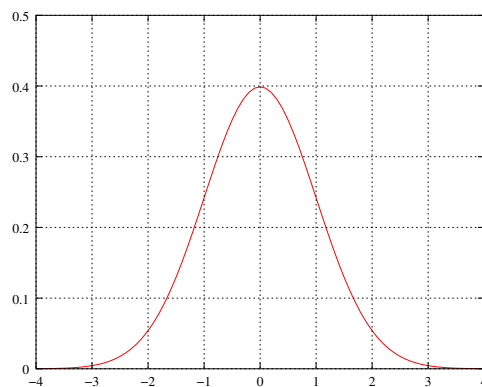


FIGURE 1.3 – Densité gaussienne centrée réduite $\mathcal{N}(0, 1)$.



Fonctions de répartition

2.1 Variables aléatoires réelles

Vous avez déjà vu les variables aléatoires à valeurs dans un espace dénombrable. Les variables aléatoires réelles sont très similaires sauf sur certains aspects techniques.

Définition 2.1. Une variable aléatoire X est une fonction de $(\Omega, \mathcal{A}, \mathbf{P})$ dans $\mathbf{R}$ telle que pour tout intervalle $]a, b[$ de $\mathbf{R}$, l'événement $(\omega : X(\omega) \in]a, b[)$ est dans $\mathcal{A}$.

En pratique et en tout cas, dans tout ce cours, toutes les applications que l'on considérera seront mesurables donc des variables aléatoires.

Définition 2.2. La loi d'une variable aléatoire réelle (v.a.r. en abrégé) X est la mesure image de $\mathbf{P}$ par X , c'est-à-dire la mesure $\mathbf{P}_X$ sur $\mathbf{R}$ définie par

$$\mathbf{P}_X(A) = \mathbf{P}(\omega : X(\omega) \in A) = \mathbf{P}(X \in A).$$

Comme $\mathbf{P}_X$ est une mesure sur $\mathbf{R}$, on sait qu'elle est totalement caractérisée par les valeurs de $\mathbf{P}_X([-\infty, b])$ pour b parcourant $\mathbf{R}$.

Définition 2.3. Soit X une v.a.r., la fonction

$$\begin{aligned} F_X : \mathbf{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mathbf{P}_X([-\infty, x]) = \mathbf{P}(X \leq x) \end{aligned}$$

s'appelle la fonction de répartition de X .

En vertu des propriétés de monotonie des mesures, F_X possède les propriétés suivantes :

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$,
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$,
- F_X est croissante, continue à droite, i.e., $\lim_{y \downarrow x} F_X(y) = F_X(x)$.

REMARQUE.— Réciproquement, on verra plus tard que toute fonction satisfaisant ces propriétés est la fonction de répartition d'une variable aléatoire réelle.

On a, d'après les propriétés de monotonie des mesures,

$$F_X(x_-) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{n=1}^{\infty}]-\infty, x - \frac{1}{n}]\right) = \mathbf{P}_X(]-\infty, x]).$$

Par conséquent, $F_X(x_-) = \mathbf{P}_X(]-\infty, x])$ et donc

$$F_X(x) - F_X(x_-) = \mathbf{P}(X = x).$$

En d'autres termes, si F_X est continue en x , $\mathbf{P}(X = x) = 0$.

Lemme 2.1. Le nombre de ces points de discontinuité d'une fonction de répartition est au plus dénombrable.

Démonstration. Soit A l'ensemble des points de discontinuité de F_X et

$$A_n = \left\{x, \Delta F_X(x) \geq \frac{1}{n}\right\}.$$

Comme F_X varie de 0 à 1, il est clair que

$$\text{card}(A_n) \leq n.$$

Comme

$$A = \bigcup_{n=1}^{\infty} A_n,$$

l'ensemble A apparaît comme la réunion dénombrable d'ensembles finis donc il est au plus dénombrable. $\square$

Comme F_X est bornée, le nombre de ces points de discontinuité est au plus dénombrable. Soit $\{x_n, n \in \mathbf{N}^*\}$ ces points. On peut alors parler de F_X^c , la régularisée de F_X :

$$\begin{aligned} F_X^c(x) &= F_X(x) - \sum_{n=1}^{\infty} \left(F_X(x_n) - F_X(x_{n-}) \right) \mathbf{1}_{[x_n, +\infty[}(x). \\ &= F_X(x) - \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x). \end{aligned}$$

La fonction F_X^c est continue et croissante par définition. Elle est d'après un théorème de Lebesgue, dérivable sauf sur un ensemble de mesure de Lebesgue nulle. Dans la suite, nous ne nous préoccupons pas de savoir ce qui se passe si elle n'est pas dérivable en tout point.

Théorème 2.2. Soit X une v.a.r. de fonction de répartition F_X . Si F_X^c est dérivable sur $\mathbf{R}$, alors

$$d\mathbf{P}_X(x) = \sum_{n=1}^{\infty} \Delta F_X(x_i) \delta_{x_i} + (F_X^c)'(x) dx. \quad (2.1)$$

Si F_X est continue alors

$$d\mathbf{P}_X(x) = (F_X^c)'(x)dx$$

et $(F_X^c)'$ s'appelle la densité de la loi de X .

Démonstration. Remarquons que

$$\mathbf{1}_{[x, +\infty[}(x) = \delta_{x_n}(\cdot - \infty, x],$$

puisque le terme de gauche ne vaut 1 que si $x \geq x_n$, soit de manière équivalente $x_n \in]-\infty, x]$. Si F_X^c est dérivable en tout point alors on a

$$F_X(x) = F_X^c(x) + \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x)$$

soit

$$\mathbf{P}(X \in]-\infty, x]) = \int_{-\infty}^x (F_X^c)'(s)ds + \sum_{n=1}^{\infty} \Delta F_X(x_i) \delta_{x_i}(\cdot - \infty, x]).$$

Les deux mesures de part et d'autre de l'égalité (2.1) coïncident donc sur les ensembles de la forme $] - \infty, x]$ pour tout x réel. C'est suffisant pour assurer que ces deux mesures sont égales. $\square$

REMARQUE.— Dans le cas où $F_X^c \equiv 0$,

$$d\mathbf{P}_X(x) = \sum_{i=1}^{\infty} \Delta F_X(x_i) \delta_{x_i},$$

ce qui signifie que X prend un nombre dénombrable de valeurs, c'est une v.a. à valeurs discrètes et $\mathbf{P}(X = x_i) = \Delta F_X(x_i)$.

Définition 2.4. Une v.a.r. X est dite de loi exponentielle de paramètre $\lambda > 0$ lorsque

$$d\mathbf{P}_X(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbf{R}^+}(x)dx$$

ou de manière équivalente

$$F_X(x) = 1 - \exp(-\lambda x).$$

On notera dorénavant ceci par $X \sim \mathcal{E}(\lambda)$.

Définition 2.5. Une v.a.r. X est dite de loi uniforme sur $[a, b]$ lorsque

$$d\mathbf{P}_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)dx.$$

On notera dorénavant ceci par $X \sim \mathcal{U}(a, b)$.

Définition 2.6. Une v.a.r. X est dite de loi gaussienne (ou normale) de paramètres m et σ^2 lorsque

$$d\mathbf{P}_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

On notera dorénavant ceci par $X \sim \mathcal{N}(m, \sigma^2)$.

Définition 2.7. Une v.a.r. X est dite de loi de Cauchy de paramètre c lorsque

$$d\mathbf{P}_X(x) = \frac{c}{\pi(c^2 + x^2)} dx.$$

On notera dorénavant ceci par $X \sim \mathcal{C}(c)$.

EXEMPLE 14.— Soit Z une v.a.r. de loi exponentielle de paramètre λ et T un réel fixé. Soit $X = \min(Z, T)$, on veut calculer la loi de X .

Il est clair que X est positive et majorée par T donc $\mathbf{P}(X \geq T) = 0$ et $\mathbf{P}(X < 0) = 0$. Pour $0 \leq x < T$, X ne peut être inférieure à x que si $X = Z$ donc

$$\mathbf{P}(X \leq x) = \mathbf{P}(Z \leq T, Z \leq x) = \mathbf{P}(Z \leq x),$$

puisque $x < T$. Par conséquent, pour $x \in [0, T[$,

$$\mathbf{P}(X \leq x) = 1 - e^{-\lambda x}.$$

En particulier, $F_X(T_-) = 1 - \exp(-\lambda T)$. Comme $\mathbf{P}(X \leq T) = 1$, on a $\mathbf{P}(X = T) = \exp(-\lambda T)$. Le tout se résume dans le graphique 2.1.

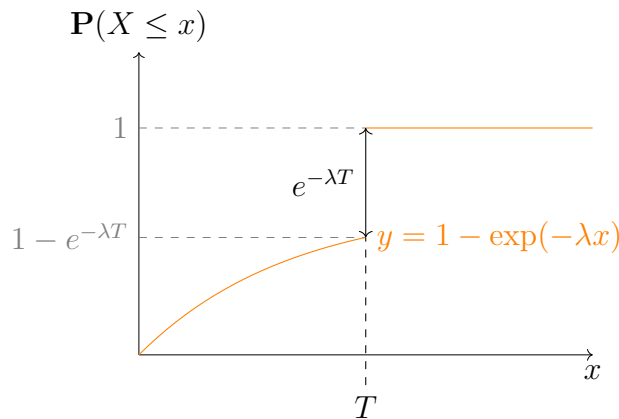


FIGURE 2.1 – Fonction de répartition de $\min(\mathcal{E}(\lambda), T)$.

2.2 Construction de variables aléatoires

Définition 2.8. Soit F une fonction de répartition. On définit

$$F^{-1}(u) = \inf \{x, F(x) \geq u\}.$$

Lemme 2.3. Pour tout $y \in \mathbf{R}$,

$$\left(u, F^{-1}(u) \leq y\right) = \left(u, u \leq F(y)\right).$$

Démonstration. Si $u \leq F(y)$, par définition de F^{-1} , on voit que $y \geq F^{-1}(u)$, d'où la première inclusion.

Si $F^{-1}(u) \leq y$, comme F est croissante

$$\begin{aligned} F(y) &\geq F(F^{-1}(u)) \\ &= F\left(\inf\{z, F(z) \geq F(u)\}\right). \end{aligned}$$

Soit $(z_n)_{n \in \mathbf{N}}$ une suite qui tend vers l'inf atteint en z . On a $F(z_n) \geq F(u)$ et comme F est continue à droite, $F(z) \geq u$, d'où le résultat. $\square$

Corollaire 2.4. Soit F une fonction de répartition et U une variable aléatoire de loi uniforme sur $[0, 1]$. La variable aléatoire

$$X = F^{-1}(U)$$

a pour fonction de répartition F .

