



PROBABILITÉS DE BASE.¹

Cours de Master 1, MM010,
Sorbonne université



Cours : Pierre TARRAGO² et Quentin BERGER³
Travaux dirigés : David GARCIA-ZELADA⁴

Premier semestre de l'année universitaire 2021-2022.

1. Texte du cours établi par T. Duquesne. Exercices corrigés de T. Lévy, T. Duquesne, L. Mazliak, P. Tarrago, Q. Berger, D. Garcia-Zelada
2. Email : pierre.tarrago@sorbonne-universite.fr
3. Email : quentin.berger@sorbonne-universite.fr
4. Email : david.garcia-zelada@sorbonne-universite.fr

Table des matières

Avant-propos pour le cours 2021/2022	vii
Avant-propos, brève bibliographie.	ix
I Le modèle probabiliste.	1
I.1 Rappels de théorie de l'intégration.	1
I.1.a Rappels sur les tribus.	1
I.1.b Rappels sur les mesures positives.	6
I.2 Premières notions de probabilité.	12
I.2.a Ce que modélise un espace de probabilité.	12
I.2.b Exemples, équi-probabilité et dénombrement.	15
I.2.c Probabilités conditionnelles.	18
I.2.d Premières notions sur l'indépendance.	20
I.3 Exercices.	24
II Variables aléatoires, espérance, lois.	27
II.1 Variables aléatoires.	27
II.1.a Rappels sur les fonctions mesurables.	27
II.1.b Variables aléatoires : interprétation probabiliste.	33
II.2 Espérance.	36
II.2.a Rappels sur la construction et les propriétés de l'intégrale.	36
II.2.b L'espérance mathématique.	45
II.3 Loi d'une variable.	46
II.3.a Cas des variables discrètes.	46
II.3.b Variables de Bernoulli, binomiales, géométriques et de Poisson.	49
II.3.c Rappel sur les mesures images.	51
II.3.d Loi d'une v.a. : cas général ; transfert.	52
II.4 Inégalités usuelles.	54
II.4.a Inégalité de Jensen, moments.	54
II.4.b Inégalités de Hölder.	55
II.4.c Inégalités de Minkowski.	55
II.4.d Inégalités de type Markov, variance et inégalité de Tchebychev.	56
II.5 Fonctions génératrices de v.a. entières.	58
II.6 Densité d'une variable aléatoire.	60
II.6.a Rappel de théorie de l'intégration : mesures à densité.	60
II.6.b Variables dont les lois admettent une densité.	63

II.7	Fonctions de répartition.	66
II.7.a	Rappel sur les atomes d'une mesure.	66
II.7.b	Continuité, limites à droite et à gauche.	67
II.7.c	Fonctions de répartition.	69
II.8	Exercices.	73
II.9	Vecteurs aléatoires.	74
II.9.a	Rappels sur les mesures produit et le théorème de Fubini.	74
II.9.b	Outil de calcul intégral en dimension n	77
II.9.c	Vecteurs aléatoires ; covariance.	81
II.10	Fonction caractéristique des vecteurs aléatoires.	83
II.10.a	Rappel sur les intégrales à paramètres.	83
II.10.b	Injectivité de la fonction caractéristique, formule d'inversion.	85
II.11	Transformée de Laplace ; fonctions génératrices.	92
II.12	Exercices	94
III	Indépendance.	97
III.1	Indépendance d'événements et de classes d'événements.	97
III.1.a	Indépendance d'événements (rappel).	97
III.1.b	Indépendance de classes d'événements.	97
III.2	Indépendance de variables aléatoires.	99
III.2.a	Résultats généraux.	99
III.2.b	Variables indépendantes de loi diffuse, statistique d'ordre.	102
III.2.c	Indépendance, espérance et covariance.	105
III.2.d	Convolution ; loi de sommes de variables indépendantes	107
III.3	Lemme de Borel-Cantelli, réciproques partielles et lois des grands nombres de Borel.	111
III.3.a	Lemme de Borel-Cantelli et loi des grands nombres de Borel.	111
III.3.b	Réciproques partielles du lemme de Borel-Cantelli.	113
III.4	Construction des suites de v.a. réelles indépendantes.	116
III.5	Exercices.	122
IV	Convergence de variables aléatoires. Lois des grands nombres.	125
IV.1	Resultats généraux de convergence.	125
IV.1.a	Convergence presque sûre.	125
IV.1.b	Convergence en probabilité.	127
IV.2	Convergence L^p	131
IV.3	Lois faibles des grands nombres.	134
IV.4	Loi Forte des grands nombres.	136
IV.4.a	Loi du 0-1 de Kolmogorov.	136
IV.4.b	La loi forte des grands nombres.	137
IV.5	Exercices.	143
V	Convergence en loi. Théorème central limite.	147
V.1	Définition, premiers résultats.	147
V.2	Convergence en loi pour des v.a. réelles.	152
V.2.a	Reformulation en termes de la fonction de répartition.	152
V.2.b	Application au théorème Central-Limite de De Moivres-Laplace	154

V.3	convergence en loi sur \mathbb{N}	156
V.3.a	Premiers résultats.	156
V.3.b	Application aux variables de Poisson et aux événements rares.	158
V.4	Convergence en loi et fonction caractéristique.	162
V.4.a	Le théorème de Paul Lévy.	162
V.4.b	Le théorème Central-Limite sur \mathbb{R} , intervalles de confiance.	164
V.4.c	Vecteurs gaussiens et TCL sur \mathbb{R}^d	168
V.5	Exercices	172
A	Table de la loi gaussienne standard.	175

Avant-propos pour le cours 2021/2022

Attention. *Ces notes ne sont destinées qu'aux étudiants du cours MM010 de l'UPMC; aucune diffusion autre qu'interne au master n'est autorisée.*

L'objectif de ce cours est d'enseigner les concepts fondamentaux de la théorie des probabilités. En particulier, nous supposerons connus les cours de théorie de la mesure et d'analyse de L3 : les notions issues de ces enseignements antérieurs et utilisées dans le cours seront la plupart du temps rappelées et les preuves refaites ou laissées en exercice.

Les exercices proposés dans le livret d'exercices ont pour but de vous familiariser avec les différentes notions de probabilités apprises pendant le cours. Certains exercices complémentaires seront également proposés en cours et en TD. Il est extrêmement important que vous passiez du temps à chercher l'ensemble de ces exercices en vous abstenant de regarder la solution dans un premier temps.

La progression du cours, les feuilles d'exercices et leur corrections seront mises à disposition sur la page Moodle du cours. Il est également important que vous fournissiez un travail régulier au fur et à mesure de l'avancée du cours, afin que vous ayez le temps de réfléchir aux différentes notions abordées durant le cours.

Sauf changement qui vous sera notifié le cas échéant, il y aura deux évaluations pendant le semestre pour l'enseignement en présence et une seule pour l'enseignement à distance : un partiel **le 29 octobre** pour l'enseignement en présence et un examen final **la semaine du 3 janvier** pour l'ensemble des étudiants. La note finale pour les étudiants en présence comprendra la note du partiel à hauteur de 30/100 et la note de l'examen final à hauteur de 70/100.

N'hésitez pas à contacter Quentin Berger, David Garcia-Zelada ou Pierre Tarrago pour toute question relative au contenu du cours ou à son organisation.

Avant-propos, brève bibliographie.

Attention. Ces notes ne sont destinées qu'aux étudiants du cours MM010 de l'UPMC ; aucune diffusion autre qu'interne au master n'est autorisée.

Les pré-requis de ce cours sont les suivants.

- Un cours d'analyse de L3. On suppose le lecteur familier avec les suites et les séries de nombres réels et complexes, la topologie de la norme de \mathbb{R}^d , la convergence de suites et de séries de fonctions pour la norme uniforme sur des sous-ensembles de \mathbb{R} et de \mathbb{R}^d . On suppose le lecteur familier avec la notion de fonctions uniformément continues (au moins pour les fonctions définies sur \mathbb{R}^d).
- Un cours d'intégration général suivant le programme du cours « Intégration I » de L3 de l'UPMC : tribus, mesures positives, intégrale contre une mesure positive, convergence monotone, Fatou, convergence dominée, interversion série/intégrale dans la version positive et dans la version intégrable. Mesures produit, théorèmes de Fubini. Mesure de Lebesgue sur \mathbb{R} et dans \mathbb{R}^d , théorème de changement de variable (faisant intervenir le Jacobien).

Durant le cours et dans ce polycopié, on rappelle les notions d'intégration nécessaires ; même si cela n'est pas le sujet principal de ce cours, des preuves de ces résultats sont rappelées. Néanmoins, pour les notions d'intégration, nous renvoyons les étudiants aux ouvrages suivants.

- BRIANE, M. ET PAGÈS G. *Théorie de l'intégration*. Vuibert, Paris (FR) 2006. 4ème édition. Licence et master de mathématiques
- GRAMAIN, A. *Intégration*. Collection Méthodes. Hermann, Paris, 1988.

Et aussi, en anglais, l'excellente référence :

- RUDIN, W. *Real and complex analysis*, third ed. McGraw-Hill Book Co., New York, 1987.

Ce cours contient 5 chapitres. Il est aussi auto-contenu que possible et la plupart des énoncés importants sont prouvés en détail. Des exemples importants sont traités ; nous fournissons aussi de nombreux exercices avec leur corrigé. Les exercices sont souvent simples : ils n'ont pas d'autre but que de permettre de vérifier si l'on a compris les notions du cours. *L'essentiel du travail consiste à comprendre puis apprendre son cours ainsi qu'un certain nombre de preuves.*

Une table de la fonction de répartition de la loi gaussienne standard se trouve à la fin du polycopié : cela sert dans plusieurs exercices concernant les applications du théorème Central-Limite. Le cours est muni d'un index assez complet. Un exemple d'examen se trouve également en appendice.

Ce photocopié ne prétend pas à l'originalité : il existe de très nombreux textes d'introduction aux probabilités. L'étudiant est invité à les consulter. Voici une brève liste non-exhaustive d'ouvrages profitables.

En français.

- BARBÉ, P. ET LEDOUX M. *Probabilité*, EDP Sciences, Paris, 2007.
- BERGER, Q., CARAVENNA, F. ET DAI PRA, P. *Introduction aux probabilités*, Dunod 2021.
- GARET, O. ET KURTZMANN, A. *De l'intégration aux probabilités*, Ellipses, 2ème édition, 2019.

En anglais.

- BILLINGSLEY, P. *Probability and measure*, third ed. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1995. A Wiley-Interscience Publication.
- DURRETT, R. *Probability : theory and examples*, fourth ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

Nous vous souhaitons un bon travail.

Le 8 Septembre 2021.

Pierre Tarrago & Quentin Berger.

Chapitre I

Le modèle probabiliste.

I.1 Rappels de théorie de l'intégration.

I.1.a Rappels sur les tribus.

Notations sur les ensembles. Les mathématiques actuelles se fondent sur la notion d'ensemble et le jeu d'axiomes dits de Zermelo-Fraenkel qui lui donne un sens. Nous ne revenons pas sur la notion d'ensemble et nous supposons que le lecteur est familier avec les notations ensemblistes usuelles : le symbole $=$, le symbole « appartient », noté \in , sa négation, notée \notin , l'inclusion, notée \subset , l'union, notée \cup , l'intersection notée \cap , l'ensemble vide, noté \emptyset .

Les ensembles seront la plupart du temps notés par des capitales scriptes du début de l'alphabet A, B, C ou E . Les points des ensembles sont notés par des lettres minuscules $x, y, z, a, c, s \dots$. Par exemple, l'assertion « x est un élément de l'ensemble A » se note « $x \in A$ ».

Soient A et B deux sous-ensembles de l'ensemble E , c'est-à-dire $A \subset E$ et $B \subset E$. L'ensemble $A \cup B$ se lit « A union B » et l'ensemble $A \cap B$ se lit « A inter B » et on rappelle que

$$A \cup B = \{x \in E : x \in A \text{ ou } x \in B\} \quad \text{et} \quad A \cap B = \{x \in E : x \in A \text{ et } x \in B\} .$$

L'ensemble B privé de A , noté $B \setminus A$, est défini par

$$B \setminus A = \{x \in E : x \in B \text{ et } x \notin A\} .$$

L'ensemble complémentaire de A dans E est parfois noté A^c lorsqu'il n'y a pas d'ambiguïté mais on lui préfère la notation explicite $E \setminus A = \{x \in E : x \notin A\}$. On voit donc que $B \setminus A = B \cap (E \setminus A)$. On vérifie aussi que pour tout sous-ensemble $C \subset E$,

$$C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B) \quad \text{et} \quad C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B) .$$

Soit $(A_i)_{i \in I}$, une famille de sous-ensembles de E , c'est-à-dire que I est un ensemble (non-vide) d'indices et que pour tout $i \in I$, $A_i \subset E$. Alors l'union des A_i et l'intersection des A_i se définissent respectivement par

$$\bigcup_{i \in I} A_i = \{x \in E : \exists i \in I : x \in A_i\} \quad \text{et} \quad \bigcap_{i \in I} A_i = \{x \in E : \forall i \in I : x \in A_i\}$$

et on rappelle que pour tout $B \subset E$,

$$C \cap \left(\bigcup_{i \in I} A_i \right) = \bigcup_{i \in I} C \cap A_i, \quad C \setminus \left(\bigcup_{i \in I} A_i \right) = \bigcap_{i \in I} C \setminus A_i \quad \text{et} \quad C \setminus \left(\bigcap_{i \in I} A_i \right) = \bigcup_{i \in I} C \setminus A_i.$$

On dit que les ensembles $A_i, i \in I$ sont *deux-à-deux disjoints* si

$$\forall i, j \in I \text{ distincts, } A_i \cap A_j = \emptyset .$$

Il est nécessaire de préciser « deux-à-deux » car l'assertion « les $A_i, i \in I$, sont disjoints » signifie plutôt que $\bigcap_{i \in I} A_i = \emptyset$, ce qui est bien évidemment moins restrictif.

On rappelle la définition suivante : on dit que les $(A_i)_{i \in I}$ forment une *partition* de E si les A_i sont disjoints deux-à-deux et si leur union est $E : \bigcup_{i \in I} A_i = E$.

Plutôt que de parler d'ensembles d'ensembles on préfère parler de *classes d'ensembles*; on les note par des majuscules cursives $\mathcal{A}, \mathcal{B}, \mathcal{P}, \mathcal{L}, \mathcal{F}, \dots$. On note $\mathcal{P}(E)$ la classe de tous les sous-ensembles de $E : \mathcal{P}(E) = \{A : A \subset E\}$. Si \mathcal{A} est une classe de sous-ensembles de E alors $\mathcal{A} \subset \mathcal{P}(E)$.

Rappelons quelques points élémentaires d'énumération : si E est un ensemble fini à n éléments on pose $\#E = n$ que l'on appelle son cardinal. Si E est infini, $\#E = \infty$. Si $\#E = 0$, alors E est l'ensemble vide \emptyset . Supposons que E soit un ensemble de n éléments, c'est-à-dire que $\#E = n$, alors

$$\#\mathcal{P}(E) = 2^n \quad \text{et} \quad \#\{A \subset E : \#A = k\} = \binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad 0 \leq k \leq n .$$

Soient $A_1, \dots, A_p \subset E$, p sous-ensembles de E . Le produit $A_1 \times \dots \times A_p$ est l'ensemble des listes (x_1, \dots, x_p) telles que $x_1 \in A_1, \dots, x_p \in A_p$. On rappelle alors que

$$\#(A_1 \times \dots \times A_p) = \prod_{1 \leq k \leq p} \#A_k .$$

Tribus. On rappelle ici la notion de tribu qui joue un rôle important dans la théorie des probabilités.

Définition I.1.1 ◀▶ Soit E un ensemble non-vide. On rappelle que la classe de tous les sous-ensembles de E est notée $\mathcal{P}(E)$.

Une *tribu* (aussi appelée *sigma-algèbre*) est une classe de sous-ensembles $\mathcal{E} \subset \mathcal{P}(E)$ qui satisfait les propriétés suivantes.

- (a) $E \in \mathcal{E}$.
- (b) $A \in \mathcal{E}, E \setminus A \in \mathcal{E}$. (*Stabilité par passage au complémentaire*)
- (c) $\forall A_n \in \mathcal{E}, n \in \mathbb{N}, \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$. (*Stabilité par union dénombrable*).

Les éléments de \mathcal{E} sont appelés les ensembles *mesurables* et le couple (E, \mathcal{E}) est appelé *espace mesurable*. □

Propriétés élémentaires. ◀▶ Soit (E, \mathcal{E}) un espace mesurable.

- (i) $\emptyset \in \mathcal{E}$
- (ii) \mathcal{E} est stable par union finie : si $A, B \in \mathcal{E}$, alors $A \cup B \in \mathcal{E}$.
- (iii) \mathcal{E} est stable par intersection dénombrable : pour tous $A_n \in \mathcal{E}, n \in \mathbb{N}$, on a $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{E}$.
De même \mathcal{E} est stable par intersection finie.

◀► **Preuve :** par (a) et (b), $\emptyset = E \setminus E \in \mathcal{E}$, ce qui prouve (i). Le point (ii) découle de (c) en prenant $A_1 = A$, $A_2 = B$, et $A_n = \emptyset$, pour $n \geq 2$. Soient $A, B \in \mathcal{E}$. Alors, $A \cap B = E \setminus ((E \setminus A) \cup (E \setminus B))$, ce qui montre que \mathcal{E} est stable par intersection finie. Plus généralement, on remarque que $\bigcap_{n \in \mathbb{N}} A_n = E \setminus \bigcup_{n \in \mathbb{N}} (E \setminus A_n)$, ce qui montre (iii). ■

Exemple I.1.1 Soit E un ensemble non-vide. Alors $\mathcal{P}(E)$ est une tribu. Toute tribu sur E est contenue dans $\mathcal{P}(E)$, qui est donc la plus grande tribu sur E . □

Exemple I.1.2 Soit E un ensemble non-vide. On pose $\mathcal{E} = \{\emptyset, E\}$. C'est une tribu. Toute autre tribu sur E contient cette tribu qui est appelée *tribu grossière* : c'est la plus petite tribu sur E . □

Exemple I.1.3 Soit E un ensemble infini. On note \mathcal{E} la classe des ensembles qui sont soit dénombrables soit de complémentaire dénombrable (par dénombrable on entend vide, fini ou en bijection avec \mathbb{N}). On vérifie facilement que \mathcal{E} est une tribu. □

Il est facile de montrer qu'une intersection *quelconque* de tribus est encore une tribu. Plus précisément, on établit le lemme suivant dont la preuve, très simple, est laissée au lecteur.