Démonstration. D'après le lemme précédent,

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

par définition de la fonction de répartition d'une variable aléatoire de loi uniforme. $\square$

Moments

3.1 Espérance

Définition 3.1. Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé et X une variable aléatoire à valeurs dans $\mathbf{R}$. On appelle espérance de X , notée $\mathbf{E}[X]$ la quantité

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

REMARQUE.— Retrouvons l'espérance d'un lancer de dé avec cette définition. Ici, $\Omega = \{1, \dots, 6\}$ et $\mathbf{P}(\{\omega\}) = 1/6$. La v.a. X qui représente le résultat du lancer d'un dé est définie par $X(\omega) = \omega$. Par conséquent,

$$\mathbf{E}[X] = \sum_{\omega=1}^6 \omega \mathbf{P}(\{\omega\}) = \sum_{\omega=1}^6 \omega \frac{1}{6} = 3,5.$$

Théorème 3.1 (Théorème de transfert). Soit X une variable aléatoire à valeurs dans $\mathbf{R}$ de loi $\mathbf{P}_X$. Pour toute fonction $f : \mathbf{R} \rightarrow \mathbf{R}^+$,

$$\mathbf{E}[f(X)] = \int_{\mathbf{R}} f(x) d\mathbf{P}_X(x). \quad (3.1)$$

On note $L^1(\mathbf{P}_X)$ l'ensemble des fonctions (mesurables) $f : \mathbf{R} \rightarrow \mathbf{R}$ telles que

$$\mathbf{E}[|f(X)|] < +\infty.$$

Pour toute fonction $f \in L^1(\mathbf{P}_X)$,

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

Preuve (Hors programme). Par définition de la loi d'une v.a.

$$\mathbf{P}(X \in A) = \mathbf{P}_X(A).$$

En termes d'intégrales, cela signifie

$$\int_{\Omega} \mathbf{1}_A(X(\omega)) d\mathbf{P}(\omega) = \int_{\mathbf{R}^k} \mathbf{1}_A(x) d\mathbf{P}_X(x).$$

Ce qui signifie que la relation (3.1) est vraie pour $f = \mathbf{1}_A$. Par linéarité, cette relation est donc satisfaite pour les fonctions étagées. Par convergence monotone, elle est satisfaite pour les fonctions mesurables positives. En prenant les parties positives et négatives de $f \circ X$, elle l'est aussi dès que f appartient à $L^1(\mathbf{P}_X)$. $\square$

REMARQUE.— On retiendra de cette preuve le résultat essentiel qui découle de la construction de l'intégrale par rapport à une mesure :

$$\mathbf{P}(X \in A) = \mathbf{E}[\mathbf{1}_A(X)]. \quad (3.2)$$

REMARQUE.— On réalise que pour calculer l'espérance de $f(X)$ on ne travaille que sur l'espace des valeurs de la variable aléatoire X : on n'a pas besoin de réellement préciser l'espace de départ, d'où le flou systématique qui entoure cet objet.

REMARQUE.— On remarque que deux v.a. qui ont la même loi, ont mêmes espérances. Réciproquement, d'après le théorème de Riesz 1.38, si $\mathbf{E}[f(X)] = \mathbf{E}[f(Y)]$ pour toute fonction f continue bornée alors X et Y ont la même loi.

Compte-tenu de la décomposition de la fonction de répartition que nous avons vue précédemment, on en déduit

Théorème 3.2. Soit X une v.a. à valeurs dans $\mathbf{R}$ telle que

$$d\mathbf{P}_X(x) = \sum_{i=1}^{\infty} \alpha_i \delta_{x_i} + f_X(x) dx,$$

avec $\alpha_1, \dots, \alpha_n$ des réels strictement positifs. Soit h une fonction mesurable de $\mathbf{R}$ dans $\mathbf{R}$. Si h vérifie l'une des deux propriétés suivantes :

- h est à valeurs positives,
- $\mathbf{E}[|h(X)|]$ est finie,

alors

$$\mathbf{E}[h(X)] = \sum_{i=1}^{\infty} \alpha_i h(x_i) + \int_{\mathbf{R}} h(x) f_X(x) dx. \quad (3.3)$$

3.2 Variance et autres moments

Si l'espérance indique la valeur « moyenne » d'une variable aléatoire, on peut s'intéresser à mesurer les écarts moyens de ces valeurs par rapport à la moyenne. On introduit ce que l'on appelle la variance :

$$\text{var}(X) = \mathbf{E}[|X - \mathbf{E}[X]|^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2.$$

REMARQUE.— Si $\text{var}(X) = 0$ alors X est p.s. constante, autrement dit, X est déterministe.

Plus généralement, le moment d'ordre p d'une variable aléatoire réelle X est défini par $\mathbf{E}[|X|^p]$.

3.3 Inégalités

Théorème 3.3 (Inégalité de Jensen). Soit X une variable aléatoire réelle et ϕ une fonction convexe définie sur $\mathbf{R}$ alors

$$\phi(\mathbf{E}[X]) \leq \mathbf{E}[\phi(X)].$$

En particulier, pour $\phi(x) = |x|$, on obtient

$$|\mathbf{E}[X]| \leq \mathbf{E}[|X|].$$

Théorème 3.4 (Inégalité de Bienaymé-Tchebycev). Soit X une variable aléatoire réelle de variance finie, pour tout réel $\eta > 0$,

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq \eta) \leq \frac{1}{\eta^2} \text{var}[X].$$

Ce théorème est un cas particulier de l'identité suivante.

Théorème 3.5 (Inégalité de Markov). Soit $p \geq 1$ et X une variable aléatoire réelle de moment d'ordre p fini, pour tout réel $\eta > 0$,

$$\mathbf{P}(|X| \geq \eta) \leq \frac{1}{\eta^p} \mathbf{E}[|X|^p].$$

Démonstration. En découpant l'intégrale, on obtient

$$\begin{aligned} \mathbf{E}[|X|^p] &= \mathbf{E}[|X|^p \mathbf{1}_{\{|X| \geq \eta\}}] + \mathbf{E}[|X|^p \mathbf{1}_{\{|X| < \eta\}}] \\ &\geq \eta^p \mathbf{P}(|X| \geq \eta) + 0. \end{aligned}$$

Le résultat s'obtient en divisant les deux membres de l'inégalité par η^p . □

Théorème 3.6 (Inégalité de Cauchy-Schwarz). Soient X et Y deux v.a. de carré intégrable,

$$|\mathbf{E}[XY]| \leq \mathbf{E}[|XY|] \leq \left(\mathbf{E}[X^2]\right)^{1/2} \left(\mathbf{E}[Y^2]\right)^{1/2}$$

Corollaire 3.7. Si X est de carré intégrable alors X est intégrable.

L'inégalité de Cauchy-Schwarz peut se démontrer en recopiant la démonstration qui existe pour tout espace préhilbertien mais elle peut aussi se voir comme un cas particulier des inégalités de Hölder.

Théorème 3.8 (Inégalités de Hölder). Soit X et Y deux v.a. réelles. Soit $p \geq 1$ et q tel que $1/p + 1/q = 1$.

$$|\mathbf{E}[XY]| \leq \mathbf{E}[|XY|] \leq \left(\mathbf{E}[|X|^p]\right)^{1/p} \left(\mathbf{E}[|Y|^q]\right)^{1/q}$$

3.4 Fonction caractéristique

La fonction caractéristique est aux variables aléatoires continues ce qu'est la fonction génératrice aux variables aléatoires discrètes : un outil de calcul puissant permettant d'identifier des lois et de calculer des moments.

Définition 3.2 (Fonction caractéristique). Soit X une variable aléatoire réelle, on pose pour tout réel t ,

$$\phi_X(t) = \mathbf{E}[e^{itX}].$$

Notons que comme l'exponentielle complexe est de module 1, on a aucun problème d'intégrabilité donc ϕ_X est toujours définie. Elle trouve tout son intérêt dans le résultat suivant dont la preuve dépasse largement ce cours.

Théorème 3.9. Soit X et Y deux variables aléatoires réelles. On a

$$\text{Loi}(X) = \text{Loi}(Y) \iff \phi_X = \phi_Y.$$

Théorème 3.10. Si $X \in L^1(\mathbf{P})$ alors ϕ_X est dérivable sur $\mathbf{R}$ et

$$\phi'_X(t) = i \mathbf{E}[X e^{itX}], \text{ en particulier } \phi'_X(0) = \mathbf{E}[X].$$

Démonstration. On va pouvoir mesurer la puissance du théorème de dérivation sous le signe somme dans sa version Lebesgue. En notation intégrale,

$$\mathbf{E}[e^{itX}] = \int_{\mathbf{R}} e^{itx} d\mathbf{P}_X(x).$$

Comme la fonction $t \mapsto e^{itx}$ est dérivable partout, de dérivée ixe^{itx} , il nous suffit de trouver

une fonction de x , ne dépendant pas de t , qui majore en module cette dérivée. Or

$$|ixe^{itx}| \leq |x|$$

et par hypothèse,

$$\infty > \mathbf{E}[|X|] = \int_{\mathbf{R}} |x| \, d\mathbf{P}_X(x).$$

Ce qui signifie que toutes les hypothèses du théorème de dérivation sous le signe somme sont satisfaites donc ϕ_X est dérivable et

$$\phi'_X(t) = \int_{\mathbf{R}} ix e^{itx} \, d\mathbf{P}_X(x) = i\mathbf{E}[X e^{itX}].$$

D'où le résultat. □

Vecteurs aléatoires

Définition 4.1. Un vecteur aléatoire X (ou variable aléatoire vectorielle) est une application

$$X : \Omega \longrightarrow \mathbf{R}^n$$

$$\omega \longmapsto X(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

mesurable, c'est-à-dire telle que

$$\left(X \in]-\infty, x_1] \times \dots \times]-\infty, x_n] \right) = \bigcap_{k=1}^n \left(X_k \in]-\infty, x_k] \right) \in \mathcal{A},$$

pour tout n -uplet $(x_1, \dots, x_n)$.

Sa loi notée $\mathbf{P}_X$ ou $\mathbf{P}_{(X_1, \dots, X_n)}$, est la mesure image de $\mathbf{P}$ par X et sa fonction de répartition est donnée par

$$F_X(x_1, \dots, x_n) = \mathbf{P}(X \in]-\infty, x_1] \times \dots \times]-\infty, x_n])$$

$$= \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

La loi d'un vecteur à n composantes est une mesure sur $\mathbf{R}^n$. Cette loi est dite à densité lorsqu'il existe $f_X : \mathbf{R}^n \longrightarrow \mathbf{R}^+$ telle que pour toute h continue bornée de $\mathbf{R}^n$ dans $\mathbf{R}$,

$$\mathbf{E}[h(X_1, \dots, X_n)] = \int_{\mathbf{R}^n} h(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Définition 4.2. Les lois des X_i pour chaque $i \in \{1, \dots, n\}$ sont appelées les lois marginales.

REMARQUE.— Si l'on connaît la loi d'un vecteur aléatoire X à valeurs dans $\mathbf{R}^n$, on peut calculer toutes les lois marginales, car

$$\mathbf{P}(X_i \in]-\infty, b]) = \mathbf{P}(X_1 \leq +\infty, \dots, X_i \leq b, \dots, X_n \leq +\infty).$$

Dans le cas où le vecteur aléatoire a une densité, on retrouve la densité marginale de l'une des composantes en intégrant par rapport aux autres :

$$f_{X_k}(x_k) = \int f_{(X_1, \dots, X_p)}(x_1, \dots, x_p) dx_1 \dots \widehat{dx_k} \dots dx_p$$

où $\widehat{dx_k}$ signifie que l'on n'intègre pas par rapport à la variable x_k .

Réciproquement, on ne peut pas, sans hypothèse supplémentaire, déterminer la loi d'un vecteur à partir de la seule connaissance des marginales. Le **seul** cas où c'est possible est lorsque les composantes de X sont supposées être indépendantes.

Définition 4.3. Soient $(X_k)_{k \leq p}$ des vecteurs aléatoires à valeur dans $\mathbf{R}^{n_k}$, définis sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. Les vecteurs aléatoires $(X_k)_{k \leq p}$ sont indépendants ssi la loi de $(X_1, \dots, X_p)$ est le produit tendoriel des lois de chaque X_k :

$$\mathbf{P}_{(X_1, \dots, X_p)} = \mathbf{P}_{X_1} \otimes \dots \otimes \mathbf{P}_{X_p}.$$

La théorie de la mesure et de l'intégration permet de décrire cette propriété de l'une des trois façons équivalentes suivantes :

1. Pour tout $(a_k, b_k, k = 1, \dots, p)$,

$$\mathbf{P}\left(X_k \in]a_k, b_k[, k = 1, \dots, p\right) = \prod_{k=1}^p \mathbf{P}\left(X_k \in]a_k, b_k[\right)$$

2. Pour toute famille de boréliens $(A_k, k = 1, \dots, p)$,

$$\mathbf{P}\left(X_k \in A_k, k = 1, \dots, p\right) = \prod_{k=1}^p \mathbf{P}\left(X_k \in A_k\right)$$

3. Pour toute famille de fonctions $\psi_k : \mathbf{R}^{n_k} \rightarrow \mathbf{R}$ mesurable bornée,

$$\mathbf{E}\left[\prod_{k=1}^p \psi_k(X_k)\right] = \prod_{k=1}^p \mathbf{E}[\psi_k(X_k)]$$

REMARQUE.— Si les $(X_k)_{k \leq p}$ ont tous des lois à densité et sont indépendants alors la loi du vecteur $(X_1, \dots, X_p)$ a une densité donnée par

$$f_{(X_1, \dots, X_p)}(x_1, \dots, x_p) = \prod_{k=1}^p f_{X_k}(x_k).$$

4.1 Changement de variables

L'un des types de calcul qui revient régulièrement dans la pratique des probabilités est celui du calcul de la loi de la transformation d'un vecteur aléatoire de loi connue. L'outil principal pour ces calculs est la formule de changement de variables dans les intégrales multiples.

Définition 4.4. Soit $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ dont toutes les dérivées partielles existent

sur O , la jacobienne de T au point x , est la matrice $J_T(x)$ où

$$J_T(x) = \left(\frac{\partial T_i}{\partial x_j}(x), 1 \leq i \leq n, 1 \leq j \leq n \right)$$

$$= \begin{pmatrix} \frac{\partial T_1}{\partial x_1}(x) & \cdots & \frac{\partial T_1}{\partial x_j}(x) & \cdots & \frac{\partial T_1}{\partial x_n}(x) \\ \vdots & & \vdots & & \vdots \\ \frac{\partial T_i}{\partial x_1}(x) & \cdots & \frac{\partial T_i}{\partial x_j}(x) & \cdots & \frac{\partial T_i}{\partial x_n}(x) \\ \vdots & & \vdots & & \vdots \\ \frac{\partial T_n}{\partial x_1}(x) & \cdots & \frac{\partial T_n}{\partial x_j}(x) & \cdots & \frac{\partial T_n}{\partial x_n}(x) \end{pmatrix}.$$

Le jacobien de T est le déterminant de J_T .

Définition 4.5. Soit O un ouvert de $\mathbf{R}^n$, $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$, T est un $\mathcal{C}^1$ -difféomorphisme de O sur $\Delta \subset \mathbf{R}^n$, lorsque

- les dérivées partielles de T existent et sont continues sur O ,
- T est une bijection de O sur Δ ,
- le jacobien de T ne s'annule pas sur O .

Théorème 4.1. Soit O un ouvert de $\mathbf{R}^n$, $T : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ un $\mathcal{C}^1$ -difféomorphisme de T sur Δ . Pour toute fonction continue bornée ψ telle que l'une des intégrales suivantes soit bien définie,

$$\int_O f(T(x)) \, dx = \int_\Delta f(y) \frac{1}{|\det J_T(T^{-1}(y))|} \, dy.$$

Corollaire 4.2. on considère l'application

$$T : (\mathbf{R}^+)^* \times [0, 2\pi[\longrightarrow \mathbf{R}^2 \setminus \{0\}$$

$$(r, \theta) \longmapsto (x = r \cos \theta, y = r \sin \theta).$$

Pour $\psi \in L^1(\mathbf{R}^2)$,

$$\int_{\mathbf{R}^2} \psi(x, y) \, dx \, dy = \int_0^{+\infty} \int_0^{2\pi} (\psi \circ T)(r, \theta) r \, dr \, d\theta.$$

Démonstration. L'application T est bijective (vu que l'on a enlevé 0) et son jacobien est donné par

$$\det J_T(r, \theta) = \det \begin{pmatrix} \cos \theta & -\sin \theta \\ r \sin \theta & r \cos \theta \end{pmatrix} = r.$$

La formule de changement de variables donne

$$\begin{aligned} \int_{\mathbf{R}^2} \psi(x, y) \, dx \, dy &= \int_{\mathbf{R}^2} (\psi \circ T) \circ T^{-1}(x, y) \, dx \, dy \\ &= \int_0^{+\infty} \int_0^{2\pi} (\psi \circ T)(r, \theta) \left| \frac{1}{\det J_{T^{-1}}(T(r, \theta))} \right| \, dr \, d\theta. \end{aligned}$$

Or $T \circ T^{-1} = \text{I}$ donc

$$\det(J_T \circ T^{-1}) \det J_{T^{-1}} = 1.$$

Par conséquent,

$$\frac{1}{\det J_{T^{-1}}(T(r, \theta))} = \det J_T(r, \theta) = r.$$

Le résultat s'ensuit. □

EXEMPLE 15.— Soit (X_1, X_2) deux variables aléatoires réelles indépendantes, de même loi

$$d\mathbf{P}(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} dx.$$

On pose $U = X_1 \cdot X_2$ et $V = X_1/X_2$.

1. Calculer la loi du vecteur (U, V) .
2. Calculer la loi de U et celle de V .
3. U et V sont-elles indépendantes ?

On part de l'hypothèse que la loi du couple (U, V) a une densité par rapport à la mesure de Lebesgue, ce qui en utilisant la caractérisation des mesures induites par le théorème de Riesz 1.38, revient à trouver $h: \mathbf{R}^2 \rightarrow \mathbf{R}^+$ telle que pour toute fonction f continue bornée de $\mathbf{R}^2$ dans $\mathbf{R}$, on ait

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y) h(x, y) dx dy.$$

Posons

$$\begin{aligned} T : \mathbf{R}^2 &\rightarrow \mathbf{R}^2 \\ (x, y) &\mapsto (xy, x/y). \end{aligned}$$

On a

$$\mathbf{E}[f(U, V)] = \mathbf{E}[(f \circ T)(X, Y)] = \int (f \circ T)(x, y) d\mathbf{P}_{X, Y}(x, y),$$

où la deuxième égalité découle du théorème de transfert. Maintenant, les v.a. X et Y sont indépendantes, ce qui équivaut à dire que

$$\mathbf{P}_{X, Y} = \mathbf{P}_X \otimes \mathbf{P}_Y.$$

Par hypothèse,

$$d\mathbf{P}_X(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} dx \text{ et } d\mathbf{P}_Y(y) = \mathbf{1}_{[1, \infty[}(y) \frac{1}{y^2} dy,$$

donc

$$d\mathbf{P}_{X,Y}(x, y) = \mathbf{1}_{[1,\infty[}(x) \frac{1}{x^2} \mathbf{1}_{[1,\infty[}(y) dx dy.$$

On a donc obtenu

$$\mathbf{E}[f(U, V)] = \int_{[1, +\infty[^2} (f \circ T)(x, y) \frac{1}{x^2 y^2} dx dy.$$

Rappelons-nous que nous voulons aboutir à une identité de la forme

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y) h(x, y) dx dy.$$

On est naturellement enclin à utiliser le théorème de changement de variables 4.1, pour cela, il nous faut calculer Δ , l'ensemble image de $[1, +\infty[^2$ par T et le jacobien de T . Posons $u = xy$ et $v = x/y$,

$$\det J_T(x, y) = \det \begin{pmatrix} y & x \\ \frac{1}{y} & -\frac{x}{y^2} \end{pmatrix} = -2 \frac{x}{y} = -2v.$$

Si x et y sont tous deux plus grands que 1 alors u l'est, et v est strictement positif. Par ailleurs,

$$\begin{cases} u = xy \\ v = x/y \end{cases} \iff \begin{cases} x^2 = uv \\ y^2 = u/v \end{cases}.$$

On déduit de ces dernières équations que $u \geq v$ et $uv \geq 1$. On vérifie alors facilement que T est une bijection de $[1, +\infty[^2$ sur

$$\Delta = \{(u, v), u \geq v \geq 0 \text{ et } uv \geq 1\}.$$

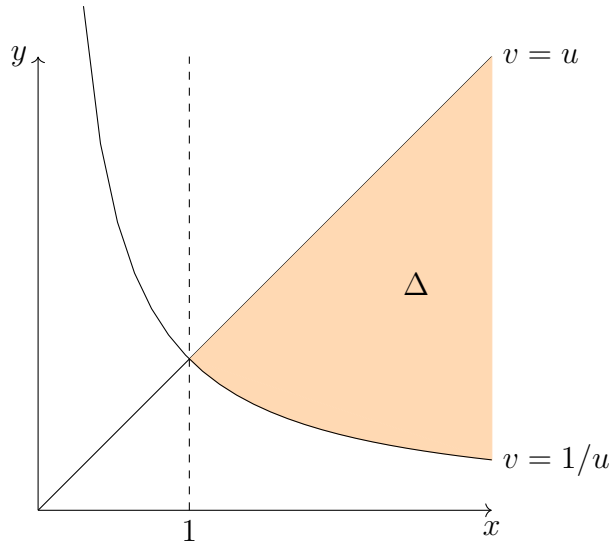


FIGURE 4.1 – Le domaine Δ dans le plan (u, v) .