Lemme I.1.1 ▶◀ Soit $\mathcal{E}_i, i \in I$, une famille de tribus sur E . Alors,

$$\bigcap_{i \in I} \mathcal{E}_i = \{A \subset E : \forall i \in I, A \in \mathcal{E}_i\}$$

est une tribu.

Définition I.1.2 ▶◀ Soit $\mathcal{R} \subset \mathcal{P}(E)$, une classe quelconque de sous-ensembles de E . On pose

$$\sigma(\mathcal{R}) = \bigcap_{\substack{\mathcal{E} \text{ tribu} \\ \mathcal{R} \subset \mathcal{E}}} \mathcal{E}.$$

C'est la *tribu engendrée* par \mathcal{R} , qui est la plus petite tribu contenant \mathcal{R} . □

Tribus Boréliennes. On rappelle tout d'abord la définition d'une topologie.

Définition I.1.3 ▶◀ Soit E un ensemble non-vide. Une *topologie* sur E est une classe d'ensembles $\mathcal{T} \subset \mathcal{P}(E)$ qui satisfait les propriétés suivantes.

- (a) \mathcal{T} est stable par union quelconque : soit $U_i \in \mathcal{T}, i \in I$, une famille quelconque d'ensembles de \mathcal{T} , alors $\bigcup_{i \in I} U_i \in \mathcal{T}$.
- (b) $\emptyset \in \mathcal{T}$ et $E \in \mathcal{T}$.
- (c) \mathcal{T} est stable par intersection simple : si $U, V \in \mathcal{T}$, alors $U \cap V \in \mathcal{T}$.

Les ensembles d'une topologie sont appelés les *ouverts de E* . Le complémentaire d'un ensemble ouvert est appelé *ensemble fermé*. (E, \mathcal{T}) est appelé *espace topologique*. □

Exemple I.1.4 ▶◀ La topologie usuelle de \mathbb{R} est décrite comme suit : un ouvert de \mathbb{R} est une réunion dénombrable d'intervalles ouverts deux-à-deux disjoints. □

Définition I.1.4 ◀▶ (Boréliens) Soit (E, \mathcal{T}) , un espace topologique. La tribu des Boréliens qui lui est associée est la tribu $\sigma(\mathcal{T})$ engendrée par la classe des ouverts : on la note $\mathcal{B}(E)$, lorsqu'il n'y a pas d'ambiguïté sur la topologie sur E . □

Lemme I.1.2 ◀▶ On pose $\mathcal{R} = \{]-\infty, a]; a \in \mathbb{R} \}$. Alors $\sigma(\mathcal{R}) = \mathcal{B}(\mathbb{R})$.

▶▶ **Preuve** : comme $\mathbb{R} \setminus]-\infty, a] =]a, \infty[$ est un ouvert, et donc un Borélien, on voit que $\mathcal{R} \subset \mathcal{B}(\mathbb{R})$, et par définition de la tribu engendrée, cela implique que $\sigma(\mathcal{R}) \subset \mathcal{B}(\mathbb{R})$.

On remarque ensuite que $]-\infty, a[= \bigcup_{n \in \mathbb{N}^*}]-\infty, a - \frac{1}{n}[$. Donc tous les intervalles ouverts infinis à gauche sont dans $\sigma(\mathcal{R})$. Soient $a \leq b$; on a $]a, b[=]a, \infty[\cap]-\infty, b[$. Cela montre que tout intervalle ouvert est dans la tribu $\sigma(\mathcal{R})$. On note \mathcal{T} la topologie de \mathbb{R} . Comme tout ouvert de \mathbb{R} est une union dénombrable d'intervalles ouverts, ce qui précède montre que $\mathcal{T} \subset \sigma(\mathcal{R})$ et, par définition de la tribu engendrée, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{T}) \subset \sigma(\mathcal{R})$, ce qui permet de conclure. ■

Remarque I.1.1 Un raisonnement analogue permet de montrer que $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles de la forme $]a, b]$, ou encore de la forme $]a, \infty[$, avec $a \in \mathbb{Q}$, etc. □

Sur un \mathbb{R} -espace vectoriel de dimension finie, toutes les normes sont équivalentes et définissent donc la même topologie : on parle donc de la tribu Borélienne de \mathbb{R}^d comme de celle engendrée par les ouverts de la topologie de n'importe quelle norme.

Par exemple, on peut choisir de munir \mathbb{R}^2 de la norme du maximum définie pour tout $x = (x_1, x_2) \in \mathbb{R}^2$ par $\|x\|_{\max} = \max(|x_1|, |x_2|)$. La boule ouverte de centre x et de rayon r est en fait le carré ouvert $]x_1 - r, x_1 + r[\times]x_2 - r, x_2 + r[$. On voit ainsi que tout ouvert de \mathbb{R}^2 est une union dénombrable de rectangles ouverts de type $]a, b[\times]c, d[$ où $a, b, c, d \in \mathbb{Q}$. En effet, pour tout point x d'un ouvert U , il existe un carré ouvert (une boule ouverte pour la norme du maximum) centré en x et de côté $2r_x$. En réduisant ce carré, il est possible de trouver $a_x, b_x, c_x, d_x \in \mathbb{Q}$ tels que $x \in]a_x, b_x[\times]c_x, d_x[\subset U$. Donc $U = \bigcup_{x \in U}]a_x, b_x[\times]c_x, d_x[$. Or $\{(a_x, b_x, c_x, d_x); x \in U\}$ est dénombrable car cet ensemble est inclus dans \mathbb{Q}^4 . En raisonnant comme dans le lemme I.1.2, on montre facilement le lemme suivant, qui est utilisé un peu plus loin dans ce cours.

Lemme I.1.3 On pose $\mathcal{R} = \{]a, b[\times]c, d[; (a, b, c, d) \in \mathbb{Q}^4 \}$. Alors $\sigma(\mathcal{R}) = \mathcal{B}(\mathbb{R}^2)$.

Remarque I.1.2 Si E est un ensemble non-vide, un théorème bien connu de Cantor montre que $\text{Card}(E) < \text{Card}(\mathcal{P}(E))$, c'est-à-dire qu'il n'existe pas de surjection de E sur $\mathcal{P}(E)$. On pourrait s'attendre à ce que $\mathcal{B}(\mathbb{R})$ et $\mathcal{P}(\mathbb{R})$ soient de même cardinal or ce n'est pas le cas : on peut en effet montrer, par des arguments de récurrence transfinie, que $\text{Card}(\mathbb{R}) = \text{Card}(\mathcal{B}(\mathbb{R}))$, c'est-à-dire qu'il existe une bijection de \mathbb{R} sur $\mathcal{B}(\mathbb{R})$. D'un point de vue ensembliste, cela signifie qu'il y a peu de Boréliens. Cela montre aussi qu'il existe beaucoup de sous-ensembles de \mathbb{R} qui ne sont pas des Boréliens. Il n'est cependant pas évident d'en exhiber par des méthodes naïves. □

Classe monotone. La remarque I.1.2 montre qu'il n'est pas facile de décrire les Boréliens de \mathbb{R} . Plus généralement, on ne dispose que rarement d'une description simple et explicite des tribus que l'on manipule : elles sont le plus souvent données comme "la tribu engendrée par quelque chose". La notion de classe monotone, donnée ci-après, est proche de celle de tribu ; à première vue, elle ne semble pas plus explicite ou plus simple mais à l'usage, elle se révèle être un concept efficace permettant de démontrer des résultats sur les tribus.

Définition I.1.5 ◀▶ Soit E un ensemble non-vide.

- $\mathcal{P} \subset \mathcal{P}(E)$ est un *pi-système* si
 - (a) $E \in \mathcal{P}$,
 - (b) pour tous $A, B \in \mathcal{P}$, on a $A \cap B \in \mathcal{P}$.
- $\mathcal{L} \subset \mathcal{P}(E)$ est une *classe monotone* si les propriétés suivantes sont vérifiées.
 - (a) $E \in \mathcal{L}$.
 - (b) Pour tous $A, B \in \mathcal{L}$ tels que $A \subset B$, on a $B \setminus A \in \mathcal{L}$ (*stabilité par différence propre*).
 - (c) Si $A_n \in \mathcal{L}$, $n \in \mathbb{N}$ sont des ensembles tels que $A_n \subset A_{n+1}$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$ (*stabilité par union dénombrable croissante*). □

Exemple I.1.5 ◀▶ $\mathcal{P} = \{\mathbb{R}\} \cup \{]-\infty, a]; a \in \mathbb{R}\}$ est un pi-système. □

Il est évident que toute tribu est une classe monotone. Le contraire étant faux en général, comme le montre l'exemple suivant.

Exemple I.1.6 Soit $E = \{1, 2, 3, 4, 5\}$. On pose $A = \{1, 2\}$ et $B = \{1, 3\}$. On vérifie que

$$\mathcal{L} = \{\emptyset, A, B, E \setminus A, E \setminus B, E\}$$

est une classe monotone. Ce n'est de toute évidence pas une tribu. □

Le lemme suivant éclaire le lien entre tribu et classe monotone.

Lemme I.1.4 Soit E un ensemble non-vide et \mathcal{L} une classe monotone sur E . Alors, \mathcal{L} est une tribu ssi \mathcal{L} est stable par intersection finie.

Preuve : supposons que \mathcal{L} soit une classe monotone stable par intersection finie. Par définition, $E \in \mathcal{L}$. Puisque \mathcal{L} est stable par différence propre, \mathcal{L} est stable par passage au complémentaire. Soient $A, B \in \mathcal{L}$. Puisque $A \cup B = E \setminus ((E \setminus A) \cap (E \setminus B))$ et puisque \mathcal{L} est stable par intersection finie, on a $A \cup B \in \mathcal{L}$. Par conséquent, \mathcal{L} est stable par union finie.

Soient $B_n \in \mathcal{L}$, $n \in \mathbb{N}$, une suite quelconque d'éléments de \mathcal{L} . On pose $A_n = \bigcup_{0 \leq p \leq n} B_p$: on a donc $A_n \in \mathcal{L}$, d'après le raisonnement précédent. On remarque que $A_n \subset A_{n+1}$. comme \mathcal{L} est stable par union dénombrable croissante, $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}$. Or $\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n$. Cela montre que \mathcal{L} est stable par union dénombrable : c'est donc une tribu. La réciproque est triviale. ■

Il est facile de montrer qu'une intersection *quelconque* de classes monotones est encore une classe monotone. Plus précisément, on établit le lemme suivant dont la preuve, très simple, est laissée au lecteur.

Lemme I.1.5 Soit $(\mathcal{L}_i, i \in I)$, une famille de classes monotones sur E . Alors,

$$\bigcap_{i \in I} \mathcal{L}_i = \{A \subset E : \forall i \in I, A \in \mathcal{L}_i\}$$

est une classe monotone.

Cette propriété permet d'introduire la notion suivante.

Définition I.1.6 Soit $\mathcal{R} \subset \mathcal{P}(E)$, une classe quelconque de sous-ensembles de E . On pose

$$\lambda(\mathcal{R}) = \bigcap_{\substack{\mathcal{L} \text{ cl. monotone,} \\ \mathcal{R} \subset \mathcal{L}}} \mathcal{L} .$$

Alors $\lambda(\mathcal{R})$ est la *classe monotone engendrée par \mathcal{R}* , qui est la plus petite classe monotone contenant \mathcal{R} . □

Théorème I.1.6 ◀▶ (Classe monotone) Soit E , un ensemble non-vide. Soit \mathcal{L} , une classe monotone sur E . Soit \mathcal{P} , un pi-système sur E . On suppose que $\mathcal{P} \subset \mathcal{L}$. Alors $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Preuve : pour démontrer le théorème, il suffit de démontrer que $\lambda(\mathcal{P})$ est une tribu : en effet, par définition de la classe monotone engendrée, on a $\mathcal{P} \subset \lambda(\mathcal{P}) \subset \mathcal{L}$ et si $\lambda(\mathcal{P})$ est une tribu, on obtient $\sigma(\mathcal{P}) \subset \lambda(\mathcal{P})$. D'après le lemme précédent, il suffit donc de montrer que $\lambda(\mathcal{P})$ est stable par intersection finie.

La preuve de cette dernière assertion se fait en plusieurs étapes : pour tout $A \subset E$, on pose tout d'abord $\mathcal{L}_A = \{B \subset E : A \cap B \in \lambda(\mathcal{P})\}$. Montrons l'implication suivante :

$$(A \in \lambda(\mathcal{P})) \implies (\mathcal{L}_A \text{ est une classe monotone}) . \quad (\text{I.1})$$

En effet, supposons que $A \in \lambda(\mathcal{P})$; on observe tout d'abord que $E \in \mathcal{L}_A$. Soient $B, C \in \mathcal{L}_A$ tels que $B \subset C$; on observe que $A \cap (C \setminus B) = (A \cap C) \setminus (A \cap B)$ qui est donc dans $\lambda(\mathcal{P})$ puisque $A \cap C$ et $A \cap B$ sont dans $\lambda(\mathcal{P})$ (car $B, C \in \mathcal{L}_A$), puisque $A \cap B \subset A \cap C$, et puisque $\lambda(\mathcal{P})$ est stable par différence propre (en tant que classe monotone). Cela montre que \mathcal{L}_A est stable par différence propre. Soient $B_n \in \mathcal{L}_A$, $n \in \mathbb{N}$, tels que $B_n \subset B_{n+1}$; puisque $B_n \cap A \in \lambda(\mathcal{P})$ pour tout n , puisque la suite $B_n \cap A$ est croissante pour l'inclusion et puisque $\lambda(\mathcal{P})$ est stable par union dénombrable croissante (en tant que classe monotone), on a bien $A \cap \bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A \cap B_n \in \lambda(\mathcal{P})$. Cela montre que \mathcal{L}_A est stable par union dénombrable croissante, ce qui termine la preuve de (I.1).

Comme \mathcal{P} est un pi-système et comme $\mathcal{P} \subset \lambda(\mathcal{P})$ (par définition), pour tout $A \in \mathcal{P}$, on a $\mathcal{P} \subset \mathcal{L}_A$; on déduit de (I.1) que $\lambda(\mathcal{P}) \subset \mathcal{L}_A$ car $\lambda(\mathcal{P})$ est la plus petite classe monotone contenant \mathcal{P} . On a donc montré :

$$\forall A \in \mathcal{P}, \forall B \in \lambda(\mathcal{P}), \quad A \cap B \in \lambda(\mathcal{P}) .$$

Par conséquent $\mathcal{P} \subset \mathcal{L}_B$ pour tout B dans $\lambda(\mathcal{P})$ et donc $\lambda(\mathcal{P}) \subset \mathcal{L}_B$, pour tout $B \in \lambda(\mathcal{P})$. Autrement dit, pour tous $B, C \in \lambda(\mathcal{P})$, $B \cap C \in \lambda(\mathcal{P})$, ce qui montre que $\lambda(\mathcal{P})$ est stable par intersection finie et ce qui termine la preuve du théorème. ■

I.1.b Rappels sur les mesures positives.

Rappels sur la convergence des suites et des séries dans $[0, \infty]$. On adjoint ∞ à l'ensemble des réels positifs : l'ensemble résultant est noté $[0, \infty]$: l'ordre usuel s'étend naturellement à $[0, \infty]$ par $x \leq \infty$, pour tout $x \in [0, \infty]$. La notion de convergence vers ∞ est la définition habituelle. On adopte les conventions suivantes : $x + \infty = \infty + x = \infty$ et on rappelle brièvement les faits suivants.

Soit $a_n \in [0, \infty]$, $n \in \mathbb{N}$.

- Si $a_n \leq a_{n+1}$, $n \in \mathbb{N}$, alors la suite $(a_n)_{n \in \mathbb{N}}$ converge dans $[0, \infty]$ et

$$\lim_{n \rightarrow \infty} a_n = \sup_{n \in \mathbb{N}} a_n .$$

- Toute série de terme général positif converge dans $[0, \infty]$: il suffit d'appliquer le point précédent aux sommes partielles. On vérifie facilement que :

$$\sum_{n \in \mathbb{N}} a_n = \sup_{\substack{S \subset \mathbb{N} \\ S \text{ fini}}} \sum_{k \in S} a_k . \quad (\text{I.2})$$

- Toute série à terme positif est commutativement convergente, c'est-à-dire que pour toute bijection $\gamma : \mathbb{N} \rightarrow \mathbb{N}$, on a $\sum_{n \in \mathbb{N}} a_n = \sum_{n \in \mathbb{N}} a_{\gamma(n)}$. Cela se déduit facilement de (I.2).
- Soit $a_{n,p} \in [0, \infty]$, $n, p \in \mathbb{N}$, une suite double-indice. On a alors

$$\sum_{n \in \mathbb{N}} \left(\sum_{p \in \mathbb{N}} a_{n,p} \right) = \sum_{p \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{n,p} \right) . \quad (\text{I.3})$$

Preuve de (I.3) : on fixe N et P dans \mathbb{N} . On a

$$\begin{aligned} \sum_{0 \leq n \leq N} \left(\sum_{0 \leq p \leq P} a_{n,p} \right) &= \sum_{0 \leq p \leq P} \left(\sum_{0 \leq n \leq N} a_{n,p} \right) = \sum_{0 \leq n \leq N} a_{n,0} + \sum_{0 \leq n \leq N} a_{n,1} + \dots + \sum_{0 \leq n \leq N} a_{n,P} \\ &\leq \sum_{n \in \mathbb{N}} a_{n,0} + \sum_{n \in \mathbb{N}} a_{n,1} + \dots + \sum_{n \in \mathbb{N}} a_{n,P} = \sum_{0 \leq p \leq P} \left(\sum_{n \in \mathbb{N}} a_{n,p} \right) \\ &\leq \sum_{p \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{n,p} \right) . \end{aligned}$$

Dans le membre de gauche, on fait tendre d'abord P vers l'infini, puis N et on obtient

$$\sum_{n \in \mathbb{N}} \left(\sum_{p \in \mathbb{N}} a_{n,p} \right) \leq \sum_{p \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{n,p} \right) .$$

On obtient l'inégalité large contraire en raisonnant de même, ce qui implique le résultat voulu. ■

Définition I.1.7 ◀▶ Soit (E, \mathcal{E}) , un espace mesurable. Une application $\mu : \mathcal{E} \rightarrow [0, \infty]$ est une *mesure positive* si elle satisfait les propriétés suivantes.