On tire du théorème de changement de variables que

$$\begin{aligned} \int_{[1, +\infty[^2} (f \circ T)(x, y) \frac{1}{x^2 y^2} dx dy &= \int_{\Delta} f(u, v) \frac{1}{uv \cdot u/v} \left| \frac{1}{-2v} \right| du dv \\ &= \int_{\Delta} f(u, v) \frac{1}{2u^2 v} du dv, \end{aligned}$$

d'où par identification,

$$d\mathbf{P}_{(U, V)}(u, v) = \frac{1}{2u^2 v} \mathbf{1}_{\Delta}(u, v) du dv.$$

Pour calculer la loi de U , on veut exprimer $\mathbf{E}[f(U)]$ pour toute fonction continue bornée de $\mathbf{R}$ dans $\mathbf{R}$. On remarque alors que l'application $\tilde{f}(x, y) = f(x)$ est continue bornée de $\mathbf{R}^2$ dans $\mathbf{R}$ donc

$$\begin{aligned} \mathbf{E}[f(U)] &= \mathbf{E}[\tilde{f}(U, V)] = \int_{\Delta} f(u) \frac{1}{2u^2 v} du dv \\ &= \int_1^{+\infty} f(u) \frac{1}{2u^2} \left(\int_{1/u}^u \frac{1}{v} dv \right) du, \end{aligned}$$

d'après le théorème de Fubini 1.35. Par conséquent,

$$\begin{aligned} d\mathbf{P}_U(u) &= \frac{1}{2u^2} \left(\int_{1/u}^u \frac{1}{v} dv \right) \mathbf{1}_{[1, +\infty[}(u) du \\ &= \frac{\ln u}{u^2} \mathbf{1}_{[1, +\infty[}(u) du. \end{aligned}$$

De même,

$$\begin{aligned} \mathbf{E}[f(V)] &= \int_{\Delta} f(v) \frac{1}{2u^2 v} du dv \\ &= \int_0^{+\infty} f(v) \frac{1}{2v} \left(\int_{\Delta_v} \frac{1}{u^2} du \right) dv, \end{aligned}$$

où

$$\begin{aligned} \Delta_v &= \{u \mid (u, v) \in \Delta\} \\ &= \begin{cases} [1/v, +\infty[& \text{si } 0 \leq v \leq 1 \\ [v, +\infty[& \text{si } v \geq 1. \end{cases} \end{aligned}$$

Par conséquent,

$$\begin{aligned} d\mathbf{P}_V(v) &= \frac{1}{2v} \left(\int_{1/v}^{+\infty} \frac{1}{u^2} du \mathbf{1}_{[0, 1]}(v) + \int_v^{+\infty} \frac{1}{u^2} du \mathbf{1}_{]v, +\infty[}(v) \right) \\ &= \frac{1}{2v} \left(v \mathbf{1}_{[0, 1]}(v) + \frac{1}{v} \mathbf{1}_{]1, +\infty[}(v) \right). \end{aligned}$$

Comme

$$d\mathbf{P}_{(U, V)} \neq d\mathbf{P}_U \otimes d\mathbf{P}_V,$$

les v.a. U et V ne sont pas indépendantes.

4.2 Formulaire

Distribution	Espérance	Variance	Loi
$B(n, p)$	np	$np(1 - p)$	$\sum_{j=0}^n \binom{n}{j} p^j (1 - p)^{n-j} \delta_j$
$\text{Geom}(p)$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$	$\sum_{j=1}^{+\infty} (1 - p)^{j-1} p \delta_j$
$P(\lambda)$	λ	λ	$\sum_{j=0}^{+\infty} e^{-\lambda} \frac{\lambda^j}{j!} \delta_j$
$\mathcal{U}([a, b])$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$	$\frac{1}{b - a} \mathbf{1}_{[a, b]}(x) dx$
$\mathcal{E}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbf{R}^+}(x) dx$
$\mathcal{N}(m, \sigma^2)$	m	σ^2	$\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$
$\mathcal{C}(c)$	non définie	non définie	$\frac{c}{\pi(c^2 + x^2)} dx$

Fonction caractéristique

On note $\langle a, b \rangle$ le produit scalaire de deux vecteurs a, b de $\mathbf{R}^d$, et $\|a\|$ la norme euclidienne de a . A^T désigne la transposée de la matrice A . Par convention, les vecteurs sont des vecteurs-colonne.

Pour un nombre complexe x , $|x|$ désigne son module et $\bar{x}$ son conjugué. On note $i = \sqrt{-1}$.

Dans ce chapitre, nous sommes amenés à utiliser l'intégrale, par rapport à des mesures de probabilité, de fonctions à valeur dans $\mathbf{C}$. Toute fonction f à valeur dans $\mathbf{C}$ peut s'écrire sous la forme :

$$f = f_R + if_I \quad (5.1)$$

où f_R, f_I sont les fonctions à valeur dans $\mathbf{R}$ désignant respectivement la partie réelle et la partie imaginaire. On dira qu'une fonction f à valeur dans $\mathbf{C}$ est *mesurable* si les fonctions f_R et f_I sont mesurables. Elle est par définition *intégrable par rapport à une mesure μ* si f_R et f_I sont intégrables. On définit alors l'intégrale de f par :

$$\int f \, d\mu := \int f_R \, d\mu + i \int f_I \, d\mu .$$

À l'aide de la décomposition (5.1) et des inégalités :

$$|f_R| \leq |f|, \quad |f_I| \leq |f|, \quad |f| \leq |f_R| + |f_I| ,$$

il est aisé de vérifier que des propriétés établies pour des fonctions à valeur dans $\mathbf{R}$ restent valables pour des fonctions à valeur dans $\mathbf{C}$. Par exemple :

- une fonction mesurable f à valeur dans $\mathbf{C}$ est intégrable par rapport à μ si et seulement

$$\int |f| \, d\mu < \infty ;$$

- le théorème de convergence dominée (voir théorème 1.33) reste vrai si les fonctions f, f_n sont supposées à valeur dans $\mathbf{C}$.

5.1 Définition et propriétés

Dans cette section, $X = (X_1, \dots, X_d)$ désigne un vecteur aléatoire à valeur dans $\mathbf{R}^d$, défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$.

Définition 5.1. On appelle *fonction caractéristique* de la variable aléatoire X la fonction $\phi_X : \mathbf{R}^d \rightarrow \mathbf{C}$ définie par

$$\begin{aligned}\phi_X(t) &:= \mathbf{E} \left[\exp(i\langle t, X \rangle) \right] = \int_{\mathbf{R}^d} \exp(i\langle t, x \rangle) d\mathbf{P}_X(x) \\ &= \mathbf{E} \left[\exp\left(i \sum_{k=1}^d t_k X_k\right) \right] = \mathbf{E} \left[\prod_{k=1}^d \exp(it_k X_k) \right] .\end{aligned}$$

Comme $|\exp(i\langle t, x \rangle)| = 1$, la fonction ϕ_X est bien définie sur $\mathbf{R}^d$. Elle est définie pour les variable aléatoire discrètes comme les variable aléatoire à valeur réelle ou vectorielle. Dans le cas particulier où X est à valeurs dans un espace au plus dénombrable E , on a

$$\phi_X(t) = \sum_{e \in E} \exp(i\langle t, e \rangle) \mathbf{P}[X = e] ;$$

dans le cas où X est une variable aléatoire à densité f_X (par rapport à la mesure de Lebesgue λ_d sur $\mathbf{R}^d$), on a

$$\phi_X(t) = \int \exp(i\langle t, x \rangle) f_X(x) dx .$$

La définition de la fonction caractéristique ne dépend que de la loi de la variable aléatoire X . En particulier, si deux variable aléatoire ont même loi alors elles ont même fonction caractéristique.

Démonstration. En conséquence du théorème de transfert (théorème 3.1), on a $\mathbf{E}(g(X)) = \mathbf{E}(g(Y))$ pour toute fonction mesurable bornée à valeur dans $\mathbf{R}$ (et donc par une extension triviale, à valeur dans $\mathbf{C}$). En particulier, en appliquant cette propriété à la fonction $g(x) = \exp(i\langle t, x \rangle)$ (t étant fixé), on en déduit que $\phi_X(t) = \phi_Y(t)$. $\square$

Proposition 5.1. La fonction caractéristique d'une variable aléatoire X vérifie les propriétés suivantes :

- a) Pour tout $t \in \mathbf{R}^d$, $|\phi_X(t)| \leq 1$. En outre, $\phi_X(0) = 1$.
- b) La fonction ϕ_X est continue sur $\mathbf{R}^d$.
- c) Si $b \in \mathbf{R}^p$ est un vecteur déterministe et A est une matrice déterministe de taille $p \times d$ alors $AX + b \in \mathbf{R}^p$ et pour tout $t \in \mathbf{R}^p$,

$$\phi_{AX+b}(t) = \exp(i\langle t, b \rangle) \phi_{AX}(t) = \exp(i\langle t, b \rangle) \phi_X(A^T t) .$$

En particulier, si X est une variable aléatoire à valeur dans $\mathbf{R}$ et que a, b sont des réels alors

$$\phi_{aX+b}(t) = \exp(itb) \phi_X(at) .$$

- d) Si X, Y sont deux v.a. indépendantes définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et à valeurs dans $\mathbf{R}^d$, alors $\phi_{X+Y} = \phi_X \phi_Y$. Plus généralement, si $X_1, \dots, X_n : \Omega \rightarrow \mathbf{R}^d$ sont des variable aléatoire indépendantes, alors :

$$\phi_{X_1 + \dots + X_n} = \prod_{k=1}^n \phi_{X_k}$$

Démonstration. **a))** On a $|\phi_X(t)| \leq \mathbf{E}(|\exp(i\langle t, X \rangle)|) = 1$. De plus, $\phi_X(0) = \mathbf{E}(\exp(0)) = 1$.

b)) La continuité en tout point $t \in \mathbf{R}^d$ est une conséquence du théorème de convergence dominée (voir théorème 1.33). Soit $t \in \mathbf{R}^d$; nous écrivons pour toute suite $(h_\ell)_\ell$ à valeur dans $\mathbf{R}^d$ telle que $\lim_{\ell \rightarrow +\infty} h_\ell = 0$,

$$\phi_X(t + h_\ell) - \phi_X(t) = \int (\exp(i\langle t + h_\ell, x \rangle) - \exp(i\langle t, x \rangle)) \, d\mathbf{P}_X(x) .$$

Par continuité du produit scalaire et de la fonction exponentielle,

$$\lim_{\ell \rightarrow +\infty} (\exp(i\langle t + h_\ell, x \rangle) - \exp(i\langle t, x \rangle)) = 0 .$$

De plus, pour tout $h \in \mathbf{R}^d$, $|\exp(i\langle t + h, x \rangle) - \exp(i\langle t, x \rangle)| \leq 2$ et $\int 2 \, d\mathbf{P}_X = 2 < \infty$. Par suite, le théorème de convergence dominée entraîne $\lim_{\ell \rightarrow +\infty} \phi_X(t + h_\ell) = \phi_X(t)$, ce qui établit la continuité en t pour tout $t \in \mathbf{R}^d$.

c)) On écrit en utilisant les propriétés du produit scalaire et de la fonction exponentielle, et en utilisant le fait que A, b sont déterministes

$$\begin{aligned} \phi_{AX+b}(t) &= \mathbf{E}(\exp(i\langle AX + b, t \rangle)) = \exp(i\langle b, t \rangle) \mathbf{E}(\exp(i\langle AX, t \rangle)) \\ &= \exp(i\langle b, t \rangle) \mathbf{E}(\exp(i\langle X, A^T t \rangle)) = \exp(i\langle b, t \rangle) \phi_X(A^T t) . \end{aligned}$$

d)) On a pour tout $t \in \mathbf{R}^d$,

$$\begin{aligned} \phi_{X+Y}(t) &= \mathbf{E}(\exp(i\langle t, X \rangle) \exp(i\langle t, Y \rangle)) = \mathbf{E}(\exp(i\langle t, X \rangle)) \mathbf{E}(\exp(i\langle t, Y \rangle)) \\ &= \phi_X(t) \phi_Y(t) \end{aligned}$$

en utilisant la caractérisation de l'indépendance donnée par le théorème ?? (établie pour des fonctions à valeur réelle et donc valable aussi pour des fonctions à valeur dans $\mathbf{C}$). La généralisation à une somme de n variable aléatoire indépendantes est immédiate.

□

5.2 Fonctions caractéristiques de variable aléatoire usuelles

La table 5.1 fournit l'expression des fonctions caractéristiques de plusieurs loi usuelles. Les sept premières expressions sont la conséquence de calculs triviaux que nous omettons ici. Nous fournissons ci-dessous la preuve des trois dernières.

Nous commençons par établir l'expression de la fonction caractéristique d'une variable aléatoire normale $\mathcal{N}(0, 1)$; nous en déduirons celle d'une loi $\mathcal{N}_d(0, I)$ puis d'une loi

Loi	Expression de $\Phi_X(t)$
Mesure de Dirac δ_a	$\exp(ita)$
Bernoulli $\mathcal{B}(p)$	$1 - p + p \exp(it)$
Binomiale $\mathcal{B}(n, p)$	$(1 - p + p \exp(it))^n$
Géométrique $\mathcal{G}(p)$ sur $\mathbf{N}$	$\frac{p}{1 - (1 - p) \exp(it)}$
Poisson $\mathcal{P}(\lambda)$	$e^{\lambda(e^{it}-1)}$
Uniforme $\mathcal{U}([a, b])$	$\frac{\exp(itb) - \exp(ita)}{it(b - a)}$
Exponentielle $\mathcal{E}(\lambda)$	$\frac{\lambda}{\lambda - it}$
Gaussienne réelle $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma^2 \geq 0$	$\exp(it\mu - 1/2 \sigma^2 t^2)$
Gaussienne multivariée $\mathcal{N}_n(\mu, \Gamma)$	$\exp(i\langle t, \mu \rangle - 1/2 t^T \Gamma t)$
Gamma $\Gamma(a, b)$	$\left(\frac{b}{b - it} \right)^a$

TABLE 5.1 – Quelques fonctions caractéristiques utiles. Se reporter aux tables 1.1 et ?? pour la définition des lois. Pour la fonction caractéristique d'une loi $\mathcal{N}_n(\mu, \Gamma)$ lorsque Γ est définie positive, voir section 5.2; lorsque Γ est positive, voir chapitre ??. Par convention, $\mathcal{N}(a, 0)$ est la mesure de Dirac en a

$\mathcal{N}_d(\mu, \Gamma)$. Soit $Y \sim \mathcal{N}(0, 1)$. Les conditions de dérivation sous l'intégrale sont vérifiées (voir exercice ??) et on a

$$\begin{aligned}\phi_Y'(t) &= \frac{1}{\sqrt{2\pi}} \int i x \exp(itx) \exp(-1/2 x^2) dx \\ &= -\frac{i}{\sqrt{2\pi}} \int (it - x) \exp(itx) \exp(-1/2 x^2) dx - \frac{t}{\sqrt{2\pi}} \int \exp(itx) \exp(-1/2 x^2) dx \\ &= -\frac{i}{\sqrt{2\pi}} \left[\exp(itx) \exp(-1/2 x^2) \right]_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}} \int \exp(itx) \exp(-1/2 x^2) dx = -t \phi_Y(t).\end{aligned}$$

La résolution de l'équation différentielle $\phi_Y'(t) = -t \phi_Y(t)$ sachant que l'on doit avoir $\phi_Y(0) = 1$, donne le résultat.

Soit $Z \sim \mathcal{N}_d(0, I)$. Alors par le théorème de Fubini (voir Chapitre 1),

$$\begin{aligned}\phi_Z(t) &= \frac{1}{\sqrt{2\pi}^d} \int_{\mathbf{R}^d} \exp(i\langle t, z \rangle) \exp(-1/2 \sum_{k=1}^d z_k^2) dz_1 \cdots dz_d \\ &= \prod_{k=1}^d \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \exp(it_k z_k) \exp(-1/2 z_k^2) dz_k \right) = \prod_{k=1}^d \phi_Y(t_k) \\ &= \prod_{k=1}^d \exp(-1/2 t_k^2) = \exp(-1/2 t^T t).\end{aligned}$$

Enfin, nous savons que si $Z \sim \mathcal{N}_d(0, I)$ alors $X = \mu + \sqrt{\Gamma}Z$ suit une loi $\mathcal{N}_d(\mu, \Gamma)$. L'expression de la fonction caractéristique d'une loi $\mathcal{N}_d(\mu, \Gamma)$ est maintenant la conséquence de la Proposition 5.1-c).

Soit $X \sim \Gamma(a, b)$. Par application du théorème de dérivation sous l'intégrale suivie d'une intégration par parties, nous avons $\phi_X'(t) = \frac{ia}{b-it} \phi_X(t)$ dont nous déduisons l'expression de ϕ_X en utilisant la condition $\phi_X(0) = 1$. A noter que si a n'est pas un entier, on prend la détermination continue valant 1 en 0.

5.3 Caractérisation de la loi

Nous avons établi (théorème 1.38) que deux lois μ, ν sont égales si et seulement si $\mathbf{E}(f(X)) = \mathbf{E}(f(Y))$ pour toute fonction continue bornée (ou pour toute fonction continue à support compact); ici $X \sim \mu$ et $Y \sim \nu$. Le théorème suivant donne une autre caractérisation : μ, ν sont égales si et seulement si $\mathbf{E}(f(X)) = \mathbf{E}(f(Y))$ pour toute fonction f de la forme $x \mapsto \exp(i\langle t, x \rangle)$, $t \in \mathbf{R}^n$. La preuve de ce résultat repose sur le lemme suivant.

Théorème 5.2. La fonction caractéristique d'une variable aléatoire détermine sa loi *i.e.*, pour deux vecteurs aléatoires X, Y à valeur dans $\mathbf{R}^d$ on a équivalence :

a) pour tout $t \in \mathbf{R}^d$, $\phi_X(t) = \phi_Y(t)$.

b) les vecteurs aléatoires X et Y ont même loi.

Démonstration. Le sens réciproque est trivial étant donnée la définition de la fonction caractéristique : on ne se préoccupe que du sens direct. Plaçons nous pour simplifier dans le cas $d = 1$. A titre de remarque, signalons que la preuve est déjà connue des élèves dans le cas particulier où X et Y sont deux lois à densité f_X et f_Y : dans le cas où les fonctions caractéristiques sont elles-mêmes intégrables, les densités se déduisent des fonctions caractéristiques par transformée de Fourier de ϕ_X et ϕ_Y . Ainsi, $\phi_X = \phi_Y$ implique que $f_X = f_Y$ presque partout, et donc que $\mathbf{P}_X = \mathbf{P}_Y$.

Dans le cas général, la preuve repose sur la formule d'inversion :

$$\mathbf{P}_X([a, b]) = \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt, \quad (5.2)$$

qui vraie pour tous a, b en lesquels $\mathbf{P}_X$ n'a pas de masse, c'est-à-dire pour tous a, b hors de l'ensemble $D_X := \{x : \mathbf{P}_X(\{x\}) > 0\}$. Cet ensemble D_X est au plus dénombrable. Donc pour tout a et b hors de $D_X \cup D_Y$, on a $\mathbf{P}_X([a, b]) = \mathbf{P}_Y([a, b])$. Puisque $F_X(b) = F_X(a) + \mathbf{P}_X([a, b])$, il suffit de faire tendre a vers $-\infty$ pour obtenir :

$$F_X(b) = F_Y(b) \quad (5.3)$$

pour tout point b hors d'un ensemble au plus dénombrable. En utilisant la continuité à droite des fonctions de répartition, on conclut que (5.3) est vraie en tout point. Puisque la fonction de répartition détermine entièrement la loi, le résultat est démontré. $\square$

5.4 Caractérisation de l'indépendance

Soient $(X_k)_{k \leq p}$ des vecteurs aléatoires à valeur dans $\mathbf{R}^{n_k}$, définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. On peut maintenant compléter la définition 4.3

Théorème 5.3. Les variables $(X_k)_{k \leq p}$ sont indépendants ssi pour tout $(t_1, \dots, t_p)$ avec $t_k \in \mathbf{R}^{n_k}$,

$$\mathbf{E} \left[\prod_{k=1}^p e^{i \langle t_k, X_k \rangle_{\mathbf{R}^{n_k}}} \right] = \prod_{k=1}^p \mathbf{E} \left[e^{i \langle t_k, X_k \rangle_{\mathbf{R}^{n_k}}} \right]$$

5.5 Calcul de moments

5.5.1 Moments et fonction caractéristique

Puisque la fonction caractéristique « caractérise la loi », elle détermine également les moments de cette loi. Il se trouve que la fonction caractéristique est un outil souvent

commode pour évaluer les moments d'une loi.