- $\mu(\emptyset) = 0$.
- Pour toute suite $A_n \in \mathcal{E}$, $n \in \mathbb{N}$, formée d'ensembles deux-à-deux disjoints,

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

Cette dernière propriété est appelée propriété de *sigma additivité*. Le triplet (E, \mathcal{E}, μ) est appelé *espace mesuré*. □

Cette définition appelle quelques commentaires.

- Si $C, D \in \mathcal{E}$ sont disjoints, c'est-à-dire si $C \cap D = \emptyset$, la sigma-additivité appliquée à la suite $A_0 = C, A_1 = D$ et $A_n = \emptyset$, $n \geq 2$, implique que $\mu(C \cup D) = \mu(C) + \mu(D)$.

• Soient $A, B \in \mathcal{E}$ tels que $A \subset B$. On a $B = A \cup (B \setminus A)$. La propriété précédente implique que $\mu(B) = \mu(A) + \mu(B \setminus A)$. Donc

$$\forall A, B \in \mathcal{E} \text{ tels que } A \subset B, \quad \mu(A) \leq \mu(B). \quad (\text{I.4})$$

Autrement dit, *une mesure positive est une application croissante pour l'inclusion*. Notons que l'on ne peut pas dire grand chose de $\mu(B \setminus A)$ a priori si $\mu(A)$ et $\mu(B)$ sont des quantités infinies.

Proposition I.1.7 ◀▶ (Propriétés élémentaires des mesures) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $A_n \in \mathcal{E}$, $n \in \mathbb{N}$, une suite d'ensembles mesurables. Les assertions suivantes sont vérifiées.

- (i) (Croissance séquentielle) Si $A_n \subset A_{n+1}$, $n \in \mathbb{N}$, alors $\lim_n \uparrow \mu(A_n) = \mu(A)$, où $\bigcup A_n = A$.
- (ii) (Décroissance séquentielle) Si $A_{n+1} \subset A_n$, $n \in \mathbb{N}$, et si $\mu(A_0) < \infty$, alors $\lim_n \downarrow \mu(A_n) = \mu(B)$, où $\bigcap A_n = B$.
- (iii) (Sigma-sous-additivité) On ne suppose rien sur les A_n . Alors : $\mu(\bigcup A_n) \leq \sum \mu(A_n)$.
- (iv) Si $\mu(A_n) = 0$, $n \in \mathbb{N}$, alors $\mu(\bigcup A_n) = 0$.

▶◀ **Preuve** : on suppose tout d'abord que la suite $(A_n)_{n \in \mathbb{N}}$ est croissante pour l'inclusion et on prouve (i). On pose $B_0 = A_0$ et $B_n = A_n \setminus A_{n-1}$ pour tout $n \geq 1$. Il est clair que les B_n sont deux-à-deux disjoints. De plus, $\bigcup_{0 \leq k \leq n} B_k = A_n$ et $\bigcup_{n \in \mathbb{N}} B_n = A$. La sigma additivité de μ entraîne alors

$$\lim_{n \rightarrow \infty} \mu(A_n) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \mu(B_k) = \sum_{n \in \mathbb{N}} \mu(B_n) = \mu(A),$$

ce qui prouve (i).

Le point (ii) se déduit de (i) par passage au complémentaire : on pose $A'_n = A_0 \setminus A_n$ pour tout $n \geq 1$. Comme la suite $(A_n)_{n \in \mathbb{N}}$ est décroissante pour l'inclusion, la suite $(A'_n)_{n \in \mathbb{N}}$ est croissante pour l'inclusion. Par ailleurs, puisque $\mu(A_n) < \mu(A_0) < \infty$, on peut écrire $\mu(A'_n) = \mu(A_0) - \mu(A_n)$. On pose alors $A' = \bigcup A'_n$ et le point (i) implique

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A'_0) - \lim_{n \rightarrow \infty} \mu(A'_n) = \mu(A_0) - \mu(A') = \mu(A),$$

car $A' = A_0 \setminus A$ et donc $\mu(A) = \mu(A_0) - \mu(A')$.

Il reste à démontrer (iii) : pour tous $B, C \in \mathcal{E}$, on remarque d'une part que $B \cup C = B \cup (C \cap (E \setminus B))$ et d'autre part que B et $C \cap (E \setminus B)$ sont disjoints. Donc

$$\mu(B \cup C) = \mu(B) + \mu(C \cap (E \setminus B)) \leq \mu(B) + \mu(C),$$

car $C \cap (E \setminus B) \subset C$. En appliquant de façon répétée cette inégalité, on voit que $\mu(\bigcup_{0 \leq k \leq n} A_k) \leq \sum_{0 \leq k \leq n} \mu(A_k)$. Or la suite d'ensembles $\bigcup_{0 \leq k \leq n} A_k$ est croissante pour l'inclusion donc (i) implique que

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=0}^n A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=0}^n \mu(A_k),$$

qui entraîne bien le résultat désiré. Le point (iv) est une conséquence immédiate du point (iii). ■

Exemple I.1.7 Soit E , un ensemble non-vidé. Pour tout $A \subset E$, on pose $\mu(A) = 0$. On vérifie facilement que $\mu : \mathcal{P}(E) \rightarrow [0, \infty]$ est une mesure positive : c'est la *mesure nulle*. \square

Exemple I.1.8 Soit E , un ensemble non-vidé. Pour tout $A \subset E$ non-vidé, on pose $\mu(A) = \infty$. On pose également $\mu(\emptyset) = 0$. On vérifie facilement que $\mu : \mathcal{P}(E) \rightarrow [0, \infty]$ est une mesure positive : c'est la *mesure infinie*. \square

Exemple I.1.9 Soit E , un ensemble non-vidé. Soit $x \in E$. On définit la fonction $\delta_x : \mathcal{P}(E) \rightarrow [0, \infty]$ par

$$\forall A \subset E, \quad \delta_x(A) = 1 \text{ si } x \in A \quad \text{et} \quad \delta_x(A) = 0 \text{ si } x \notin A.$$

Il est facile de vérifier que δ_x est une mesure positive. Cette mesure est appelée *masse de Dirac en x* ; elle est définie sur la tribu $\mathcal{P}(E)$ de tous les sous-ensembles de E . \square

La proposition suivante donne plusieurs procédés simples permettant de créer des mesures.

Proposition I.1.8 Soit (E, \mathcal{E}) , un ensemble mesurable.

(i) Soit $\mu_n : \mathcal{E} \rightarrow [0, \infty]$, $n \in \mathbb{N}$, une suite de mesures positives. Soit $c_n \in]0, \infty[$, $n \in \mathbb{N}$. Pour tout $A \in \mathcal{E}$, on pose $\mu(A) = \sum_{n \in \mathbb{N}} c_n \mu_n(A)$. Alors $\mu : \mathcal{E} \rightarrow [0, \infty]$ est une mesure positive notée $\mu = \sum c_n \mu_n$.

(ii) Soit $\mu : \mathcal{E} \rightarrow [0, \infty]$, une mesure positive et soit $B \in \mathcal{E}$. Pour tout $A \in \mathcal{E}$, on pose $\nu(A) = \mu(B \cap A)$. Alors ν est une mesure positive appelée *mesure-trace de μ sur B* .

Preuve : il est clair que $\mu(\emptyset) = 0$. Soit $A_p \in \mathcal{E}$, $p \in \mathbb{N}$, une suite d'ensembles disjoints deux-à-deux. La sigma additivité des mesures μ_n et (I.3) impliquent

$$\mu\left(\bigcup_{p \in \mathbb{N}} A_p\right) = \sum_{n \in \mathbb{N}} c_n \mu_n\left(\bigcup_{p \in \mathbb{N}} A_p\right) = \sum_{n \in \mathbb{N}} \sum_{p \in \mathbb{N}} c_n \mu_n(A_p) = \sum_{p \in \mathbb{N}} \sum_{n \in \mathbb{N}} c_n \mu_n(A_p) = \sum_{p \in \mathbb{N}} \mu(A_p),$$

ce qui prouve (i). La preuve de (ii), très simple, est laissée au lecteur. \blacksquare

◀► **Vocabulaire.** Soit (E, \mathcal{E}, μ) , un espace mesuré.

- Si $\mu(E) = 1$, la mesure μ est appelée *loi* ou *mesure de probabilité*. L'espace (E, \mathcal{E}, μ) est appelé *espace de probabilité*.
- On appelle $\mu(E)$ la *masse de la mesure μ* . Si $\mu(E) < \infty$ on parle alors de *mesure finie* (ou encore de mesure de masse finie).
- La mesure μ est dite *sigma finie* s'il existe une suite d'ensembles mesurables $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, tels que

$$\forall n \in \mathbb{N}, \mu(E_n) < \infty \quad \text{et} \quad \bigcup_{n \in \mathbb{N}} E_n = E.$$

- Un point $x \in E$ est un *atome* de μ si $\mu(\{x\}) > 0$. La quantité $\mu(\{x\})$ est la masse de l'atome x . On remarque que cette notion n'a de sens que si les singletons sont dans la tribu \mathcal{E} , ce qui n'est a priori pas toujours le cas.
- Une mesure sans atome est dite *diffuse*.

Le lemme suivant donne deux formulations équivalentes de la propriété d'être sigma-finie pour une mesure.

Lemme I.1.9 Soit (E, \mathcal{E}, μ) , un espace mesuré. Les assertions suivantes sont équivalentes.

(a) μ est sigma-finie.

(b) Il existe $A_n \in \mathcal{E}$, $n \in \mathbb{N}$, $A_n \subset A_{n+1}$, $\mu(A_n) < \infty$ et $\bigcup_{n \in \mathbb{N}} A_n = E$.

(c) Il existe $B_n \in \mathcal{E}$, $n \in \mathbb{N}$, deux à deux disjoints tels que $\mu(B_n) < \infty$ et $\bigcup_{n \in \mathbb{N}} B_n = E$.

De plus, μ probabilité $\implies \mu$ finie $\implies \mu$ sigma-finie $\implies \mu$ somme de mesures finies.

Preuve : il est clair que (b) ou (c) impliquent (a). Montrons (a) \implies (b) : soit $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, tels que $\bigcup_{n \in \mathbb{N}} E_n = E$ et $\mu(E_n) < \infty$. On pose alors $A_n = \bigcup_{0 \leq k \leq n} E_k$. On a bien $A_n \in \mathcal{E}$, $\bigcup_{n \in \mathbb{N}} A_n = E$ et $A_n \subset A_{n+1}$. Par sous-additivité des mesures, $\mu(A_n) \leq \sum_{0 \leq k \leq n} \mu(E_k) < \infty$, ce qui prouve (b).

On pose ensuite $B_0 = A_0$ et pour tout $n \geq 1$, $B_n = A_n \setminus A_{n-1}$. Il est clair que les B_n sont dans \mathcal{E} et sont deux-à-deux disjoints. De plus $A_n = \bigcup_{0 \leq k \leq n} B_k$, ce qui implique d'une part que $\bigcup_{n \in \mathbb{N}} B_n = E$. Comme $B_n \subset A_n$, on a d'autre part $\mu(B_n) \leq \mu(A_n) < \infty$, ce qui termine la preuve de (c). ■

Nous rappelons, sans preuve, le théorème d'existence et d'unicité de la mesure de Lebesgue sur \mathbb{R} et plus généralement des mesures de Stieltjes.

Théorème I.1.10 ◀▶ On rappelle que $\mathcal{B}(\mathbb{R})$ désigne la tribu des Boréliens de \mathbb{R} .

(i) Il existe une **unique** mesure positive $\ell : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$, appelée **mesure de Lebesgue**, telle que pour tous $a < b$, on ait $\ell(]a, b]) = b - a$.

(ii) Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une application croissante et continue à droite. Il existe une **unique** mesure positive $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$ telle que pour tous $a < b$, on ait $\mu(]a, b]) = F(b) - F(a)$. Cette mesure est appelée **mesure de Stieltjes associée à F** et elle est souvent notée dF .

Remarque I.1.3 On voit que la mesure de Lebesgue ℓ est la mesure de Stieltjes associée à la fonction identité : $F(x) = x$, $x \in \mathbb{R}$. □

Remarque I.1.4 Dans le théorème I.1.10, on suppose F croissante et continue à droite. C'est en effet une nécessité : supposons que $F : \mathbb{R} \rightarrow \mathbb{R}$ soit une fonction telle qu'il existe une mesure positive $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$ telle que $\mu(]a, b]) = F(b) - F(a)$, pour tous $a < b$. On voit immédiatement que $F(a) \leq F(b)$ dès que $a < b$, ce qui montre que F est croissante. Fixons a ; soit $h_n > 0$, $n \in \mathbb{N}$, une suite décroissante tendant vers 0; on voit que $F(a + h_n)$, $n \in \mathbb{N}$, est une suite décroissante minorée par $F(a)$ car F croît; sa limite est la limite à droite de F , notée $F(a+)$. Or on observe que d'une part $F(a + h_n) - F(a) = \mu(]a, a + h_n])$. De plus on a $]a, a + h_{n+1}] \subset]a, a + h_n]$ et $\bigcap_{n \geq 0}]a, a + h_n] = \emptyset$. Comme $\mu(]a, a + h_0]) = F(a + h_0) - F(a) < \infty$. La proposition I.1.7 (ii) implique que $\lim_n \downarrow \mu(]a, a + h_n]) = 0$, ce qui montre que $F(a) = \lim_n F(a + h_n) = F(a+)$. En d'autres termes, F est continue à droite. □

Les mesures de Stieltjes permettent d'obtenir une grande variété de mesures sur $\mathcal{B}(\mathbb{R})$. Si on prend par exemple $F(x) = \lfloor x \rfloor$, la fonction partie entière, il est facile de voir que $dF = \sum_{n \in \mathbb{Z}} \delta_n$.

Nous rappelons ensuite l'existence et l'unicité de la mesure de Lebesgue en dimension n qui correspond à la mesure des surfaces, volumes etc.

Théorème I.1.11 ◀▶ Soit $n \in \mathbb{N}^*$. Il existe une unique mesure positive $\ell_n : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty]$ telle que pour tout pavé ouvert $R =]a_1, b_1[\times \dots \times]a_n, b_n[$, on a

$$\ell_n(R) = (b_1 - a_1) \dots (b_n - a_n).$$

Pourvu que l'on ait prouvé l'existence de la mesure de Lebesgue (dans \mathbb{R} ou \mathbb{R}^n), son unicité est une conséquence du théorème suivant montrant qu'il suffit de connaître une mesure sur un pi-système pour la caractériser.

Théorème I.1.12 ◀▶ (Unicité de prolongement des mesures) Soit (E, \mathcal{E}) , un espace mesurable. Soit \mathcal{P} , un pi-système engendrant \mathcal{E} . Soient μ_1 et μ_2 , deux mesures sur (E, \mathcal{E}) qui coïncident sur \mathcal{P} :

$$\forall A \in \mathcal{P}, \quad \mu_1(A) = \mu_2(A) .$$

Alors, les assertions suivantes sont vérifiées.

(i) Si $\mu_1(E) = \mu_2(E) < \infty$, alors $\mu_1 = \mu_2$.

(ii) S'il existe $E_n \in \mathcal{P}$, $n \in \mathbb{N}$, tels que $\mu_1(E_n) = \mu_2(E_n) < \infty$, tels que $E_n \subset E_{n+1}$ et tels que $E = \bigcup_{n \in \mathbb{N}} E_n$, alors $\mu_1 = \mu_2$.

◀▶ **Preuve** : on prouve d'abord (i). On pose $\mathcal{L} = \{B \in \mathcal{E} : \mu_1(B) = \mu_2(B)\}$. Par hypothèse, on a $\mathcal{P} \subset \mathcal{L}$. Montrons ensuite que \mathcal{L} est une classe monotone : on voit tout d'abord que $E \in \mathcal{L}$. Soient $B, C \in \mathcal{L}$ tels que $B \subset C$. Comme μ_1 et μ_2 sont supposées finies, on a

$$\mu_1(C \setminus B) = \mu_1(C) - \mu_1(B) = \mu_2(C) - \mu_2(B) = \mu_2(C \setminus B),$$

qui entraîne bien que $C \setminus B \in \mathcal{L}$. Soit $B_n \in \mathcal{L}$, $n \in \mathbb{N}$, une suite d'ensembles tels que $B_n \subset B_{n+1}$. Par la proposition I.1.7 (i),

$$\mu_1\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \mu_1(B_n) = \lim_{n \rightarrow \infty} \mu_2(B_n) = \mu_2\left(\bigcup_{n \in \mathbb{N}} B_n\right),$$

ce qui montre que $\bigcup B_n \in \mathcal{L}$. La classe \mathcal{L} est donc une classe monotone et le théorème I.1.6 de la classe monotone entraîne que $\mathcal{E} = \sigma(\mathcal{P}) \subset \mathcal{L} \subset \mathcal{E}$. Donc $\mathcal{L} = \mathcal{E}$, qui prouve (i).