Théorème 5.4. Soit X une variable aléatoire réelle possédant un moment d'ordre p ($p > 0$). Alors ϕ_X est de classe C^p et l'on a

$$\frac{\partial^p \phi_X(t)}{\partial t^p} = i^p \mathbf{E}(X^p \exp(itX)) .$$

En particulier, les moments sont liés aux dérivées en zero :

$$\mathbf{E}(X^p) = (-1)^p i^p \frac{\partial^p \phi_X(0)}{\partial t^p} .$$

Démonstration. Nous établissons le résultat suivant :

$$\phi_X(t) = 1 + it\mathbf{E}(X) - \frac{t^2}{2}\mathbf{E}(X^2) + \cdots + i^p \frac{t^p}{p!}\mathbf{E}(X^p) + \xi(t) \quad \text{où } \lim_{t \rightarrow 0} \xi(t)/t^p = 0.$$

Cette égalité se justifie par le développement :

$$\begin{aligned} \exp(itx) &= \sum_{k=0}^p i^k \frac{t^k}{k!} x^k \\ &= \sum_{k \geq p} i^k \frac{t^k}{(k-p)!} x^k \frac{1}{(k-p+1) \cdots (k-1)k} \\ &= \sum_{k \geq p} i^k \frac{t^k}{(k-p)!} x^k \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} u_1^{k-p} du_1 du_2 \cdots du_p \\ &= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} \sum_{k \geq p} i^{k-p} \frac{t^{k-p}}{(k-p)!} x^{k-p} u_1^{k-p} du_1 du_2 \cdots du_p \\ &= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} \left(\exp(iu_1 t x) - 1 \right) du_1 du_2 \cdots du_p , \end{aligned}$$

où l'on a utilisé le théorème Fubini. On montre que l'espérance de ce dernier terme est $o(t^p)$ par application du théorème de convergence dominée. $\square$

Généralisation aux vecteurs aléatoires

De même, on peut montrer le résultat suivant :

Théorème 5.5. Soit X un vecteur aléatoire admettant un moment d'ordre p ($p > 0$). $t \mapsto \phi_X(t)$ est de classe C^p et on a

$$\frac{\partial^p \phi_X(t)}{\partial t_1 \cdots \partial t_p} = i^p \mathbf{E} \left(X_{t_1} X_{t_2} \cdots X_{t_p} \exp(i \langle t, X \rangle) \right) .$$

On en déduit aussi une méthode pour le calcul de moments à partir de l'expression de la fonction caractéristique.

5.5.2 Applications

Application 1

Nous montrons que si X est une variable aléatoire réelle gaussienne d'espérance μ et de variance σ^2 ($\sigma^2 > 0$) *i.e.*, $X \sim \mathcal{N}(\mu, \sigma^2)$ alors tous les moments impairs de $X - \mu$ sont nuls et les moments pairs s'expriment à l'aide de σ^2 : pour tout $q \in \mathbf{N}$,

$$\mathbf{E}((X - \mu)^{2q+1}) = 0, \quad (5.4)$$

$$\mathbf{E}((X - \mu)^{2q}) = \sigma^{2q} (2q - 1)(2q - 3) \cdots 3 = \sigma^{2q} \frac{(2q)!}{2^q q!}. \quad (5.5)$$

Démonstration. On veut calculer $\mathbf{E}((X - \mu)^q)$ pour tout $q \in \mathbf{N}$. Nous avons établi (exercice ??) que si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\sigma^{-1}(X - \mu) \sim \mathcal{N}(0, 1)$. Par suite, nous allons établir

$$\mathbf{E}(Y^{2q+1}) = 0, \quad \mathbf{E}(Y^{2q}) = (2q - 1)(2q - 3) \cdots 3 = \frac{(2q)!}{2^q q!}, \quad (5.6)$$

où $Y \sim \mathcal{N}(0, 1)$. On écrit

$$\mathbf{E}(\exp(itY)) = \mathbf{E}\left(\sum_{n \geq 0} \frac{(itY)^n}{n!}\right) = \mathbf{E}\left(\lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} Y^k\right).$$

Pour permuter limite et espérance, on applique le théorème de convergence dominée (théorème 1.33) en remarquant que $|\sum_{k=0}^N \frac{(it)^k}{k!} Y^k| \leq \exp(|t||Y|)$ et que l'espérance de ce majorant est finie. Par suite,

$$\mathbf{E}(\exp(itY)) = \lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} \mathbf{E}(Y^k) = \sum_{k \geq 0} \frac{(it)^k}{k!} \mathbf{E}(Y^k).$$

D'autre part, d'après le Tableau 5.1

$$\mathbf{E}(\exp(itY)) = \exp(-1/2 t^2) = \sum_{n \geq 0} \frac{(-1)^n t^{2n}}{2^n n!}.$$

On en déduit (5.6) par identification. □

Par convention, une loi gaussienne d'espérance μ et de variance nulle est une masse de Dirac en μ ($\mu \in \mathbf{R}$). Si $X \sim \delta_\mu$ alors tous ses moments centrés sont nuls et les égalités ci-dessus restent vraies.

Ainsi, on a établi que si $X \sim \mathcal{N}(\mu, \sigma^2)$, avec $\sigma^2 \geq 0$, on a (5.4) et (5.5).

Application 2

Soit $m \in \mathbf{R}^d$ un vecteur déterministe et Γ une matrice $d \times d$ positive déterministe. Montrons

a) que la fonction

$$t \mapsto \exp(i\langle t, m \rangle - 1/2 t^T \Gamma t) \quad (5.7)$$

est la fonction caractéristique d'un vecteur aléatoire,

b) et que ce vecteur aléatoire a pour espérance m et pour matrice de covariance Γ .

Démonstration. (a) Nous avons établi que lorsque Γ est définie positive, la fonction donnée par (5.7) est la transformée de Fourier d'une loi $\mathcal{N}_d(m, \Gamma)$ (voir Tableau 5.1). L'exercice suivant justifie le fait que lorsque Γ est juste supposée positive, la fonction définie par (5.7) peut encore être lue comme la transformée de Fourier d'une loi.

(b) Nous montrons l'expression de l'espérance et de la matrice de covariance par application des résultats du Théorème 5.5.1.

Soit $k \in \{1, \dots, n\}$. En dérivant $t \mapsto \phi_X(t)$ par rapport à la k -ième composante t_k , on sait d'une part que

$$\partial_{t_k} \phi_X|_{t=0} = i\mathbf{E}(X_k)$$

et d'autre part,

$$\partial_{t_k} \phi_X|_{t=0} = \partial_{t_k} \left(\exp(it^T m - 1/2 t^T \Gamma t) \right) |_{t=0} = i m_k .$$

Ainsi, $\mathbf{E}(X_k) = m_k$ pour tout $k \in \{1, \dots, n\}$.

On considère la dérivée partielle d'ordre 2 : $\partial_{t_j t_k} \phi_X(t)$. D'une part, on sait que

$$\partial_{t_j t_k} \phi_X|_{t=0} = -\mathbf{E}(X_j X_k) .$$

D'autre part,

$$\partial_{t_j t_k} \phi_X|_{t=0} = \partial_{t_j t_k} \left(\exp(it^T m - 1/2 t^T \Gamma t) \right) |_{t=0} = -\Gamma_{j,k} - m_j m_k .$$

Par suite,

$$\Gamma_{j,k} = \mathbf{E}(X_j X_k) - m_j m_k = \mathbf{E}(X_j X_k) - \mathbf{E}(X_j) \mathbf{E}(X_k) = \text{Cov}(X_j, X_k) .$$

Ainsi, Γ est la matrice de covariance du vecteur aléatoire X . □

Vecteurs gaussiens

On note $\langle a, b \rangle$ le produit scalaire de deux vecteurs a, b de $\mathbf{R}^d$; et on note M^T la transposée de la matrice M . Par convention, les vecteurs sont des vecteurs-colonne. Rappelons que la transposée d'une matrice satisfait la relation :

$$\langle Ma, b \rangle = \sum_{i=1}^d \left(\sum_{j=1}^d M_{i,j} \right) a_j b_i = \langle a, M^T b \rangle.$$

Définition 6.1. La loi gaussienne (ou normale) de paramètres m, σ^2 ($\sigma \geq 0$, $m \in \mathbf{R}$) - notée $\mathcal{N}(m, \sigma^2)$ - est définie comme étant :

si $\sigma > 0$: la loi de densité par rapport à la mesure de Lebesgue donnée par

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp \left(-\frac{1}{2} \frac{(x - m)^2}{\sigma^2} \right). \quad (6.1)$$

si $\sigma = 0$: la mesure de Dirac en m .

Dans tous les cas, la fonction caractéristique d'une loi gaussienne $\mathcal{N}(m, \sigma^2)$ est donnée par

$$t \mapsto \exp(itm - \frac{1}{2} t^2 \sigma^2). \quad (6.2)$$

6.1 Définition et propriétés

Définition 6.2. X est un *vecteur gaussien* (ou *variable gaussienne multivariée* ou *variable normale multivariée*) si et seulement si pour tout $a \in \mathbf{R}^d$, la loi de $\langle a, X \rangle$ est une loi gaussienne (éventuellement de variance nulle), i.e. toute combinaison linéaire des composantes de X est une gaussienne.

REMARQUE.— On remarque que par définition toutes les composantes de X sont nécessairement des variables aléatoires gaussiennes (prendre $a = e_i$ le i -ième vecteur de la base canonique de $\mathbf{R}^d$).

On a déjà vu que la somme de gaussiennes indépendantes suit une loi gaussienne, donc un vecteur composé de gaussiennes indépendantes est un vecteur gaussien.

L'un des objectifs de ce chapitre est de voir qu'en revanche, un vecteur de gaussiennes (sans hypothèse d'indépendance) n'est pas nécessairement un vecteur gaussien.

L'autre est de voir quelles sont les transformations qui permettent à un vecteur aléatoire de rester un vecteur gaussien.

De la définition 6.1, on peut déduire la forme générale de la fonction caractéristique d'un vecteur gaussien.

Définition 6.3. Pour un vecteur $X : \Omega \rightarrow \mathbf{R}^d$, on note

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

$$\Gamma_X = (\text{Cov}(X_i, X_j), 1 \leq i, j \leq d)$$

où

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

Théorème 6.1. Soit X un vecteur gaussien dans $\mathbf{R}^d$. Pour tout $t \in \mathbf{R}^d$,

$$\mathbf{E}[e^{i\langle t, X \rangle}] = \exp\left(i\langle t, \mathbf{E}[X] \rangle - \frac{1}{2} \langle \Gamma_X t, t \rangle\right). \quad (6.3)$$

6.1.1 Fonction caractéristique

Nous avons vu au chapitre 5 que la fonction caractéristique déterminait la loi de X . Le théorème suivant peut être lu comme une alternative à la définition de vecteur gaussien.

Théorème 6.2. Les deux conditions sont équivalentes

- a) X est un vecteur gaussien d'espérance m et de matrice de covariance Γ .
- b) la fonction caractéristique du vecteur aléatoire X est $t \mapsto \exp(i\langle t, m \rangle - \frac{1}{2} t^T \Gamma t)$.

Dans ce cas, on écrira $X \sim \mathcal{N}_d(m, \Gamma)$.

Démonstration. Supposons a)). Alors pour tout $t \in \mathbf{R}^d$, $\langle t, X \rangle$ est une v.a.r. gaussienne d'espérance $\langle t, m \rangle$ et de variance $t^T \Gamma t$. On en déduit l'expression de la fonction caractéristique en appliquant le formulaire, tableau 5.1.

Réciproquement, supposons b)). Identifions la loi de $\langle t, X \rangle$, pour tout $t \in \mathbf{R}^d$, en calculant

la fonction caractéristique de cette v.a. à valeur dans $\mathbf{R}$. Soit $y \in \mathbf{R}$:

$$\begin{aligned}\phi_{\langle t, X \rangle}(y) &= \mathbf{E}[\exp(iy \langle t, X \rangle)] \\ &= \mathbf{E}[\exp(i \langle yt, X \rangle)] = \phi_X(yt) = \exp(i \langle yt, m \rangle - \frac{1}{2} (yt)^T \Gamma (yt)) \\ &= \exp(iy \langle t, m \rangle - \frac{1}{2} y^2 (t^T \Gamma t)) .\end{aligned}$$

On reconnaît la fonction caractéristique d'une loi gaussienne réelle d'espérance $\langle t, m \rangle$ et de variance (éventuellement nulle) $t^T \Gamma t$ (voir le formulaire Tableau 5.1). Donc X est un vecteur gaussien ; l'expression de son espérance et de sa matrice de covariance sont une conséquence de la section 5.5.2. $\square$

Ce théorème, combiné au théorème 5.2, montre que la loi d'un vecteur gaussien est entièrement caractérisée par son espérance m et sa matrice de covariance Γ .

L'expression de la fonction caractéristique d'un vecteur gaussien est à rapprocher de l'expression obtenue dans la section 5.2 pour les variables gaussiennes multivariées $\mathcal{N}_d(m, \Gamma)$ dans le cas où Γ est définie positive.

Puisque la fonction caractéristique caractérise la loi, il est légitime de se demander si, réciproquement, la loi d'un vecteur gaussien possède une densité par rapport à la mesure de Lebesgue sur $\mathbf{R}^d$. Ce n'est pas toujours le cas et tout dépend si Γ est inversible ou pas. Nous reviendrons sur ce point en section 6.5.

6.1.2 Exemples et contre-exemple

Puisque les v.a. constantes sont des lois gaussiennes (de variance nulle), tout vecteur constant est un exemple de vecteur gaussien.

Un exemple de vecteur gaussien moins trivial est obtenu en considérant des v.a. $X_1, \dots, X_d$ indépendantes de même loi $\mathcal{N}(0, 1)$ et en posant $X = (X_1, \dots, X_d)$.

Démonstration. X est une loi gaussienne multivariée $\mathcal{N}_d(0, I)$ (voir exercice ??) donc, d'après le tableau 5.1, sa fonction caractéristique est donnée par $\exp(-\frac{1}{2} \|t\|^2)$. D'après le théorème 6.2, X est un vecteur gaussien. $\square$

Plus généralement, on peut obtenir un vecteur gaussien par concaténation de v.a. gaussiennes indépendantes (on peut établir rapidement ce résultat en appliquant le résultat énoncé en section 6.4).

Si X est un vecteur gaussien, alors tout sous-vecteur est encore un vecteur gaussien. En particulier, toute composante d'un vecteur gaussien est un vecteur gaussien réel *i.e.*, c'est une loi gaussienne (donc soit une v.a. constante, soit une v.a. de densité de la forme (6.1)).

La proposition suivante précise le lien entre les paramètres du vecteur gaussien et les paramètres de la loi gaussienne de chaque composante.

Proposition 6.3. Soit $X \sim \mathcal{N}_d(m, \Gamma)$. Pour tout $k \in \{1, \dots, n\}$, $X_k \sim \mathcal{N}(m_k, \Gamma_{k,k})$.

Démonstration. On a pour tout $t \in \mathbf{R}$:

$$\phi_{X_k}(t) = \phi_X((0, \dots, 0, t, 0, \dots, 0)) = \exp(itm_k - \frac{1}{2} t^2 \Gamma_{k,k}) .$$

D'après le théorème 5.2, la fonction caractéristique caractérise la loi et à droite, on reconnaît la fonction caractéristique d'une loi $\mathcal{N}(m_k, \Gamma_{k,k})$ (voir Eq. (6.2)). $\square$

Réciproquement, est-il vrai qu'un vecteur aléatoire tel que toutes ses composantes sont des v.a. gaussiennes est un vecteur gaussien ? la réponse est non comme le montre l'exercice ??.

6.2 Caractérisation de l'indépendance

Nous savons que si les composantes $X_1, \dots, X_d$ d'un vecteur aléatoire X sont indépendantes, alors ces v.a. sont décorrélées et la matrice de covariance de X est une matrice diagonale.

Le théorème suivant établit un résultat plus fort en considérant la réciproque : si les composantes $X_1, \dots, X_d$ du vecteur gaussien X sont décorrélées (*i.e.*, la matrice de covariance de X est diagonale) alors ces composantes sont indépendantes.

Nous insistons sur le fait que la décorrélation deux à deux d'une famille de v.a. n'entraîne pas nécessairement l'indépendance mutuelle de ces v.a. (voir par exemple, l'exercice ??) et que le résultat est ici établi sous des hypothèses précises sur la loi jointe de cette famille de v.a.

Théorème 6.4. Soient $(X_k)_{k \leq d}$ des v.a. réelles définies sur $(\Omega, \mathcal{F}, \mathbf{P})$. Les deux conditions sont équivalentes :

- a) Le vecteur aléatoire $(X_1, \dots, X_d)$ est un vecteur gaussien et $\text{Cov}(X_i, X_j) = 0$ pour tout $i \neq j$.
- b) Les v.a. $X_1, \dots, X_d$ sont des v.a. gaussiennes **indépendantes**.

Démonstration. Soit Γ la matrice du vecteur aléatoire $X = (X_1, \dots, X_d)$. Supposons que Γ est de la forme $\text{diag}(\sigma_1^2, \dots, \sigma_d^2)$, $\sigma_k \geq 0$. Alors, d'après le théorème 6.2, la fonction caractéristique de X est

$$\phi_X(t) = \exp(i \sum_{k=1}^d t_k m_k) \exp(-\frac{1}{2} \sum_{k=1}^d t_k^2 \sigma_k^2) = \prod_{k=1}^d \exp(it_k m_k - \frac{1}{2} t_k^2 \sigma_k^2) .$$

Or on sait que chaque composante X_k suit une loi $\mathcal{N}(m_k, \sigma_k^2)$ (cf. Proposition 6.3). Donc $\phi_X(t) = \prod_{k=1}^d \phi_{X_k}(t_k)$. Ainsi, par le théorème ??, les v.a. sont indépendantes.

Réciproquement, si les v.a. $(X_k)_{k \leq d}$ sont indépendantes, alors leur covariance est nulle. Le fait que pour tout $t \in \mathbf{R}^d$, $\langle t, X \rangle$ est une v.a. gaussienne est établi en section 6.1.2. $\square$

Corollaire 6.5. Soit $X = (X_1, \dots, X_d) \sim \mathcal{N}_d(m, \Gamma)$. Les v.a. $X_1, \dots, X_d$ sont indépendantes si et seulement si la matrice de covariance Γ est diagonale.

6.3 Stabilité par transformation affine

Proposition 6.6. Soient $b \in \mathbf{R}^p$ un vecteur déterministe et A une matrice $p \times d$ déterministe. Soit $X \sim \mathcal{N}_d(m, \Gamma)$. Alors $AX + b$ est un vecteur gaussien (à valeur dans $\mathbf{R}^p$), d'espérance $Am + b$ et de matrice de covariance $A\Gamma A^T$.

Démonstration. Calculons la fonction caractéristique de $AX + b$. Soit $t \in \mathbf{R}^p$. On a

$$\phi_{AX+b}(t) = \exp(i\langle t, b \rangle) \phi_X(A^T t) .$$

Par le Théorème 6.2, il vient

$$\begin{aligned} \phi_X(A^T t) &= \exp(i\langle A^T t, m \rangle) \exp\left(-\frac{1}{2} (A^T t)^T \Gamma (A^T t)\right) \\ &= \exp(i\langle t, Am \rangle) \exp\left(-\frac{1}{2} t^T (A\Gamma A^T) t\right) . \end{aligned}$$

Ainsi,

$$\phi_{AX+b}(t) = \exp(i\langle t, b + Am \rangle) \exp\left(-\frac{1}{2} t^T (A\Gamma A^T) t\right) ,$$

et en utilisant encore le Théorème 6.2, on en déduit que $AX + b$ est un vecteur gaussien d'espérance $Am + b$ et de matrice de covariance $A\Gamma A^T$. $\square$

Construction d'un vecteur gaussien

Le théorème 6.4 prouve l'existence d'un vecteur gaussien centré réduit (*i.e.*, d'espérance nulle et de matrice de covariance égale à l'identité). Nous montrons comment n'importe quel vecteur gaussien $\mathcal{N}_d(m, \Gamma)$ s'obtient par transformation affine d'un vecteur gaussien centré réduit.

- Etape 1 : construction d'un vecteur de loi $\mathcal{N}_d(0, I)$. Le théorème 6.4 montre que le vecteur $Y := (Y_1, \dots, Y_d)$ où $(Y_k)_{k \leq d}$ sont des v.a. gaussiennes centrées réduites indépendantes a pour loi $\mathcal{N}_d(0, I)$.