Montrons (ii) : pour tout $i \in \{1, 2\}$, on pose $\mu_{i,n}(B) = \mu_i(B \cap E_n)$, $B \in \mathcal{E}$. Par la proposition I.1.8 (iii), $\mu_{i,n}$ est une mesure. Comme $E_n \in \mathcal{P}$, $\mu_{1,n}$ et $\mu_{2,n}$ coïncident sur \mathcal{P} et par le point (i) elles sont égales. Soit $B \in \mathcal{E}$. On a $B \cap E_n \subset B \cap E_{n+1}$, et que $\bigcup B \cap E_n = B$. Par la proposition I.1.7 (i)

$$\mu_1(B) = \lim_{n \rightarrow \infty} \mu_1(B \cap E_n) = \lim_{n \rightarrow \infty} \mu_{1,n}(B) = \lim_{n \rightarrow \infty} \mu_{2,n}(B) = \lim_{n \rightarrow \infty} \mu_2(B \cap E_n) = \mu_2(B) ,$$

ce qui termine la preuve. ■

En application du théorème précédent, montrons l'unicité de la mesure de Lebesgue sur \mathbb{R} : on pose

$$\mathcal{P} = \{\mathbb{R}\} \cup \{]a, b]; a, b \in \mathbb{R}, a \leq b\} .$$

Il est facile de vérifier que \mathcal{P} est un pi-système générant $\mathcal{B}(\mathbb{R})$. Soit $\lambda : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}$, une mesure telle que $\lambda(]a, b]) = \ell(]a, b]) = b - a$, pour tous $a \leq b$. On pose $E_n =]-n, n]$, $n \in \mathbb{N}$. On a bien $\ell(\mathbb{R}) = \lambda(\mathbb{R}) = \infty = \lim_{n \rightarrow \infty} \lambda(E_n) = \lim_{n \rightarrow \infty} \ell(E_n)$. On a donc $E_n \subset E_{n+1}$, $\bigcup E_n = \mathbb{R}$ et pour tout $n \in \mathbb{N}$, $\lambda(E_n) = \ell(E_n)$; de plus λ et ℓ coïncident sur \mathcal{P} : le théorème I.1.12 (ii) implique donc que $\lambda = \ell$.

Exemple I.1.10 Afin de montrer l'importance des hypothèses du point (ii) du théorème I.1.12 dans le cas des mesures de masse infinies. On discute du contreexemple suivant : on pose

$$\mathcal{P} = \{\mathbb{R}\} \cup \{]-\infty, a]; a \in \mathbb{R}\} ;$$

c'est un pi-système et on a montré au lemme I.1.2 page 4 que $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$; on pose ensuite $\mu_1 = \ell$ et $\mu_2 = 2\ell$. Ces deux mesures coïncident sur \mathcal{P} et pourtant $\mu_1 \neq \mu_2$. □

I.2 Premières notions de probabilité.

I.2.a Ce que modélise un espace de probabilité.

Définition I.2.1 ◀▶ (*Espace de probabilité*) Soit (E, \mathcal{E}, μ) , un espace mesuré. C'est un espace de probabilité si la mesure positive μ est de masse totale 1, c'est-à-dire $\mu(E) = 1$. La mesure μ est appelée mesure de probabilité ou loi. \square

◀▶ **Interprétation, notations.** Un espace de probabilité est en général plutôt noté de la manière suivante :

$$(\Omega, \mathcal{F}, \mathbf{P}).$$

Il modélise un phénomène ou une expérience aléatoire de la manière suivante.

- (a) Ω est un ensemble qui représente tous les résultats possibles du phénomène aléatoire.
- (b) Les sous-ensembles de Ω qui constituent la tribu \mathcal{F} sont appelés les *événements*. Un événement $B \in \mathcal{F}$ est décrit par tous les résultats qui le réalisent. Autrement dit, un résultat possible particulier de l'expérience aléatoire $\omega \in \Omega$ réalise l'événement B si et seulement si $\omega \in B$. C'est-à-dire que la notation mathématique $\omega \in B$ s'interprète par la phrase « ω réalise l'événement B ».
- (c) À chaque événement $B \in \mathcal{F}$, le nombre $\mathbf{P}(B) \in [0, 1]$ représente la probabilité que l'événement B a de survenir lors du phénomène aléatoire que l'on étudie. Autrement dit, *l'événement B a $\mathbf{P}(B)\%$ de chance de survenir.*

Exemple I.2.1 (*Le jet d'un dé à 6 faces*) Considérons le jet d'un dé à six faces (le dé est bien secoué dans un gobelet avant d'être lancé sur le tapis). On ne s'intéresse qu'au résultat final du jet de dé, c'est-à-dire au chiffre montré par la face supérieure du dé lorsqu'il s'immobilise. On peut donc prendre $\Omega = \{1, \dots, 6\}$, qui représente les six résultats possibles de cette expérience aléatoire (ce choix n'est en rien unique). La tribu des événements est $\mathcal{P}(\Omega)$, la classe de tous les sous-ensembles de $\Omega = \{1, \dots, 6\}$. Par exemple, l'événement "*le résultat est pair*" est modélisé par l'ensemble $\{2, 4, 6\}$ et l'événement "*le résultat est divisible par trois*" est modélisé par l'ensemble $\{3, 6\}$. On considère que le dé n'est pas pipé : chaque résultat a donc la même probabilité d'apparaître (pourquoi en serait-il autrement ?) Chaque chiffre a une chance sur six de se produire. La mesure \mathbf{P} est donc satisfaisante

$$\forall \omega \in \{1, \dots, 6\}, \quad \mathbf{P}(\{\omega\}) = \frac{1}{6}.$$

On voit alors que pour tout $B \subset \{1, \dots, 6\}$

$$\mathbf{P}(B) = \mathbf{P}\left(\bigcup_{\omega \in B} \{\omega\}\right) = \sum_{\omega \in B} \mathbf{P}(\{\omega\}) = \frac{1}{6} \#B,$$

c'est-à-dire

$$\mathbf{P} = \frac{1}{6}(\delta_1 + \dots + \delta_6)$$

qui est la mesure de probabilité uniforme sur $\{1, \dots, 6\}$. On a donc

$$\text{Probabilité d'avoir un résultat pair} = \mathbf{P}(\{2, 4, 6\}) = \frac{1}{2},$$

ce qui est conforme à l'intuition. \square

Très bref historique. Les anciens avaient imaginé de nombreux jeux de hasard (que l'on se rappelle la phrase de César au franchissement du Rubicon : *alea jacta est*, les dés sont jetés). Le mot hasard nous vient de l'arabe. Ni les croyances et ni les philosophies antiques et médiévales n'ignoraient les phénomènes aléatoires. Ce n'est cependant qu'à l'époque moderne que certains phénomènes aléatoires ont été envisagés de manière objective et quantifiable. Il semble qu'un des premiers scientifiques à avoir conçu et éprouvé une première théorie de l'aléatoire soit le physicien Galilée dans son *Traité des dés*. Les différents concepts de probabilité ont progressivement émergés à partir du XVIIème siècle. La formalisation et la mathématisation du hasard s'étale sur trois cents ans et aboutit à une théorie rigoureuse grâce notamment aux travaux d'Emile Borel, d'Henri Lebesgue et de Nikolai Kolmogorov. Cette théorie n'est pas l'unique façon d'appréhender l'aléatoire : les systèmes dynamiques chaotiques et l'aléa de la mécanique quantiques sont de nature différente et ne sont pas vraiment réductibles à la théorie classique des probabilités, dont l'exposé est l'objet de ce cours. L'histoire des probabilités pose des problèmes épistémologiques nombreux et intéressants mais nous n'en parlerons pas dans ce cours. Disons qu'en substance, la théorie moderne des probabilités s'appuie sur la théorie générale de l'intégration et développe une notion propre : l'*indépendance*, ainsi que les nombreux concepts d'analyse en dimension infinie qui en découlent. Elle forme en cela une branche des mathématiques pures. Par ailleurs, ses résultats ont très souvent un rapport étroit avec la réalité : ils se vérifient expérimentalement et donnent lieu à des interprétations effectives : la théorie des probabilités forme donc une branche très utile des mathématiques appliquées et elle est fortement liée dans ses applications à la statistique mathématique.

◀► **Vocabulaire.** Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. On introduit le vocabulaire suivant.

- L'événement \emptyset est l'événement impossible.
- L'événement Ω est l'événement certain.
- Si $A \in \mathcal{F}$, est tel que $\mathbf{P}(A) = 1$, on dit que l'événement A est *presque sûr*.
- On rappelle que la notation mathématique $\omega \in A$ s'interprète par « ω réalise l'événement A ».
- Si $A \in \mathcal{F}$, alors $\Omega \setminus A$ est « l'événement contraire de A ».
- Soient $A, B \in \mathcal{F}$, deux événements. L'événement $A \cup B$ se comprend comme l'événement " A ou B " et l'événement $A \cap B$ se comprend comme l'événement " A et B ". On s'autorisera souvent à utiliser les mots "ou" et "et" dans les calculs :

$$\mathbf{P}(A \text{ ou } B) = \mathbf{P}(A \cup B) \quad \text{et} \quad \mathbf{P}(A \text{ et } B) = \mathbf{P}(A \cap B).$$

Très souvent, le « et » (c'est-à-dire l'intersection ensembliste « \cap ») est remplacé par un point virgule « ; » :

$$\mathbf{P}(A ; B) = \mathbf{P}(A \text{ et } B) = \mathbf{P}(A \cap B).$$

◀► **Propriétés d'un espace de probabilité.** Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Les propriétés de \mathbf{P} incluent celles de toute mesure positive. Nous les ré-exposons bien que cela soit redondant avec la section I.1.b sur les mesures positives. On se donne $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, une suite d'événements de la tribu \mathcal{F} .

Propriété I.2.1 ◀► (*Sigma-additivité*) Si les A_n sont disjoints deux-à-deux, alors

$$\mathbf{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n). \quad (\text{I.5})$$

◀► **Preuve** : puisque \mathbf{P} est une mesure positive cette propriété fait partie de sa définition. ■

Propriété I.2.2 ▶◀ (*Additivité*) On suppose $A, B \in \mathcal{F}$ disjoints. Alors

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) . \quad (\text{I.6})$$

◀► **Preuve** : C est une conséquence de la propriété de sigma-additivité avec $A_0 = A$, $A_1 = B$ et $A_n = \emptyset$, pour tout $n \geq 2$. ■

Propriété I.2.3 ▶◀ (*Croissance pour l'inclusion*) On suppose $C, D \in \mathcal{F}$ tels que $C \subset D$:

$$\mathbf{P}(D \setminus C) = \mathbf{P}(D) - \mathbf{P}(C) \quad \text{et donc} \quad \mathbf{P}(C) \leq \mathbf{P}(D) . \quad (\text{I.7})$$

◀► **Preuve** : on applique l'additivité aux deux événements disjoints C et $D \setminus C$ dont l'union est D . ■

Propriété I.2.4 ▶◀ (*Formule des probabilités totales*) Soit $B \in \mathcal{F}$. On suppose que les A_n forment une *partition* de Ω , c'est-à-dire qu'ils sont disjoints deux-à-deux et que $\bigcup_{n \in \mathbb{N}} A_n = \Omega$. Alors

$$\mathbf{P}(B) = \sum_{n \in \mathbb{N}} \mathbf{P}(B \cap A_n) . \quad (\text{I.8})$$

◀► **Preuve** : soient deux entiers $m \neq n$. On a

$$(B \cap A_m) \cap (B \cap A_n) = B \cap (A_m \cap A_n) = B \cap \emptyset = \emptyset ,$$

car les A_n sont supposés deux-à-deux disjoints. Donc les événements $(B \cap A_n)_{n \in \mathbb{N}}$ sont deux-à-deux disjoints. La sigma-additivité de \mathbf{P} implique que $\mathbf{P}(\bigcup_{n \in \mathbb{N}} B \cap A_n) = \sum_{n \in \mathbb{N}} \mathbf{P}(B \cap A_n)$. On rappelle ensuite la propriété ensembliste élémentaire suivante : $\bigcup_{n \in \mathbb{N}} B \cap A_n = B \cap (\bigcup_{n \in \mathbb{N}} A_n)$. Or $\bigcup_{n \in \mathbb{N}} A_n = \Omega$ donc $\bigcup_{n \in \mathbb{N}} B \cap A_n = B \cap \Omega = B$, ce qui termine la preuve de la formule des probabilités totales. ■

Propriété I.2.5 ▶◀ (*Croissance séquentielle*) on suppose $A_n \subset A_{n+1}$, $n \in \mathbb{N}$. Alors

$$\mathbf{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \uparrow \mathbf{P}(A_n) . \quad (\text{I.9})$$

◀► **Preuve** : il s'agit de la propriété (i) de la proposition I.1.7 qui est prouvée page 8. ■

Propriété I.2.6 ▶◀ (*Décroissance séquentielle*) On suppose $A_{n+1} \subset A_n$, $n \in \mathbb{N}$. Alors

$$\mathbf{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \downarrow \mathbf{P}(A_n) . \quad (\text{I.10})$$

◀► **Preuve** : il s'agit de la propriété (ii) de la proposition I.1.7 qui est prouvée page 8. ■

Propriété I.2.7 ▶◀ (*Sigma-sous-additivité*) On ne suppose rien sur les $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. Alors,

$$\mathbf{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mathbf{P}(A_n). \quad (\text{I.11})$$

◀► **Preuve** : il s'agit de la propriété (iii) de la proposition I.1.7 qui est prouvée page 8. ■

Propriété I.2.8 ▶◀ Soient $A, B, C \in \mathcal{F}$. On montre facilement que

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \quad (\text{I.12})$$

et aussi

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) + \mathbf{P}(A \cap B \cap C). \quad (\text{I.13})$$

◀► **Preuve** : les trois événements $A_1 = A \cap B$, $A_2 = A \setminus (A \cap B)$ et $A_3 = B \setminus (A \cap B)$ sont dans \mathcal{F} et sont disjoints deux-à-deux. On a $A_1 \cup A_2 = A$ et $B = A_1 \cup A_3$ et donc $A \cup B = A_1 \cup A_2 \cup A_3$. Par additivité de \mathbf{P} , $\mathbf{P}(A \cup B) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3)$. Or par la propriété I.2.3, $\mathbf{P}(A_2) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$ et $\mathbf{P}(A_3) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$ donc

$$\mathbf{P}(A \cup B) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) = \mathbf{P}(A_1) + \mathbf{P}(A) - \mathbf{P}(A \cap B) + \mathbf{P}(B) - \mathbf{P}(A \cap B),$$

ce qui prouve (I.12) car $A_1 = A \cap B$. On laisse en exercice la preuve directe de (I.13). ■

Plus généralement, on a le résultat suivant que l'on montre plus loin dans le cours.

Lemme I.2.1 (Formule du crible) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A_1, \dots, A_n \in \mathcal{F}$. Alors

$$\mathbf{P}\left(\bigcup_{1 \leq j \leq n} A_j\right) = \sum_{\emptyset \neq J \subset \{1, \dots, n\}} (-1)^{1+\#J} \mathbf{P}\left(\bigcap_{j \in J} A_j\right).$$

Preuve : plusieurs preuves sont possible, la plus élémentaire consistant en une récurrence. Voir l'exercice I.3.3, page 24 et son corrigé. ■

I.2.b Exemples, équi-probabilité et dénombrement.

On considère un espace fini $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega) := \{A : A \subset \Omega\}$, et de la probabilité $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$ avec

$$\mathbf{P}(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{n}, \quad \forall A \in \mathcal{F}.$$

Cette probabilité s'appelle *équi-probabilité* sur Ω : elle ne privilège aucun élément particulier de Ω .

Exemple I.2.2 (*Un jeu de pile ou face*) On jette trois fois une pièce de monnaie parfaite. On peut représenter l'espace Ω comme l'ensemble des applications de $\{1, 2, 3\}$ (trois jets) dans $\{P, F\}$ ($P =$ pile, $F =$ face). Autrement dit, $\Omega = \{P, F\}^3$, donc $n = \#(\Omega) = 2^3 = 8$.

Il est alors facile de voir par exemple que

$$\mathbf{P}(\text{on obtient exactement une fois } P) = 3/8$$

$$\mathbf{P}(\text{on obtient au moins une fois } P) = 1 - \mathbf{P}(\text{on obtient trois fois } F) = 1 - 1/8 = 7/8.$$

□

Exemple I.2.3 (*Le Loto*) Le jeu du loto consiste à choisir 6 numéros distincts parmi $\{1, 2, \dots, 49\}$. On suppose que les boules qui portent les 49 numéros sont toutes parfaites. Ici on ne s'intéresse qu'aux résultats des 6 boules, sans leur ordre de tirage. L'espace Ω est

$$\Omega = \{\{a_1, a_2, \dots, a_6\} : 1 \leq a_i \leq 49, \text{ les } a_i \text{ sont distincts}\}.$$

On a donc $n := \#\Omega = \binom{49}{6}$. Par conséquent,

$$\mathbf{P}(\text{on gagne le premier prix avec un bulletin}) = \frac{1}{\binom{49}{6}} \approx 7 \times 10^{-8},$$

ce qui fait réfléchir.

□

Exemple I.2.4 (*Permutation aléatoires*) On considère l'arrivée d'une course de chevaux, avec dix partants, numérotés de 1 à 10. On note l'ordre d'arrivée. On suppose que les concurrents sont de force égale et qu'il n'y a pas d'ex-aequo. L'espace Ω est l'ensemble des permutations de $\{1, 2, \dots, 10\}$. Donc $n = \#\Omega = 10!$. On a alors

$$\begin{aligned} \mathbf{P}(\text{le numéro 10 arrive dernier}) &= \frac{\#\{\omega \in \Omega : \omega(10) = 10\}}{10} \\ &= \frac{\#\{\text{permutations de } \{1, \dots, 9\}\}}{10!} \\ &= 9!/10! = 1/10. \end{aligned}$$

Cela est conforme à l'intuition. Si l'on s'intéresse à l'événement

$$A = \{\text{le numéro 10 arrive dans les trois premiers}\},$$

alors on peut considérer $A_k = \{\text{le numéro 10 arrive à la } k\text{-ième place}\}$ ($k = 1, 2$ ou 3). Les A_k sont deux-à-deux disjoints, on a

$$\mathbf{P}(A) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) = 3/10.$$

□

Exemple I.2.5 (*Permutations aléatoires*) Imaginons l'expérience aléatoire suivante : on dispose de n tiroirs numérotés de 1 à n , et de n cartes numérotées de 1 à n ; dans chaque tiroir le meneur de jeu dépose une et une seule carte. On suppose qu'il le fait totalement au hasard. *Quelle est la probabilité qu'aucun tiroir ne contienne une carte ayant le même numéro ?*

On modélise facilement cette situation de la manière suivante : on note $\sigma(k)$ le numéro de la carte contenue dans le tiroir k . On voit que $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ est une bijection, c'est-à-dire une permutation de $\{1, \dots, n\}$. On note \mathbf{S}_n l'ensemble des permutations de $\{1, \dots, n\}$. Le

résultat de l'expérience aléatoire se modélise par une permutation : on peut donc prendre $\Omega := \mathbf{S}_n$. Les événements relatifs à l'expérience peuvent être n'importe quel sous-ensemble de \mathbf{S}_n . On prend donc $\mathcal{F} := \mathcal{P}(\mathbf{S}_n)$ qui est la tribu de tous les sous-ensembles de \mathbf{S}_n . Enfin, on suppose que les cartes sont distribuées totalement au hasard, donc la probabilité d'obtenir une permutation spécifique $\gamma \in \mathbf{S}_n$ ne dépend pas de γ . La probabilité \mathbf{P} sur (Ω, \mathcal{F}) correspondant à l'expérience est la probabilité uniforme :

$$\forall \gamma \in \mathbf{S}_n, \quad \mathbf{P}(\{\gamma\}) = 1/n! .$$

On a donc

$$\forall B \subset \mathbf{S}_n, \quad \mathbf{P}(B) = \#B/n! .$$

Autrement dit $\mathbf{P} = \frac{1}{n!} \sum_{\gamma \in \mathbf{S}_n} \delta_\gamma$. On s'intéresse à l'événement

$$\begin{aligned} D &:= \text{« aucun tiroir ne contient une carte de même numéro »} \\ &= \{\gamma \in \mathbf{S}_n : \forall j \in \{1, \dots, n\}, \gamma(j) \neq j\}. \end{aligned}$$

Les permutations n'ayant aucun point fixe sont appelées des *dérangements* et D est donc l'ensemble des dérangements. On répond donc à la question en calculant $\mathbf{P}(D)$.