- Etape 2 : transformation affine de Y . Soit une matrice $\sqrt{\Gamma}$ telle que $\sqrt{\Gamma}\sqrt{\Gamma}^T = \Gamma$ (comme Γ est une matrice positive, $\sqrt{\Gamma}$ existe toujours¹). La proposition 6.6 entraîne que $m + \sqrt{\Gamma}Y$ pour loi $\mathcal{N}_d(m, \Gamma)$.

6.4 Somme de vecteurs gaussiens indépendants

La proposition suivante montre que la somme de **vecteurs gaussiens indépendants** est encore un vecteur gaussien, dont l'espérance (resp. la matrice de covariance) est la somme des espérances (resp. des matrices de covariance).

Il est important de noter que ce résultat n'est vrai que si les variables sont indépendantes ; considérons en effet le contre-exemple suivant. Soit $X \sim \mathcal{N}(0, 1)$ et $Y = -X$; notons que $Y \sim \mathcal{N}(0, 1)$ de sorte que X et Y sont deux vecteurs gaussiens. Alors $X + Y = 0$ (avec probabilité 1) et donc $X + Y$ est un vecteur gaussien de variance nulle (la variance n'est donc pas égale à la somme des variances). Mais X et Y ne sont pas indépendantes puisque

$$\mathbf{E}(XY) = -\mathbf{E}(X^2) = -1 \neq 0 = \mathbf{E}(X)\mathbf{E}(Y) .$$

Proposition 6.7. Soient $X^{(1)}, \dots, X^{(p)}$ des vecteurs aléatoires indépendants tels que $X^{(\ell)} \sim \mathcal{N}_d(m^{(\ell)}, \Gamma^{(\ell)})$. Alors

$$X^{(1)} + \dots + X^{(p)} \sim \mathcal{N}_d\left(\sum_{\ell=1}^p m^{(\ell)}; \sum_{\ell=1}^p \Gamma^{(\ell)}\right) .$$

Démonstration. Pour tout $t \in \mathbf{R}^d$,

$$\begin{aligned} \phi_{X^{(1)} + \dots + X^{(p)}}(t) &= \mathbf{E}\left(\exp(i\langle t, X^{(1)} \rangle + \dots + i\langle t, X^{(p)} \rangle)\right) \\ &= \prod_{\ell=1}^p \mathbf{E}\left(\exp(i\langle t, X^{(\ell)} \rangle)\right) = \prod_{\ell=1}^p \phi_{X^{(\ell)}}(t) \end{aligned}$$

puisque les v.a. $(X^{(\ell)})_{\ell \leq p}$ sont indépendantes. En utilisant le formulaire Tableau 5.1

$$\phi_{X^{(\ell)}}(t) = \exp\left(i\langle t, m^{(\ell)} \rangle - \frac{1}{2} t^T \Gamma^{(\ell)} t\right) ,$$

donc

$$\phi_{X^{(1)} + \dots + X^{(p)}}(t) = \exp\left(i\langle t, \sum_{\ell=1}^p m^{(\ell)} \rangle - \frac{1}{2} t^T \left\{\sum_{\ell=1}^p \Gamma^{(\ell)}\right\} t\right) ,$$

et l'on conclut par application du théorème 5.2 et du formulaire Tableau 5.1. $\square$

1. puisque Γ est une matrice de covariance, il existe une matrice orthogonale Q et une matrice diagonale Δ dont les éléments diagonaux $\Delta_{j,j}$ sont positifs ou nuls telles que $\Gamma = Q\Delta Q^T$. On peut prendre $\sqrt{\Gamma} = Q\sqrt{\Delta}Q^T$ où $\sqrt{\Delta}$ est la matrice diagonale dont les éléments diagonaux sont $\sqrt{\Delta_{j,j}}$.

6.5 La loi d'un vecteur gaussien admet-elle une densité ?

Soit $X \sim \mathcal{N}_d(m, \Gamma)$. On distingue deux cas :

Lorsque Γ est inversible. Nous avons établi au chapitre 5 (voir Tableau 5.1 et théorème 5.2) que la loi d'un vecteur gaussien admet une densité :

$$\frac{1}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det(\Gamma)}} \exp\left(-\frac{1}{2} (x - m)^T \Gamma^{-1} (x - m)\right) .$$

Lorsque Γ est non-inversible. La loi du vecteur gaussien n'admet pas de densité. On peut montrer que

$$\mathbf{P}[X - m \in \text{Im}(\Gamma)] = 1 ; \quad (6.4)$$

autrement dit, la v.a. X prend ses valeurs dans l'espace $m + \text{Im}(\Gamma)$ avec probabilité 1.

Démonstration. Notons r le rang de Γ et $u_{r+1}, \dots, u_d$ une base orthonormal de l'espace orthogonal de $\text{Im}(\Gamma)$. Alors pour tout $r + 1 \leq k \leq d$,

$$\text{var}(\langle u_k, (X - m) \rangle) = u_k^T \mathbf{E} \left((X - m)(X - m)^T \right) u_k = u_k^T \Gamma u_k = 0 .$$

Donc la v.a. $\langle u_k, X - m \rangle$ est constante avec probabilité 1, et comme son espérance est nulle, elle vaut zero avec probabilité 1. Ainsi, $X - m$ est, avec probabilité 1, orthogonal au vecteur u_k pour tout $r + 1 \leq k \leq d$. Ce qui conclut la démonstration. $\square$

Par suite, la loi ne peut pas posséder de densité par rapport à la mesure de Lebesgue sur $\mathbf{R}^d$. On dit dans ce cas que le vecteur gaussien est *dégénéré*.

Démonstration. En effet, supposons qu'elle en possède une, notée f . On a alors :

$$1 = \mathbf{P}[X - m \in \text{Im}(\Gamma)] = \int_{m + \text{Im}(\Gamma)} f(x) d\lambda_d(x) = 0$$

où la dernière égalité vient du fait que $\dim(\text{Im}(\Gamma)) < d$ et donc $\lambda_d(m + \text{Im}(\Gamma)) = 0$. Cela conduit à une contradiction. $\square$

Convergences

7.1 Loi des grands nombres

Définition 7.1. On dit qu'une suite $(X_n, n \geq 1)$ de v.a. converge **P**-presque-sûrement (ou **P**-presque-partout) vers une v.a. X lorsqu'il existe un ensemble A tel que $\mathbf{P}(A^c) = 0$ et pour tout $\omega \in A$,

$$X_n(\omega) \xrightarrow{n \rightarrow +\infty} X(\omega).$$

En d'autres termes, il s'agit de la convergence simple à un ensemble de mesure nulle près.

Théorème 7.1 (Loi forte des grands nombres). Soit $(X_n, n \geq 1)$ une suite de v.a. indépendantes, identiquement distribuées telles que $\mathbf{E}[|X_1|] < \infty$ alors

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{n \rightarrow +\infty} \mathbf{E}[X_1], \quad \mathbf{P} - \text{p.p.}$$

7.2 Limite centrée

Définition 7.2. Pour un ensemble ouvert $A \in \mathbf{R}^k$, on note ∂A sa frontière définie par

$$\partial A = \bar{A} - A.$$

Pour un intervalle $]a, b[$, on a alors $\partial A = \{a, b\}$. Pour un pavé ouvert de $\mathbf{R}^k$, la frontière au sens topologique correspond à la notion intuitive de bord.

REMARQUE.— Si Y a même loi que X et si $(X_n, n \geq 1)$ converge en loi vers X alors $(X_n, n \geq 1)$ converge aussi vers Y . La convergence en loi, malgré sa présentation, n'est pas une convergence de variables aléatoires mais une convergence des mesures associées aux v.a..

Théorème 7.2. La convergence presque sûre implique la convergence en loi mais la réciproque est fausse.

Démonstration. Si $(X_n, n \geq 1)$ converge p.s. vers X alors pour toute fonction continue bornée,

- $f(X_n) \xrightarrow{n \rightarrow +\infty} f(X)$, presque-sûrement,
- pour tout $n \geq 1$, $|f(X_n)| \leq \|f\|_\infty$
- et $\mathbf{E}[\|f\|_\infty] < \infty$,

donc toutes les hypothèses du théorème de convergence dominée sont satisfaites, d'où

$$\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[f(X)].$$

D'après la première caractérisation de la convergence en loi, cela signifie que $(X_n, n \geq 1)$ converge en loi vers X . $\square$

REMARQUE.— Construisons un contre-exemple à la réciproque. Soit X une v.a. gaussienne de moyenne nulle. Comme la densité gaussienne est paire, $-X$ suit la même loi que X . Considérons pour tout $n \geq 1$, la suite $X_n = X$. Il est clair que X_n converge en loi vers X donc vers $-X$. En revanche, X_n ne converge vers $-X$ que sur l'ensemble $(X = -X)$, c'est-à-dire l'ensemble $(X = 0)$, qui est de probabilité nulle puisque la loi gaussienne est absolument continue.

Définition 7.3. On dit qu'une suite $(X_n, n \geq 1)$ de v.a., à valeurs dans $\mathbf{R}^k$, converge en loi vers X lorsque l'une des propriétés équivalentes suivantes est vérifiée :

- Pour toute fonction continue bornée f de $\mathbf{R}^k$ dans $\mathbf{R}$,

$$\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[f(X)],$$

- pour tout ensemble ouvert $A \in \mathbf{R}^k$ tel que $\mathbf{P}(X \in \partial A) = 0$,

$$\mathbf{P}(X_n \in A) \xrightarrow{n \rightarrow +\infty} \mathbf{P}(X \in A),$$

- pour tout $t \in \mathbf{R}^k$,

$$\mathbf{E}[e^{it \cdot X_n}] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[e^{it \cdot X}].$$

Le théorème le plus important de ce chapitre est connu sous le nom de Théorème de la Limite Centrée (TLC en français, TCL en anglais).

Théorème 7.3. Soit $(X_n, n \geq 1)$ une suite de v.a. indépendantes, identiquement

distribuées telles que $\mathbf{E}[|X_1|^2] < \infty$ alors

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{j=1}^n X_j - \mathbf{E}[X_1] \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1), \text{ en loi}$$

où

$$\sigma^2 = \text{var}(X_1).$$

Notations

$\mathcal{C}(c)$, 40

$\mathcal{N}(m, \sigma^2)$, 40

$\mathcal{U}(a, b)$, 39

$\mathcal{E}(\lambda)$, 39

F_X , 37

Index des sujets

changement
variables, 50

fonction
répartition, 37

loi
Cauchy, 40
exponentielle, 39

gaussienne, 40
marginale, 49
normale, 40
uniforme, 39

variable aléatoire
vectorielle, 49
variable aléatoire
réelle, 37