Pour cela, pour tout $j \in \{1, \dots, n\}$, on pose

$$A_j = \{\gamma \in \mathbf{S}_n : \gamma(j) = j\}.$$

A_j est l'événement « le tiroir j renferme la carte numéro j ». On a donc

$$\Omega \setminus D = \bigcup_{1 \leq j \leq n} A_j ,$$

si bien que

$$\mathbf{P}(D) = 1 - \mathbf{P}\left(\bigcup_{1 \leq j \leq n} A_j\right) .$$

On utilise la formule du crible (le lemme I.2.1 page 15) pour calculer cette probabilité. Pour tout $J \subset \{1, \dots, n\}$ non-vidé, on remarque que

$$\#\bigcap_{j \in J} A_j = \#\{\gamma \in \mathbf{S}_n : \forall j \in J, \gamma(j) = j\} = (n - \#J)! .$$

On rappelle qu'il y a $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ sous-ensemble $J \subset \{1, \dots, n\}$ tels que $\#J = k$. Par conséquent,

$$\begin{aligned} \mathbf{P}(D) &= 1 + \sum_{\emptyset \neq J \subset \{1, \dots, n\}} (-1)^{\#J} \mathbf{P}\left(\bigcap_{j \in J} A_j\right) = 1 + \sum_{\emptyset \neq J \subset \{1, \dots, n\}} (-1)^{\#J} \frac{1}{n!} \#\bigcap_{j \in J} A_j \\ &= 1 + \sum_{1 \leq k \leq n} (-1)^k \binom{n}{k} \frac{1}{n!} (n-k)! = \sum_{0 \leq k \leq n} \frac{(-1)^k}{k!} \end{aligned} \quad (\text{I.14})$$

On voit notamment que

$$|\mathbf{P}(D) - e^{-1}| \leq \frac{1}{(n+1)!} ,$$

donc $\mathbf{P}(D)$ est très proche de e^{-1} qui vaut environ 36%. Ce résultat est remarquable car il dépend peu de n . \square

I.2.c Probabilités conditionnelles.

Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité modélisant un phénomène aléatoire. Soit $A \in \mathcal{F}$, un événement tel que $\mathbf{P}(A) > 0$. Dans le phénomène aléatoire que l'on observe on possède une information : *on sait que l'événement A s'est produit*. Cela modifie évidemment les probabilités des autres événements qui peuvent survenir car ces événements peuvent être influencés par A . Par exemple, considérons l'expérience aléatoire qui consiste à sélectionner au hasard un habitant d'un immeuble et à lui demander son âge ; dans cette expérience on cherche à évaluer la probabilité de l'événement $B =$ « la personne sélectionnée a plus de 85 ans ». Supposons que l'on ait une information supplémentaire représenté par l'événement $A =$ « la personne sélectionnée au hasard dans l'immeuble pratique le triathlon ». On voit que l'événement A influe sur l'événement B et que la probabilité de B sachant A est bien plus faible que la probabilité de B sans aucun renseignement (en revanche si l'événement A est « le nombre de lettres du prénom et du nom de la personne sélectionnée est pair », cette information n'a aucune influence sur la probabilité que B ait de survenir).

Cherchons maintenant comment la connaissance de l'événement A modifie les probabilités des autres événements : l'événement certain est donc A et pour tout autre événement B , évaluer la probabilité B sachant A est la même chose que d'évaluer la probabilité de $A \cap B$. Cela conduit à définir la probabilité de B sachant A comme la quantité $\mathbf{P}(A \cap B)/\mathbf{P}(A)$. On note cette quantité par $\mathbf{P}(\cdot|A)$. On vérifie dans la proposition-définition suivante qu'il s'agit bien d'une mesure de probabilité.

Proposition I.2.2 ◀► Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité Soit $A \in \mathcal{F}$, un événement tel que $\mathbf{P}(A) > 0$. on définit la $\mathbf{P}(\cdot|A) : \mathcal{F} \rightarrow [0, 1]$ par

$$\forall B \in \mathcal{F}, \quad \mathbf{P}(B|A) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)}.$$

On vérifie que $\mathbf{P}(\cdot|A)$ est une mesure de probabilité. Cette probabilité $\mathbf{P}(\cdot|A)$ est appelée probabilité sachant A , ou conditionnellement à A .

◀► **Preuve** : il est clair que $\mathbf{P}(\emptyset|A) = 0$. Par ailleurs, comme $\Omega \cap A = A$, on a $\mathbf{P}(\Omega|A) = \mathbf{P}(A)/\mathbf{P}(A) = 1$. Enfin, soit $B_n \in \mathcal{F}$, $n \in \mathbb{N}$, une suite d'événements disjoints deux-à-deux ; pour simplifier les notations on pose $B = \bigcup_{n \in \mathbb{N}} B_n$; on vérifie alors que les événements $B_n \cap A$, $n \in \mathbb{N}$, sont également disjoints deux-à-deux et que $B \cap A = \bigcup_{n \in \mathbb{N}} B_n \cap A$. La sigma-additivité de \mathbf{P} implique alors que $\mathbf{P}(B \cap A) = \sum_{n \in \mathbb{N}} \mathbf{P}(B_n \cap A)$ en divisant cette égalité par $\mathbf{P}(A)$, on obtient alors $\mathbf{P}(B|A) = \sum_{n \in \mathbb{N}} \mathbf{P}(B_n|A)$. Cela prouve que $\mathbf{P}(\cdot|A)$ est sigma-additive et donc que c'est une mesure de probabilité. ■

Toutes les propriétés pour les probabilités sont valables pour les probabilités conditionnelles. Par exemple, si $(A_n)_{n \geq 1}$ est une suite croissante d'éléments de \mathcal{F} , alors

$$\mathbf{P}\left(\bigcup_{n \geq 1} A_n \mid B\right) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n \mid B).$$

Exemple I.2.6 On joue avec deux dés, un rouge et un blanc. Soit $A = \{ \text{on obtient deux fois } 6 \}$. Alors $\mathbf{P}(A) = 1/36$.

Soit $B = \{ \text{le dé blanc donne } 6 \}$. On a $\mathbf{P}(B) = 1/6$. Donc la probabilité d'avoir deux 6, sachant que le dé blanc a donné 6, est

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{1/36}{1/6} = \frac{1}{6}.$$

On remarque ici que $\mathbf{P}(A|B) \neq \mathbf{P}(A)$. □

Propriété I.2.9 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient A et B deux événements de \mathcal{F} tels que $\mathbf{P}(A)\mathbf{P}(B) > 0$. On a alors

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

Preuve : par définition, $\mathbf{P}(B|A) = \mathbf{P}(A \cap B)/\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B)/\mathbf{P}(A)$. ■

Propriété I.2.10 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A_n \in \mathcal{F}$, $n \geq 1$, une partition dénombrable de Ω , avec $\mathbf{P}(A_n) > 0$ pour tout n . Alors

$$\mathbf{P}(B) = \sum_n \mathbf{P}(A_n)\mathbf{P}(B|A_n), \quad \forall B \in \mathcal{F}.$$

◀▶ **Preuve :** la formule des probabilités totales pour \mathbf{P} (voir la propriété I.2.4, page 14) donne $\mathbf{P}(B) = \sum_n \mathbf{P}(B \cap A_n)$. Le résultat découle alors de la relation $\mathbf{P}(B \cap A_n) = \mathbf{P}(A_n)\mathbf{P}(B|A_n)$. ■

Exemple I.2.7 On tire deux cartes d'un paquet de 52, en retenant l'ordre du tirage : il y a donc une première carte tirée puis une seconde. *Quelle est la probabilité que la deuxième carte soit « Dame » ?*

La situation dépend du résultat de la première carte. Si celle-ci est Dame, alors la probabilité (conditionnelle) que la deuxième carte soit Dame est $3/51$. Si celle-ci ne l'est pas, alors la probabilité (conditionnelle) en question devient $4/51$. On est donc dans le cadre de probabilités conditionnelles, et l'on fait une discussion sur le résultat de la première carte.

Soient

$$\begin{aligned} A_1 &:= \{\text{la première carte est Dame}\}, \\ A_2 &:= \{\text{la première carte n'est pas Dame}\}. \end{aligned}$$

On a $A_1 \cup A_2 = \Omega$ et $A_1 \cap A_2 = \emptyset$. Soit $B := \{\text{la deuxième carte est Dame}\}$. Par la Propriété 5.6,

$$\mathbf{P}(B) = \mathbf{P}(A_1)\mathbf{P}(B|A_1) + \mathbf{P}(A_2)\mathbf{P}(B|A_2) = \frac{4}{52} \times \frac{3}{51} + \frac{48}{52} \times \frac{4}{51} = \frac{1}{13}.$$

On constate que la deuxième carte a exactement la même chance d'être Dame que la première, ce qui est rassurant par exemple pour le tirage au sort dans un tournoi sportif. □

Propriété I.2.11 (Formule de Bayes). Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A_0, \dots, A_n \in \mathcal{F}$ tels que $A_0 \subset \dots \subset A_n$ et tels que $\mathbf{P}(A_1) > 0$, on a

$$\mathbf{P}(A_0|A_n) = \prod_{0 \leq k < n} \mathbf{P}(A_k|A_{k+1}).$$

Preuve : c'est immédiat. ■

Propriété I.2.12 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $A_n \in \mathcal{F}$, $n \geq 1$, une partition dénombrable de Ω avec $\mathbf{P}(A_n) > 0$ pour tout n . Alors pour tout $B \in \mathcal{F}$ tel que $\mathbf{P}(B) > 0$,

$$\mathbf{P}(A_n|B) = \frac{\mathbf{P}(A_n)\mathbf{P}(B|A_n)}{\sum_m \mathbf{P}(A_m)\mathbf{P}(B|A_m)}.$$

Preuve : par définition,

$$\mathbf{P}(A_n|B) = \frac{\mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_n)\mathbf{P}(B|A_n)}{\mathbf{P}(B)},$$

ce qui complète la preuve à l'aide de la propriété I.2.10, page 19. ■

Exemple I.2.8 Dans un concours de pronostic sportif pour un match entre l'équipe E et l'équipe F (pas de match nul), les participants sont composés à 50% d'étudiants de Jussieu, à 20% d'étudiants de Dauphine, et à 30% d'étudiants d'Orsay. On constate que 60% des étudiants de Jussieu prévoient l'équipe E gagnante, ainsi que 30% des étudiants de Dauphine, et 90% des étudiants d'Orsay. On tire une personne au hasard parmi les participants : elle pronostique sur l'équipe F gagnante. *Quelle est la probabilité qu'il s'agit de quelqu'un de Jussieu ?*

On définit

$$A_1 := \{\text{Jussieu}\}, A_2 := \{\text{Dauphine}\} \text{ et } A_3 := \{\text{Orsay}\}.$$

On sait que ces trois événements forment une partition de Ω , avec $\mathbf{P}(A_1) = 0,5$, $\mathbf{P}(A_2) = 0,2$, et $\mathbf{P}(A_3) = 0,3$. Soit

$$B := \{\text{supporteur de l'équipe F}\}.$$

Par hypothèse, $\mathbf{P}(B|A_1) = 0,4$, $\mathbf{P}(B|A_2) = 0,7$, et $\mathbf{P}(B|A_3) = 0,1$. D'après la propriété I.2.12,

$$\begin{aligned} \mathbf{P}(A_1|B) &= \frac{\mathbf{P}(A_1)\mathbf{P}(B|A_1)}{\mathbf{P}(A_1)\mathbf{P}(B|A_1) + \mathbf{P}(A_2)\mathbf{P}(B|A_2) + \mathbf{P}(A_3)\mathbf{P}(B|A_3)} \\ &= \frac{0,5 \times 0,4}{0,5 \times 0,4 + 0,2 \times 0,7 + 0,3 \times 0,1} = \frac{20}{37} > \frac{1}{2}. \end{aligned}$$

□

I.2.d Premières notions sur l'indépendance.

Informellement, un événement B est indépendant de A si la réalisation de A n'influe pas sur la probabilité de la réalisation ou la non-réalisation de B , c'est-à-dire que l'on doit avoir $\mathbf{P}(A|B) = \mathbf{P}(A)$, autrement dit $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On remarque que cette notion est symétrique en A et B . Ces considérations élémentaires conduisent à formaliser la notion d'indépendance d'événements comme suit.

Définition I.2.2 ◀▶ (Indépendance de 2 événements) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A, B \in \mathcal{F}$. Les événements A et B sont dits *indépendants sous \mathbf{P}* si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

□

Propriété I.2.13 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A, B \in \mathcal{F}$. Les assertions suivantes sont équivalentes.

- (a) A et B sont indépendants sous \mathbf{P} .
- (b) $\Omega \setminus A$ et B sont indépendants sous \mathbf{P} .
- (c) A et $\Omega \setminus B$ sont indépendants sous \mathbf{P} .
- (d) $\Omega \setminus A$ et $\Omega \setminus B$ sont indépendants sous \mathbf{P} .

◀► **Preuve :** on suppose (a) donc $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On a donc

$$\mathbf{P}(B \setminus (A \cap B)) = \mathbf{P}(B) - \mathbf{P}(A \cap B) = \mathbf{P}(B) - \mathbf{P}(A)\mathbf{P}(B) = (1 - \mathbf{P}(A))\mathbf{P}(B) = \mathbf{P}(\Omega \setminus A)\mathbf{P}(B).$$

On observe ensuite que $B \setminus (A \cap B) = B \cap (\Omega \setminus A)$. Donc finalement,

$$\mathbf{P}(B \cap (\Omega \setminus A)) = \mathbf{P}(\Omega \setminus A)\mathbf{P}(B).$$

On a donc montré que (a) implique (b). En raisonnant de même en inversant les rôles de A et de B on montre que (a) implique (c). En appliquant l'implication (a) \Rightarrow (b) à A et $\Omega \setminus B$, on montre que (a) implique (d) et les réciproques se montrent par passage au complémentaire. ■

Propriété I.2.14 ▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. On vérifie les deux propriétés suivantes.

- (i) Soit $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 0$ (presque sûrement A ne se produit pas). Alors pour tout $B \in \mathcal{F}$, A et B sont indépendants.
- (ii) Soit $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 1$ (l'événement A se produit presque sûrement). Alors pour tout $B \in \mathcal{F}$, A et B sont indépendants.

◀► **Preuve :** soit $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 0$. Soit $B \in \mathcal{F}$. Alors, $A \cap B \subset A$ et donc $0 \leq \mathbf{P}(A \cap B) \leq \mathbf{P}(A) = 0$ par croissance des mesures de probabilité pour l'inclusion. Donc $\mathbf{P}(A \cap B) = 0 = \mathbf{P}(A)\mathbf{P}(B)$, ce qui montre que A et B sont indépendants. Cela montre (i).

Montrons (ii) : soit $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 1$. Alors $\mathbf{P}(\Omega \setminus A) = 0$. Le (i) implique que pour tout $B \in \mathcal{F}$, $\Omega \setminus A$ et B sont indépendants sous \mathbf{P} . La propriété I.2.13 (a) \Leftrightarrow (b) implique alors que A et B sont indépendants sous \mathbf{P} . ■

Exemple I.2.9 On jette un dé rouge et un dé blanc, tous les deux sont cubiques. Construisons rapidement un espace de probabilité modélisant cette expérience aléatoire : on note par exemple les deux résultats dans un vecteur du plan dont la première composante est le résultat du dé blanc et la seconde composante est le résultat du dé rouge. On a donc

$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}, \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \text{et} \quad \mathbf{P} = \frac{1}{36} \sum_{\omega \in \Omega} \delta_{\omega}.$$

Ici, \mathbf{P} est l'équi-probabilité sur Ω . On considère les deux événements suivants :

$$A = \{\text{le résultat du dé blanc est pair}\} \quad \text{et} \quad B = \{\text{le résultat du dé rouge est divisible par 3}\}.$$

D'un point de vue intuitif ces deux événements sont indépendants : le premier concerne uniquement le dé blanc et le second uniquement le dé rouge. Or le résultat d'un dé n'influence pas l'autre. Démontrons cela.

$$A = \{2, 4, 6\} \times \{1, \dots, 6\}, \quad B = \{1, \dots, 6\} \times \{3, 6\} \quad \text{et} \quad A \cap B = \{2, 4, 6\} \times \{3, 6\}.$$

On a donc $\mathbf{P}(A \cap B) = 6/36 = 1/6$, $\mathbf{P}(A) = 18/36 = 1/2$ et $\mathbf{P}(B) = 12/36 = 1/3$. Donc $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, ce qui montre que A et B sont indépendants. □

On introduit ensuite la définition de l'indépendance mutuelle d'un nombre fini d'événement.

Définition I.2.3 ◀▶ (Indépendance mutuelle d'une suite finie d'événements) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $A_1, \dots, A_n \in \mathcal{F}$. Les événements A_1, \dots, A_n sont dits *mutuellement indépendants sous \mathbf{P}* si pour tout $2 \leq k \leq n$ et tous $1 \leq j_1 < \dots < j_k \leq n$, on

$$\mathbf{P}(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \mathbf{P}(A_{j_1})\mathbf{P}(A_{j_2}) \dots \mathbf{P}(A_{j_k}) .$$

Attention ! Il y a donc $2^n - 1 - n$ égalités à vérifier. □

Remarque I.2.1 L'indépendance mutuelle implique l'indépendance deux-à-deux mais la réciproque est fautive en général : on peut trouver un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et trois événements $A, B, C \in \mathcal{F}$ tels que, sous \mathbf{P} , A et B soient indépendants, B et C soient indépendants, A et C soient indépendants mais A, B, C ne soient pas mutuellement indépendants.

En effet on peut prendre $\Omega = \{1, \dots, 4\}$, $\mathcal{F} = \mathcal{P}(\{1, \dots, 4\})$ et $\mathbf{P} = \frac{1}{4} \sum_{1 \leq j \leq 4} \delta_j$, la loi uniforme sur Ω . On prend alors

$$A = \{1, 2\}, \quad B = \{1, 3\} \quad \text{et} \quad C = \{1, 4\} .$$

On voit alors que $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/2$, que $A \cap B = A \cap C = B \cap C = \{1\}$. Donc

$$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A)\mathbf{P}(C) = \mathbf{P}(B)\mathbf{P}(C) ,$$

ce qui montre que A, B, C sont indépendants deux-à-deux sous \mathbf{P} . En revanche $A \cap B \cap C = \{1\}$ et donc

$$\mathbf{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C),$$

ce qui montre que A, B, C ne sont pas mutuellement indépendants sous \mathbf{P} . □

L'analogie de la propriété I.2.13 pour l'indépendance mutuelle de n événements. Elle peut se prouver directement mais nous l'expliquons dans un chapitre ultérieur approfondissant la notion d'indépendance.

Proposition I.2.3 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $n \geq 2$, un entier. On considère n événements $A_1, \dots, A_n \in \mathcal{F}$. Soit $A'_1, \dots, A'_n \in \mathcal{F}$ tels que pour tout $k \in \{1, \dots, n\}$,

$$A'_k = A_k \text{ ou bien } \Omega \setminus A_k .$$

Alors, les deux assertions suivantes sont équivalentes.

- (a) Les événements A_1, \dots, A_n sont mutuellement indépendants sous \mathbf{P} .
- (b) Les événements A'_1, \dots, A'_n sont mutuellement indépendants sous \mathbf{P} .

Enfin, il sera utile d'introduire la notion d'indépendance mutuelle d'une famille (possiblement infinie) d'événements.

Définition I.2.4 ◀▶ (Indépendance mutuelle d'une suite finie d'événements) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $A_i \in \mathcal{F}$, $i \in I$, une famille d'événements. On dit qu'ils sont *mutuellement indépendants sous \mathbf{P}* si pour tout sous-ensemble d'indices $J \subset I$ qui est fini et non-vidé, on a

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j) .$$

On remarque que les $(A_i)_{i \in I}$ sont mutuellement indépendants sous \mathbf{P} si et seulement si pour tout sous-ensemble d'indices $J \subset I$ qui est fini et qui compte au moins deux éléments, les événements $(A_j)_{j \in J}$ sont mutuellement indépendants sous \mathbf{P} comme défini plus haut. \square

◀► **Convention** Pour simplifier les énoncés il est *sous-entendu* que « l'indépendance » signifie la « mutuelle indépendance ». On dit donc simplement « les événements A_1, \dots, A_n sont indépendants sous \mathbf{P} » au lieu de dire « les événements A_1, \dots, A_n sont mutuellement indépendants sous \mathbf{P} ». Enfin, lorsqu'il n'y a pas d'ambiguïté sur la mesure de probabilité que l'on considère, on omet le terme « sous \mathbf{P} ». Donc finalement on dit simplement « les événements A_1, \dots, A_n sont indépendants » au lieu de dire « les événements A_1, \dots, A_n sont mutuellement indépendants sous \mathbf{P} ». \square

Exemple I.2.10 (Une application arithmétique concernant la fonction d'Euler). Désignons par $\varphi(n)$ la fonction d'Euler de la théorie des nombres, c'est-à-dire, $\varphi(n)$ est le nombre des entiers plus petits que n sans diviseur commun avec n . Alors

$$\varphi(n) = n \prod_{\substack{p \text{ premier,} \\ p \text{ divise } n}} \left(1 - \frac{1}{p}\right), \quad (\text{I.15})$$

Pour redémontrer cette formule bien connue de la théorie des nombres, on considère le modèle probabiliste suivant : on choisit au hasard un nombre parmi $\{1, 2, \dots, n\}$ avec équi-probabilité. On peut donc choisir

$$\Omega = \{1, 2, \dots, n\}, \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \text{et} \quad \mathbf{P} = \frac{1}{n} \sum_{1 \leq k \leq n} \delta_k.$$

Pour tout nombre premier p , soit l'événement

$$A_p = \{ \text{le nombre choisi est divisible par } p \}.$$

Soient p_1, p_2, \dots, p_m les facteurs premiers de n . Montrons d'abord que $A_{p_1}, A_{p_2}, \dots, A_{p_m}$ sont des événements indépendants. D'après la II il suffit de montrer que

$$\mathbf{P}(A_{p_{i_1}} \cap \dots \cap A_{p_{i_k}}) = \mathbf{P}(A_{p_{i_1}}) \cdots \mathbf{P}(A_{p_{i_k}})$$

pour tous $1 \leq i_1 < \dots < i_k \leq m$. Or, il est clair que

$$\begin{aligned} \mathbf{P}(A_{p_i}) &= \mathbf{P}\left(\text{le nombre est un élément de } \{p_i, 2p_i, 3p_i, \dots, \frac{n}{p_i}p_i\}\right) \\ &= \frac{n/p_i}{n} = \frac{1}{p_i}, \end{aligned}$$

On pose $q := p_{i_1} \cdots p_{i_k}$. Comme les p_{i_k} sont premiers, un entier N est divisible par p_{i_1} , par p_{i_2}, \dots et par p_{i_k} si et seulement si N est divisible par q . Donc

$$\begin{aligned} \mathbf{P}(A_{p_{i_1}} \cap \dots \cap A_{p_{i_k}}) &= \mathbf{P}\left(\text{le nombre choisi est un élément de } \{q, 2q, 3q, \dots, \frac{n}{q}q\}\right) \\ &= \frac{n/q}{n} = \frac{1}{p_{i_1} \cdots p_{i_k}} \\ &= \mathbf{P}(A_{p_{i_1}}) \cdots \mathbf{P}(A_{p_{i_k}}), \end{aligned}$$

Ce qui prouve que $A_{p_1}, A_{p_2}, \dots, A_{p_m}$ sont indépendants. Par la proposition I.2.3, page 22, les complémentaires $A_{p_1}^c, A_{p_2}^c, \dots, A_{p_m}^c$ sont aussi indépendants. On a donc

$$\mathbf{P}(A_{p_1}^c \cap \dots \cap A_{p_k}^c) = \mathbf{P}(A_{p_1}^c) \dots \mathbf{P}(A_{p_k}^c) = \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right).$$

D'autre part $A_{p_1}^c \cap \dots \cap A_{p_k}^c$ est l'ensemble des entiers de $\{1, \dots, n\}$ qui ne sont pas divisibles par p_1 , ni par p_2, \dots et ni par p_n : ils n'ont donc aucun diviseur commun avec n . Autrement dit $A_{p_1}^c \cap \dots \cap A_{p_k}^c$ est l'ensemble des entiers de $\{1, \dots, n\}$ sans diviseur commun avec n et donc $\#(A_{p_1}^c \cap \dots \cap A_{p_k}^c) = \varphi(n)$. On en déduit que $\mathbf{P}(A_{p_1}^c \cap \dots \cap A_{p_k}^c) = \varphi(n)/n$, ce qui entraîne l'identité (I.15) \square

I.3 Exercices.

Exercice I.3.1 Quelle est la probabilité pour que parmi vingt-trois personnes, deux aient la même date d'anniversaire ?

Exercice I.3.2 On tire deux cartes d'un jeu de 32. Quelle est la probabilité d'obtenir une paire ? Si l'on n'a pas obtenu une paire, on a le choix entre jeter l'une des deux cartes tirées et en retirer une parmi les 30 restantes, ou jeter les deux cartes tirées et en retirer deux parmi les 30 restantes. Quelle stratégie donne la plus grande probabilité d'avoir une paire à la fin ?

Exercice I.3.3 a. Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités. Soit $n \geq 1$. Soient A_1, \dots, A_n des événements. Montrer que

$$\mathbf{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

C'est la formule d'*inclusion-exclusion*.

b. En appliquant cette formule à un espace de probabilités et à des événements bien choisis, calculer le nombre de surjections de $\{1, \dots, p\}$ dans $\{1, \dots, n\}$ pour tous n et p entiers.

Exercice I.3.4 Que peut-on dire d'un événement qui est indépendant de lui-même ?

Exercice I.3.5 Soit (Ω, \mathcal{F}) un espace mesurable. Soit $\omega \in \Omega$. Montrer que la formule

$$\delta_\omega(A) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

définit une mesure de probabilités sur (Ω, \mathcal{F}) . On l'appelle la *mesure de Dirac* en ω .

Donner un exemple d'espace de probabilités et d'événement négligeable non vide ainsi que d'événement presque sûr différent de l'espace tout entier.

Exercice I.3.6 Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités. On rappelle la propriété de sigma-sous-additivité qui affirme que pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{F} , on a

$$\mathbf{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 0} \mathbf{P}(A_n).$$

Montrer que $\mathcal{N} = \{A \in \mathcal{F} : \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}$ est une tribu sur Ω .

Exercice I.3.7 Sur la page de Wikipedia consacrée aux probabilités au poker¹ on trouve un tableau qui indique la probabilité que dans cinq cartes tirées au hasard parmi trente-deux, on trouve une configuration plus forte (selon les règles du poker) que telle ou telle figure, en particulier qu'une paire de rois. Après ce tableau on lit la chose suivante :

Ce tableau est indépendant du nombre de joueurs, mais n'est pas exploité directement ainsi. L'utilisation typique de ce tableau est de répondre à des questions comme : J'ai une paire de roi servie, nous jouons à quatre à 32 cartes, quelle est la probabilité a priori pour que ma main soit la meilleure ? Pour ce type de question, les étapes de calcul sont :

La probabilité pour un joueur d'avoir plus qu'une paire de roi dans ces conditions est : 36,8%. Il aura moins avec une probabilité de 63,2%. Pour que la paire de roi soit la plus forte, il faut que le premier adversaire ait moins ET le second ait moins ET le troisième ait moins. La probabilité est le produit des trois : $63,2\% \times 63,2\% \times 63,2\% = 25,2\%$. On peut donc parier à un contre trois que ma paire de rois n'est pas la meilleure main des quatre.

Que pensez-vous de cet extrait ?

Exercice I.3.8 On rappelle qu'un ensemble est *dénombrable* s'il peut être mis en bijection avec une partie finie ou infinie de l'ensemble \mathbb{N} des entiers naturels.

a. Montrer qu'un ensemble E est dénombrable si et seulement s'il existe une injection de E dans \mathbb{N} .

b. Montrer qu'un ensemble non vide E est dénombrable si et seulement s'il existe une surjection de \mathbb{N} sur E .

c. Montrer que si E et F sont deux ensembles dénombrables, alors $E \times F$ est dénombrable. Plus généralement, montrer que si E_1, \dots, E_n sont dénombrables, alors $E_1 \times \dots \times E_n$ est dénombrable.

d. Montrer que si I est dénombrable et $(E_i)_{i \in I}$ est une famille d'ensembles dénombrables, alors $\bigcup_{i \in I} E_i$ est un ensemble dénombrable.

On retiendra qu'un produit cartésien fini d'ensembles dénombrables et une réunion dénombrable d'ensembles dénombrables sont dénombrables.

e. Parmi les ensemble suivants, dire lesquels sont dénombrables : \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} , l'ensemble des suites finies de longueur quelconque de 0 et de 1, l'ensemble des suites infinies de 0 et de 1, l'ensemble des suites finies d'entiers naturels, l'ensemble des polynômes à une indéterminée à coefficients rationnels.

Exercice I.3.9 Soit Ω un ensemble non vide.

a. Soit $(\mathcal{F}_i)_{i \in I}$ une famille quelconque de tribus sur Ω . On pose

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i.$$

Montrer que \mathcal{F} est une tribu sur Ω .

b. Montrer que pour toute partie $\mathcal{G} \subset \mathcal{P}(\Omega)$, il existe une tribu qui contient \mathcal{G} et qui est minimale pour l'inclusion parmi toutes les tribus qui ont cette propriété. On appelle cette tribu la *tribu engendrée par \mathcal{G}* et on la note $\sigma(\mathcal{G})$.

c. Déterminer la tribu engendré par une partition finie ou dénombrable de Ω , c'est-à-dire par une collection $\mathcal{G} = \{A_1, A_2, \dots\}$ de parties deux à deux disjointes et dont la réunion vaut Ω .

1. http://fr.wikipedia.org/wiki/Probabilité_au_poker#Tableau_de_synthèse

Exercice I.3.10 Soit (Ω, \mathcal{F}) un espace mesurable.

a. Montrer qu'on définit une relation d'équivalence sur Ω en posant

$$(x \sim y) \Leftrightarrow (\forall A \in \mathcal{F}, x \in A \Rightarrow y \in A).$$

Les classes de cette relation s'appellent les *atomes* de la tribu \mathcal{F} .

b. On suppose désormais que Ω est fini ou dénombrable. Montrer que \mathcal{F} coïncide avec la tribu engendrée par la partition de Ω en les atomes de \mathcal{F} .

Exercice I.3.11 Soit Ω un ensemble.

a. On note \mathcal{F} l'ensemble des parties de Ω qui sont dénombrables ou dont le complémentaire est dénombrable. Montrer que \mathcal{F} est une tribu sur Ω .

b. Plus généralement, soit $\Pi = (\Pi_i)_{i \in I}$ une partition de Ω . On ne fait aucune hypothèse sur le cardinal de I . On note \mathcal{G} l'ensemble des parties de Ω qui sont réunion d'un nombre fini ou dénombrable de blocs de Π , ou dont le complémentaire est réunion d'un nombre fini ou dénombrable de blocs de Π . Montrer que \mathcal{G} est une tribu sur Ω . Montrer que c'est la tribu engendrée par la partition Π .

c. Montrer que si une tribu sur Ω est engendrée par une partition, alors les blocs de cette partition sont ses atomes.

d. La tribu borélienne sur \mathbb{R} (c'est-à-dire la tribu engendrée par la topologie usuelle de \mathbb{R}) est-elle engendrée par une partition ?

Chapitre II

Variables aléatoires, espérance, lois.

II.1 Variables aléatoires.

II.1.a Rappels sur les fonctions mesurables.

Soient E et E' , deux ensembles et une fonction $f: E \rightarrow E'$. Pour tout $B \subset E'$, on utilise la notion

$$f^{-1}(B) := \{x \in E : f(x) \in B\} \subset E.$$

L'ensemble $f^{-1}(E)$ est la *pré-image de E par f* . On utilise également la notation

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

Rappelons le résultat ensembliste élémentaire suivant : soit $B_j \subset E'$, $j \in J$, une famille (quelconque) de sous-ensembles de E' . Alors,

$$f^{-1}\left(\bigcup_{j \in J} B_j\right) = \bigcup_{j \in J} f^{-1}(B_j) \quad \text{et} \quad f^{-1}\left(\bigcap_{j \in J} B_j\right) = \bigcap_{j \in J} f^{-1}(B_j). \quad (\text{II.1})$$

Autrement dit, la pré-image d'une union est l'union des pré-images et l'intersection d'une pré-image est l'intersection des pré-images. Pour illustrer cette propriété, on démontre le lemme suivant qui montre que la propriété d'être une tribu pour une classe d'ensembles est préservé par l'opération « *pré-image par une fonction* »

Lemme II.1.1 ◀▶ (Stabilité des tribus par pré-image) Soient E et E' des ensembles non-vides. Soit une fonction $f: E \rightarrow E'$. Soit \mathcal{E}' , une tribu sur E' . Alors, la pré-image de cette tribu par f , c'est-à-dire la classe d'ensembles sur E suivante :

$$f^{-1}(\mathcal{E}') := \{f^{-1}(B); B \in \mathcal{E}'\}$$

est une tribu sur E .

◀▶ **Preuve** : pour simplifier on pose $\mathcal{E} = f^{-1}(\mathcal{E}')$. Comme $f^{-1}(E') = E$, on a $E \in \mathcal{E}$. Soit $B \in \mathcal{E}'$; comme $f^{-1}(E' \setminus B) = E \setminus f^{-1}(B)$, on voit que \mathcal{E} est stable par passage au complémentaire. Soit $B_n \in \mathcal{E}'$, $n \in \mathbb{N}$; on remarque $f^{-1}(\bigcup_{n \in \mathbb{N}} B_n) = \bigcup_{n \in \mathbb{N}} f^{-1}(B_n)$, ce qui montre que \mathcal{E} est stable par union dénombrable. Donc \mathcal{E} est une tribu. ■

Définition II.1.1 ◀▶ (*Fonctions mesurables*) Soient (E, \mathcal{E}) et (E', \mathcal{E}') , deux espaces mesurables. Une fonction $f : E \rightarrow E'$ est dite $(\mathcal{E}, \mathcal{E}')$ -mesurable si

$$\forall B \in \mathcal{E}' , \quad f^{-1}(B) \in \mathcal{E} . \quad (\text{II.2})$$

◀▶ **Convention importante.** Lorsque E' est $\mathbb{R}, [0, \infty], [-\infty, \infty], \mathbb{C}, \mathbb{R}^n$ etc, on dit simple que « f est \mathcal{E} -mesurable » pour dire que « f est $(\mathcal{E}, \mathcal{E}')$ -mesurable » où il est sous-entendu que \mathcal{E}' est la tribu des Boréliens de E' . Ainsi lorsque l'on écrit que $f : E \rightarrow \mathbb{R}$ est \mathcal{E} -mesurable cela signifie en fait que f est $(\mathcal{E}, \mathcal{B}(\mathbb{R}))$ -mesurable. ◻

Pour montrer qu'une fonction est mesurable, il semble a priori nécessaire d'effectuer autant de vérifications qu'il y a d'ensembles dans la tribu sur l'espace d'arrivée. Ce n'est heureusement pas le cas en général : la proposition suivante, d'un usage constant, permet de réduire le nombre de vérifications en se restreignant à une classe plus petite ou plus commode.

Proposition II.1.2 ◀▶ (*Simplification de la mesurabilité*) Soient (E, \mathcal{E}) et (E', \mathcal{E}') , deux espaces mesurables. Soit $\mathcal{C} \subset \mathcal{E}'$, une classe quelconque d'ensembles engendrant \mathcal{E}' , c'est-à-dire que

$$\sigma(\mathcal{C}) = \mathcal{E}' .$$

Soit $f : E \rightarrow E'$. Alors, f est $(\mathcal{E}, \mathcal{E}')$ -mesurable si et seulement si

$$\forall C \in \mathcal{C} , \quad f^{-1}(C) \in \mathcal{E} . \quad (\text{II.3})$$

◀▶ **Preuve :** on suppose que f satisfait (II.3) et on pose $\mathcal{G} = \{B \in \mathcal{E}' : f^{-1}(B) \in \mathcal{E}\}$. Par (II.3), on a $\mathcal{C} \subset \mathcal{G}$. Il suffit de montrer que \mathcal{G} est une tribu car alors $\mathcal{E}' = \sigma(\mathcal{C}) \subset \mathcal{G} \subset \mathcal{E}'$, et donc $\mathcal{G} = \mathcal{E}'$ qui est le résultat voulu.

Montrons donc que \mathcal{G} est une tribu : comme $f^{-1}(E') = E$, on a $E' \in \mathcal{G}$. Soit $B \in \mathcal{G}$; comme $f^{-1}(E' \setminus B) = E \setminus f^{-1}(B)$, on a $E' \setminus B \in \mathcal{G}$. Soit $B_n \in \mathcal{G}, n \in \mathbb{N}$; on a

$$f^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \bigcup_{n \in \mathbb{N}} f^{-1}(B_n) ,$$

ce qui entraîne que $\bigcup B_n \in \mathcal{G}$, achevant la preuve de la proposition. ■

Soit $f : E \rightarrow \mathbb{R}$. Pour tout $a \in \mathbb{R}$, on note l'ensemble $f^{-1}(]-\infty, a])$ (qui est par définition l'ensemble $\{x \in E : f(x) \leq a\}$) simplement par $\{f \leq a\}$. En application de la proposition précédente, on obtient la propriété suivante.

Propriété II.1.1 ◀▶ Soit (E, \mathcal{E}) , un espace mesurable et soit $f : E \rightarrow \mathbb{R}$. Alors,

$$f \text{ est } \mathcal{E}\text{-mesurable} \iff \forall a \in \mathbb{R} , \quad \{f \leq a\} \in \mathcal{E} .$$

◀▶ **Preuve :** on pose $\mathcal{C} = \{]-\infty, a] ; a \in \mathbb{R}\}$. Par le lemme I.1.2 page 4, on sait que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. La proposition II.1.2 qui précède implique alors le résultat voulu. ■

Propriété II.1.2 Soit $f : \mathbb{R} \rightarrow \mathbb{R}$, une fonction monotone. Alors elle est $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ -mesurable.

$$f \text{ est } \mathcal{E}\text{-mesurable} \iff \forall a \in \mathbb{R} , \quad \{f \leq a\} \in \mathcal{E} .$$

Preuve : un intervalle ouvert de \mathbb{R} est un borélien de \mathbb{R} car $\mathcal{B}(\mathbb{R})$ est engendrée par les ouverts : elle les contient donc. Un fermé est le complémentaire d'un ouvert : il est également dans la tribu borélienne. On en déduit que les singletons sont dans la tribu borélienne de \mathbb{R} . Enfin tout intervalle est la réunion d'un intervalle ouvert et d'au plus deux singletons : on en déduit que tout intervalle est dans $\mathcal{B}(\mathbb{R})$. On observe ensuite que pour tout $a \in \mathbb{R}$, $\{f \leq a\}$ est un intervalle de \mathbb{R} car f est supposée monotone. Donc pour tout $a \in \mathbb{R}$, $\{f \leq a\} \in \mathcal{B}(\mathbb{R})$ ce qui montre que le résultat voulu par la propriété II.1.1 qui précède. ■

Donnons plusieurs applications importantes de ce résultat.

Exemple II.1.1 ◀▶ Soit (E, \mathcal{E}) , un espace mesurable et soit $f : E \rightarrow \mathbb{R}$. Comme $\sigma(\{]-\infty, a] ; a \in \mathbb{R} \}) = \mathcal{B}(\mathbb{R})$, la proposition II.1.2 montre donc que

$$f \text{ est } \mathcal{E}\text{-mesurable} \iff \forall a \in \mathbb{R}, f^{-1}(] -\infty, a]) \in \mathcal{E}.$$

On rappelle que $f^{-1}(] -\infty, a]) = \{x \in E : f(x) \leq a\}$. Cet ensemble est souvent noté $\{f \leq a\}$.

Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est monotone, alors $\{f \leq a\}$ est un intervalle de \mathbb{R} , qui est donc un Borélien de \mathbb{R} . Cela montre que toute fonction monotone est $\mathcal{B}(\mathbb{R})$ -mesurable. □

Rappelons ensuite la définition générale des fonctions continues.

Définition II.1.2 ◀▶ (*Continuité en général*) Soient (E, \mathcal{T}) et (E', \mathcal{T}') , deux espaces topologiques. Une fonction $f : E \rightarrow E'$ est $(\mathcal{T}, \mathcal{T}')$ -continue si $f^{-1}(U') \in \mathcal{T}$, pour tout $U' \in \mathcal{T}'$. □

La proposition II.1.2 implique immédiatement le lemme suivant.

Lemme II.1.3 ◀▶ (*Les fonctions continues sont mesurables*) Soient (E, \mathcal{T}) et (E', \mathcal{T}') , deux espaces topologiques et $f : E \rightarrow E'$, une fonction $(\mathcal{T}, \mathcal{T}')$ -continue. Alors, elle est $(\mathcal{B}(E), \mathcal{B}(E'))$ -mesurable.

La stabilité de la mesurabilité par composition est une conséquence immédiate de la définition des fonctions mesurables.

Proposition II.1.4 ◀▶ (*Composition de fonctions mesurables*) Soient (E, \mathcal{E}) , (E', \mathcal{E}') , (E'', \mathcal{E}'') , trois espaces mesurables, et soient $f : E \rightarrow E'$ et $g : E' \rightarrow E''$, deux fonctions respectivement $(\mathcal{E}, \mathcal{E}')$ et $(\mathcal{E}', \mathcal{E}'')$ -mesurables. Alors, $g \circ f$ est $(\mathcal{E}, \mathcal{E}'')$ -mesurable.

Proposition II.1.5 ◀▶ (*Addition et produit de fonctions mesurables*) Soit (E, \mathcal{E}) un espace mesurable et $f, g : E \rightarrow \mathbb{R}$, deux fonctions \mathcal{E} -mesurables. Alors, $f + g$ et fg sont également \mathcal{E} -mesurables. De même, si $f, g : E \rightarrow [0, \infty]$ sont \mathcal{E} -mesurables, alors $f + g$ également.

Preuve : on montre d'abord que $h := (f, g) : E \rightarrow \mathbb{R}^2$, est $(\mathcal{E}, \mathcal{B}(\mathbb{R}^2))$ -mesurable : soient $a, b, c, d \in \mathbb{Q}$ tels que $a \leq b$ et $c \leq d$. On pose $R =]a, b[\times]c, d[$. On remarque que

$$h^{-1}(R) = \{x \in E : a < f(x) < b \text{ et } c < g(x) < d\} = f^{-1}(]a, b[) \cap g^{-1}(]c, d[) \in \mathcal{E}.$$

Le lemme I.1.3 page 4 et la proposition II.3 impliquent alors que h est $(\mathcal{E}, \mathcal{B}(\mathbb{R}^2))$ -mesurable.

On remarque ensuite que l'addition **Add** ou la multiplication **Mul** sont deux applications continues de \mathbb{R}^2 dans \mathbb{R} . Par conséquent, elles sont toutes les deux $(\mathcal{B}(\mathbb{R}^2), \mathcal{B}(\mathbb{R}))$ -mesurables, d'après le lemme II.1.3. Il suffit ensuite d'observer que $\mathbf{Add} \circ h = f + g$ et que $\mathbf{Mul} \circ h = fg$, et la proposition II.1.4 entraîne le résultat voulu. La mesurabilité de la somme de deux fonctions à valeurs dans $[0, \infty]$, se démontre de la même façon. ■

Brefs rappels sur la droite numérique achevée. On adjoint ∞ et $-\infty$ à l'ensemble des réels. L'ensemble résultant est noté $[-\infty, \infty]$: il est souvent appelé la *droite numérique achevée*. L'ordre usuel s'étend naturellement à $[-\infty, \infty]$ par $-\infty \leq x \leq \infty$, pour tout $x \in [-\infty, \infty]$. Les notions de convergence vers ∞ et $-\infty$ sont les définitions habituelles.

Soient $a, b \in [-\infty, \infty]$. On utilise souvent les notations $a \wedge b = \min(a, b)$ et $a \vee b = \max(a, b)$. La *partie positive* de a est donnée par $(a)_+ = \max(0, a)$ et la *partie négative* de a est donnée par $(a)_- = \max(0, -a)$ et on rappelle que $|a|$ désigne la valeur absolue de a , avec des conventions évidentes lorsque a est infini. On rappelle que

$$0 \leq (a)_+ \leq |a|, \quad 0 \leq (a)_- \leq |a|, \quad |a| = (a)_+ + (a)_- \quad \text{et} \quad a = (a)_+ - (a)_-,$$

la dernière égalité ayant bien un sens aussi lorsque a est infini.

Soit $a_n \in [-\infty, \infty]$, $n \in \mathbb{N}$. Pour tout $n \in \mathbb{N}$, on pose $r_n = \inf_{p \geq n} a_p$ et $s_n = \sup_{p \geq n} a_p$, qui sont bien définies dans $[-\infty, \infty]$. On remarque que $r_n \leq r_{n+1}$ et $s_n \geq s_{n+1}$. Ces deux suites ont donc des limites dans $[-\infty, \infty]$: la limite de $(r_n)_{n \in \mathbb{N}}$ est la *limite inférieure* de $(a_n)_{n \in \mathbb{N}}$, notée $\liminf_{n \rightarrow \infty} a_n$, et la limite de $(s_n)_{n \in \mathbb{N}}$ est la *limite supérieure* de $(a_n)_{n \in \mathbb{N}}$, notée $\limsup_{n \rightarrow \infty} a_n$. En résumé

$$\liminf_{n \rightarrow \infty} a_n = \sup_{n \in \mathbb{N}} \inf_{p \geq n} a_p = \lim_{n \rightarrow \infty} \uparrow \inf_{p \geq n} a_p \quad \text{et} \quad \limsup_{n \rightarrow \infty} a_n = \inf_{n \in \mathbb{N}} \sup_{p \geq n} a_p = \lim_{n \rightarrow \infty} \downarrow \sup_{p \geq n} a_p.$$

Par définition, on a $\liminf_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} a_n$. Il est facile de voir que $\liminf_{n \rightarrow \infty} a_n$ est la plus petite valeur d'adhérence de $(a_n)_{n \in \mathbb{N}}$ et que $\limsup_{n \rightarrow \infty} a_n$ est la plus grande valeur d'adhérence de $(a_n)_{n \in \mathbb{N}}$. Donc

$$\lim_{n \rightarrow \infty} a_n \text{ existe et vaut } a \in [-\infty, \infty] \quad \iff \quad \liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = a.$$

On vérifie facilement que

$$\liminf_{n \rightarrow \infty} (-a_n) = - \limsup_{n \rightarrow \infty} a_n$$

et de manière analogue $\limsup_{n \rightarrow \infty} (-a_n) = - \liminf_{n \rightarrow \infty} a_n$. Soit $b_n \in [-\infty, \infty]$, $n \in \mathbb{N}$. Les propriétés élémentaires suivantes sont utilisées dans la suite du cours.

- Si $a_n \leq b_n$, pour tout n assez grand, alors

$$\liminf_{n \rightarrow \infty} a_n \leq \liminf_{n \rightarrow \infty} b_n \quad \text{et} \quad \limsup_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} b_n.$$

- On suppose que $\lim_{n \rightarrow \infty} b_n = b \in \mathbb{R}$, alors

$$\limsup_{n \rightarrow \infty} (a_n + b_n) = b + \limsup_{n \rightarrow \infty} a_n \quad \text{et} \quad \liminf_{n \rightarrow \infty} (a_n + b_n) = b + \liminf_{n \rightarrow \infty} a_n.$$

Soit E , un ensemble E et $f, g : E \rightarrow [-\infty, \infty]$. On définit $f \wedge g : E \rightarrow [-\infty, \infty]$ comme la fonction donnée pour tout $x \in E$ par $(f \wedge g)(x) = f(x) \wedge g(x)$. On définit de même $f \vee g$.

La partie positive de f , notée f_+ , est alors donnée par $f_+ = \max(0, f)$ et sa partie négative par $f_- = \max(0, -f)$. La notation $|f|$ est également évidente. Si $f_n, n \in \mathbb{N}$, est une suite de fonctions de E dans $[-\infty, \infty]$, on définit également $\inf_{n \in \mathbb{N}} f_n, \sup_{n \in \mathbb{N}} f_n, \liminf_{n \rightarrow \infty} f_n$ et $\limsup_{n \rightarrow \infty} f_n$. On rappelle que dire que la suite de fonctions $(f_n)_{n \in \mathbb{N}}$ converge ponctuellement vers la fonction $f : E \rightarrow [-\infty, \infty]$ signifie que pour tout $x \in E, \lim_{n \rightarrow \infty} f_n(x) = f(x)$. On voit donc que

$$\lim_{n \rightarrow \infty} f_n = f \text{ ponctuellement} \iff \liminf_{n \rightarrow \infty} f_n = \limsup_{n \rightarrow \infty} f_n = f. \quad (\text{II.4})$$

Les propriétés des limites supérieures et inférieures s'étendent aux fonctions de manière évidente.

Proposition II.1.6 ◀► Soit (E, \mathcal{E}) , un espace mesurable. Soient $f, g, f_n : E \rightarrow [-\infty, \infty], n \in \mathbb{N}$, des fonctions \mathcal{E} -mesurables. Les assertions suivantes sont vérifiées.

- (i) $f \wedge g, g \vee f, f_+, f_-$ et $|f|$ sont \mathcal{E} -mesurables.
- (ii) $\inf f_n, \sup f_n, \liminf f_n$ et $\limsup f_n$ sont également \mathcal{E} -mesurables.
- (iii) Si $\lim_{n \rightarrow \infty} f_n = f_\infty$ ponctuellement alors f_∞ est \mathcal{E} -mesurable.
- (iv) On suppose que les f_n sont à valeurs dans $[0, \infty]$. Alors $\sum_{n \in \mathbb{N}} f_n : E \rightarrow [0, \infty]$ est bien définie et est une fonction \mathcal{E} -mesurable.

Preuve : il est facile de montrer que $\{[-\infty, a]; a \in \mathbb{R}\}$ engendre les Boréliens de $[-\infty, \infty]$. On observe ensuite que pour tout $a \in \mathbb{R}, \{f \vee g \leq a\} = \{f \leq a\} \cap \{g \leq a\} \in \mathcal{E}$. La proposition II.1.2 implique alors que $f \vee g$ est \mathcal{E} -mesurable. Comme $f \wedge g = -((-f) \vee (-g))$, la fonction $f \wedge g$ est également \mathcal{E} -mesurable ainsi que f_+ et f_- . On rappelle que la valeur absolue est continue sur $[-\infty, \infty]$. Elle est donc Borel-mesurable, ce qui entraîne que $|f|$ est \mathcal{E} -mesurable par composition de fonctions mesurables. Cela montre (i).

Prouvons (ii) : on remarque que pour tout $a \in \mathbb{R}, \{\sup_{n \in \mathbb{N}} f_n \leq a\} = \bigcap_{n \in \mathbb{N}} \{f_n \leq a\} \in \mathcal{E}$, ce qui montre que $\sup_{n \in \mathbb{N}} f_n$ est \mathcal{E} -mesurable par la proposition II.1.2. Il en est de même pour $\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}} (-f_n)$. Pour tout $n \in \mathbb{N}$, on pose $g_n = \inf_{p \geq n} f_p$. Ce qui précède montre que g_n est \mathcal{E} -mesurable. Or par définition, $\sup_{n \in \mathbb{N}} g_n = \liminf_{n \rightarrow \infty} f_n$: ce qui précède entraîne donc que $\liminf_{n \rightarrow \infty} f_n$ est \mathcal{E} -mesurable. Il en est de même pour $\limsup_{n \rightarrow \infty} f_n = -\liminf_{n \rightarrow \infty} (-f_n)$, ce qui prouve (ii). Le point (iii) est une conséquence immédiate de (II.4) et de (ii).

Le point (iv) découle de la proposition II.1.5 et de (ii) appliqué à $h_n = \sum_{0 \leq k \leq n} f_k$ qui satisfait $\sum_{n \in \mathbb{N}} f_n = \sup_{n \in \mathbb{N}} h_n$. ■

◀► **Fonctions étagées.** Soit (E, \mathcal{E}) , un espace mesurable. Soit $A \subset E$. La fonction indicatrice de $A, \mathbf{1}_A : E \rightarrow \mathbb{R}$ est définie par

$$\forall x \in E, \quad \mathbf{1}_A(x) = 1 \text{ si } x \in A \quad \text{et} \quad \mathbf{1}_A(x) = 0 \text{ si } x \notin A.$$

On voit que pour tout $a \in \mathbb{R}, \{\mathbf{1}_A \leq a\} = \emptyset$ si $a < 0, \{\mathbf{1}_A \leq a\} = E \setminus A$ si $0 \leq a < 1$ et $\{\mathbf{1}_A \leq a\} = E$ si $a \geq 1$. Cela montre que $\mathbf{1}_A$ est \mathcal{E} -mesurable ssi $A \in \mathcal{E}$.

Une fonction $s : E \rightarrow \mathbb{R}$ est étagée (ou simple) si c'est une combinaison linéaire de fonctions indicatrices, c'est-à-dire s'il existe $A_1, \dots, A_n \subset E$ et $c_1, \dots, c_n \in \mathbb{R}$ tels que

$$s = \sum_{1 \leq k \leq n} c_k \mathbf{1}_{A_k}. \quad (\text{II.5})$$

Si les A_k dans l'écriture ci-dessus sont dans \mathcal{E} , alors le lemme II.1.5 implique que s est \mathcal{E} -mesurable. On remarque que l'écriture d'une fonction étagée s sous la forme (II.5) n'est pas unique : par exemple, pour tout $A \subset E$ on peut écrire $\mathbf{1}_E = \mathbf{1}_A + \mathbf{1}_{E \setminus A}$.

On remarque également qu'une fonction étagée n'a qu'un nombre fini de valeurs possibles. Réciproquement, si une fonction $s : E \rightarrow \mathbb{R}$ est telle que $s(E)$ soit un ensemble fini, on liste ses éléments, sans répétition, par $s(E) = \{c_1, \dots, c_n\}$, on pose $A_k = s^{-1}(\{c_k\})$ et on vérifie que s s'écrit comme dans (II.5). Dans ce cas les A_k sont disjoints deux-à-deux.

Cela montre également que si $s : E \rightarrow \mathbb{R}$ est étagée \mathcal{E} -mesurable, alors il est possible d'écrire s sous la forme (II.5) avec les A_k dans \mathcal{E} et disjoints deux-à-deux. Si de plus s est positive, alors les coefficients c_k peuvent être choisis positifs.

◀► **Approximations des fonctions mesurables positives par des fonctions étagées.** Nous expliquons une procédure d'approximation qui joue un rôle technique essentiel. Soit (E, \mathcal{E}) , un espace mesurable ; soit $f : E \rightarrow [0, \infty]$, une fonction \mathcal{E} -mesurable. On pose

$$\forall n, k \in \mathbb{N}, \quad A_{n,k} := f^{-1}([k2^{-n}, (k+1)2^{-n}[) = \{k2^{-n} \leq f < (k+1)2^{-n}\}.$$

On pose également

$$B_n = f^{-1}([n, \infty]) = \{f \geq n\}.$$

Il est clair que si on fixe $n \in \mathbb{N}$, les ensembles B_n et $A_{n,k}$, $0 \leq k < n2^n$, sont deux-à-deux disjoints et dans \mathcal{E} . On pose alors

$$s_n = n\mathbf{1}_{B_n} + \sum_{0 \leq k < n2^n} k2^{-n}\mathbf{1}_{A_{n,k}}. \quad (\text{II.6})$$

Il est clair que $s_n : E \rightarrow [0, n]$ est une fonction simple \mathcal{E} -mesurable, positive et bornée (par n). On note $[\cdot]$ la fonction partie entière et on vérifie pour tout $x \in E$ les assertions suivantes.

– Si $f(x) < n$, alors

$$s_n(x) = 2^{-n}[2^n f(x)] \leq f(x) < 2^{-n}[2^n f(x)] + 2^{-n} = s_n(x) + 2^{-n}. \quad (\text{II.7})$$

– Si $f(x) \geq n$, alors on a $s_n(x) = n$ et donc $f(x) \geq s_n(x)$.

Dans tous les cas on a $f \geq s_n$, pour tout $n \in \mathbb{N}$. De plus, on vérifie les assertions suivantes.

- Si $f(x) < \infty$, pour tout $n \in \mathbb{N}$ tel que $f(x) < n$, (II.7) implique que $\lim_n s_n(x) = f(x)$.
- Si $f(x) = \infty$, alors on a pour tout $n \in \mathbb{N}$, $s_n(x) = n$ et on a également $\lim_n s_n(x) = f(x) = \infty$.

Cela montre donc que $\lim_{n \rightarrow \infty} s_n = f$ ponctuellement. On montre ensuite que la suite de fonctions s_n croît en n . Pour cela on fixe $x \in E$ et $n \in \mathbb{N}$.

- Supposons d'abord que $k2^{-n-1} \leq f(x) < (k+1)2^{-n-1} \leq n$. On a donc $2^{-n-1}k = s_{n+1}(x)$.
 - Si $k = 2\ell$ est pair, alors $\ell 2^{-n} = k2^{-n-1} \leq f(x) < (k+1)2^{-n-1} < (\ell+1)2^{-n}$. Donc $s_n(x) = s_{n+1}(x)$.
 - Si $k = 2\ell + 1$ est impair, alors $\ell 2^{-n} < k2^{-n-1} \leq f(x) < (k+1)2^{-n-1} = (\ell+1)2^{-n}$ et donc $s_n(x) = \ell 2^{-n} < k2^{-n-1} = s_{n+1}(x)$.

Dans les deux cas, $s_n(x) \leq s_{n+1}(x)$.

- Supposons que $n \leq f(x) < n+1$. Alors $s_n(x) = n$ et $s_{n+1}(x) = k2^{-n-1}$, où $n2^{-n-1} \leq k < (n+1)2^{-n-1}$ et donc $s_n(x) \leq s_{n+1}(x)$.

- Supposons que $f(x) \geq n + 1$, alors $s_n(x) = n < n + 1 = s_{n+1}(x)$.

On a donc montré la proposition suivante.

Proposition II.1.7 ◀▶ Soit $f : E \rightarrow [0, \infty]$, une fonction \mathcal{E} -mesurable. Il existe $s_n : E \rightarrow \mathbb{R}_+$, $n \in \mathbb{N}$, une suite de fonctions étagées \mathcal{E} -mesurables positives telles que

$$\lim_{n \rightarrow \infty} s_n = f \text{ ponctuellement et } 0 \leq s_n \leq s_{n+1} \leq f, n \in \mathbb{N}. \quad (\text{II.8})$$

Comme conséquence de la proposition précédente, nous démontrons le lemme suivant qui s'avère être d'un usage très pratique.

Lemme II.1.8 ◀▶ (dit « technique ») Soit (E, \mathcal{E}) , un espace mesurable. Soit $f : E \rightarrow [0, \infty]$, une fonction \mathcal{E} -mesurable. Alors, il existe $B_n \in \mathcal{E}$ et $c_n \in \mathbb{R}_+$, $n \in \mathbb{N}$, tels que $f = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}$.

◀▶ **Preuve :** on considère la suite de fonctions étagées s_n , $n \in \mathbb{N}$, donnée par (II.6). On pose $t_0 = s_0$ et $t_n = s_n - s_{n-1}$, pour tout $n \in \mathbb{N}^*$. Les fonctions t_n sont étagées positives : elles sont donc combinaison linéaire à coefficients positifs de fonctions indicatrices d'ensembles mesurables. De plus on a $f = \sum_{n \in \mathbb{N}} t_n$, ce qui permet de conclure. ■

Une première application de cette approximation est donnée par la version fonctionnelle du théorème de la classe monotone.

Théorème II.1.9 (Classe monotone, version fonctionnelle) Soit (E, \mathcal{E}) , un espace mesurable. Soit \mathcal{P} , un π -système sur E tel que $\sigma(\mathcal{P}) = \mathcal{E}$. Soit \mathcal{H} , un ensemble d'applications de E dans \mathbb{R} . On suppose que l'ensemble de fonctions \mathcal{H} satisfait les propriétés suivantes.

- \mathcal{H} est un \mathbb{R} -espace vectoriel (il contient donc la fonction nulle).
- Pour tout $C \in \mathcal{P}$, on a $\mathbf{1}_C \in \mathcal{H}$.
- Si $f_n \in \mathcal{H}$, $n \in \mathbb{N}$, et $M \in \mathbb{R}_+$ sont tels que $0 \leq f_n \leq f_{n+1} \leq M$, alors, $\sup_{n \in \mathbb{N}} f_n \in \mathcal{H}$.

Alors, \mathcal{H} contient toutes les fonctions réelles \mathcal{E} -mesurables bornées.

Preuve : on pose $\mathcal{L} = \{B \subset E : \mathbf{1}_B \in \mathcal{H}\}$. Les propriétés (a), (b) et (c) permettent facilement de vérifier que \mathcal{L} est une classe monotone contenant \mathcal{P} . Le théorème de la classe monotone I.1.6 (page 6) implique que $\mathcal{E} = \sigma(\mathcal{P}) \subset \mathcal{L}$. Donc pour tout $B \in \mathcal{E}$, $\mathbf{1}_B \in \mathcal{H}$.

Soit $f : E \rightarrow [0, M]$, une fonction \mathcal{E} -mesurable. Le lemme II.1.8 implique qu'il existe $c_n \in \mathbb{R}_+$ et $B_n \in \mathcal{E}$, $n \in \mathbb{N}$, tels que $f = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}$. Pour tout $n \in \mathbb{N}$, on pose $f_n = \sum_{0 \leq k \leq n} c_k \mathbf{1}_{B_k}$. Comme \mathcal{H} est un espace vectoriel, $f_n \in \mathcal{H}$, pour tout $n \in \mathbb{N}$. De plus on a $0 \leq f_n \leq f_{n+1} \leq M$. Donc $f = \sup_{n \in \mathbb{N}} f_n \in \mathcal{H}$, par (c). L'espace vectoriel \mathcal{H} contient donc toutes les fonctions \mathcal{E} -mesurables positives bornées. Si $f : E \rightarrow \mathbb{R}$ est \mathcal{E} -mesurable bornée, alors que qui précède montre que $f_+, f_- \in \mathcal{H}$ et puisque \mathcal{H} est un espace vectoriel $f = f_+ - f_-$ est dans \mathcal{H} . ■

II.1.b Variables aléatoires : interprétation probabiliste.

La définition suivante n'en est pas vraiment une : il s'agit plutôt d'introduire le vocabulaire probabiliste.

Définition II.1.3 ◀▶ (Variable aléatoire) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit (E, \mathcal{E}) , un espace mesurable. Une fonction $X : \Omega \rightarrow E$ qui est $(\mathcal{F}, \mathcal{E})$ -mesurable est, dans le contexte des probabilités, plutôt appelée une *variable aléatoire* $(\mathcal{F}, \mathcal{E})$ -mesurable. On utilise l'abréviation « v.a. » pour variable aléatoire. □

◀► **Notations importantes.** Les variables aléatoires sont souvent notées par des lettres capitales scriptes de la fin de l'alphabet X, Y, Z . Soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable. On utilise la notation suivante

$$\{X \in C\} := X^{-1}(C) = \{\omega \in \Omega : X(\omega) \in C\}.$$

On omet les accolades $\{ \}$ le plus souvent possible. Par exemple on écrit « $\mathbf{P}(X \in C)$ » plutôt que « $\mathbf{P}(\{X \in C\})$ » et on écrit rarement « $\mathbf{P}(X^{-1}(C))$ ». Si Y est une autre variable aléatoire $(\mathcal{F}, \mathcal{E})$ -mesurable, pour tout $D \in \mathcal{E}$, on écrit

$$\mathbf{P}(X \in C ; Y \in D) \text{ au lieu de } \mathbf{P}(X^{-1}(C) \cap Y^{-1}(D)).$$

◀► **Conventions importantes.** Lorsque $E = \mathbb{R}, \mathbb{C}, [0, \infty], [-\infty, \infty], \mathbb{R}^d$, et que \mathcal{E} est la tribu des Boréliens correspondante, on omet de préciser la tribu de l'espace d'arrivée : on parle simplement d'une variable aléatoire \mathcal{F} -mesurable (réelle, complexe, positive ...).

◀► **Interprétation.** Une variable aléatoire sélectionne un aspect du phénomène aléatoire : un résultat possible $\omega \in \Omega$ peut être assez compliqué et contenir beaucoup d'informations, alors que $X(\omega)$ n'extrait qu'un aspect de ω .

Exemple II.1.2 (*Jet de deux dés*) On jette deux dés, un rouge et un bleu. L'espace de probabilité peut correspondre à

$$\Omega = \{1, \dots, 6\}^2, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbf{P} = \text{equi-probabilité sur } \Omega.$$

On note X_1 le résultat du dé rouge et X_2 le résultat du dé bleu. On a donc

$$\forall \omega = (r, b) \in \Omega, \quad X_1(\omega) = r \quad \text{et} \quad X_2(\omega) = b.$$

On voit donc que $X_1 : \Omega \rightarrow \{1, \dots, 6\}$ est une v.a. \mathcal{F} -mesurable. De même, pour $X_2 : \Omega \rightarrow \{1, \dots, 6\}$. La somme des résultats des dés est également une variable aléatoire $S : \Omega \rightarrow \{2, \dots, 12\}$ donnée par $S = X_1 + X_2$. \square

Tribu engendrée par des variables aléatoires.

Définition II.1.4 ◀► (*tribu engendrée par une famille de v.a.*) Soit (Ω, \mathcal{F}) , un espace mesurable. Soient $(E_i, \mathcal{E}_i), i \in I$, une famille d'espaces mesurables. Pour tout $i \in I$, soit $X_i : \Omega \rightarrow E_i, i \in I$, une v.a. $(\mathcal{F}, \mathcal{E}_i)$ -mesurable. On pose

$$\sigma(X_i, i \in I) = \bigcap \{ \mathcal{G} \subset \mathcal{P}(E) : \mathcal{G} \text{ tribu} : \forall i \in I, X_i \text{ est } (\mathcal{G}, \mathcal{E}_i)\text{-mesurable} \}.$$

La tribu $\sigma(X_i, i \in I)$ sur Ω est appelée la *tribu engendrée par les variables aléatoires* $(X_i)_{i \in I}$. C'est la plus petite tribu sur Ω rendant mesurables toutes les variables aléatoires X_i . La tribu $\sigma(X_i, i \in I)$ dépend des tribus $\mathcal{E}_i, i \in I$, bien que cela n'apparaisse pas dans la notation. \square

Il est facile de vérifier que

$$\sigma(f_i, i \in I) = \sigma(\{ \{X_i \in B\} ; i \in I, B \in \mathcal{E}_i \}).$$

Lorsqu'on ne considère qu'une seule fonction, la tribu qu'elle engendre est particulièrement simple, comme le montre le lemme suivant qui n'est qu'une reformulation, avec des notations probabilistes, du lemme II.1.1, page 27.

Lemme II.1.10 ◀▶ Soit (Ω, \mathcal{F}) , un espace mesurable. Soit (E, \mathcal{E}) , un autre espace mesurable. Soit une fonction $X : \Omega \rightarrow E$. Alors,

$$\sigma(X) = \{\{X \in B\} ; B \in \mathcal{E}\} . \quad (\text{II.9})$$

◀▶ **Preuve** : elle est déjà faite au lemme II.1.1 (page 27) mais on la répète avec les notations probabilistes à titre « d'exercice ». On note $\mathcal{G} = \{\{X \in B\} ; B \in \mathcal{E}\}$. Il est clair que $\mathcal{G} \subset \sigma(X)$. Montrons que \mathcal{G} est une tribu : tout d'abord on observe que $\Omega = \{X \in E\}$ et donc $\Omega \in \mathcal{G}$; soit $B \in \mathcal{E}$; on observe que $\Omega \setminus \{X \in B\} = \{X \in E \setminus B\}$ et comme $E \setminus B \in \mathcal{E}$, cela implique que $\Omega \setminus \{X \in B\} \in \mathcal{G}$; enfin soient $B_n \in \mathcal{G}$, $n \in \mathbb{N}$. Par (II.1), page 27, on a

$$\bigcup_{n \in \mathbb{N}} \{X \in B_n\} = \bigcup_{n \in \mathbb{N}} X^{-1}(B_n) = X^{-1}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \{X \in B\}$$

où $B := \bigcup_{n \in \mathbb{N}} B_n \in \mathcal{E}$; donc $\bigcup_{n \in \mathbb{N}} \{X \in B_n\} \in \mathcal{G}$. On a montré que \mathcal{G} est une classe d'événements contenant l'événement sûr Ω , stable par complémentaire et stable par union dénombrable : c'est donc une tribu. Il est clair que X est $(\mathcal{G}, \mathcal{E})$ -mesurable. Or comme $\mathcal{G} \subset \sigma(X)$ et comme $\sigma(X)$ est la plus petite tribu rendant mesurable X , on en déduit que $\mathcal{G} = \sigma(X)$. ■

Montrons le théorème suivant qui trouve de nombreuses applications en probabilité.

Théorème II.1.11 ◀▶ (Théorème de « représentation ») Soit (Ω, \mathcal{F}) , un espace mesurable. Soit (E, \mathcal{E}) , un autre espace mesurable. Soit $X : \Omega \rightarrow E$, une variable $(\mathcal{F}, \mathcal{E})$ -mesurable. On note $\sigma(X)$ la tribu sur Ω qui est engendrée par X . On considère $Y : \Omega \rightarrow \mathbb{R}$, une variable aléatoires qui est supposée $(\sigma(X), \mathcal{B}(\mathbb{R}))$ -mesurable bornée. Alors, il existe une fonction $\varphi : E \rightarrow \mathbb{R}$ qui est $(\mathcal{E}, \mathcal{B}(\mathbb{R}))$ -mesurable et telle que

$$\forall \omega \in \Omega, \quad Y(\omega) = \varphi(X(\omega)).$$

Autrement dit une v.a. est $\sigma(X)$ -mesurable ssi elle est une fonction (déterministe) de X .

Preuve : on suppose d'abord que Y est $\sigma(X)$ -mesurable à valeurs dans \mathbb{R}_+ . Par le lemme II.1.8, il existe $c_n \in \mathbb{R}_+$ et $B_n \in \sigma(X)$, $n \in \mathbb{N}$, tels que $Y = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}$. Par le lemme II.1.10, il existe ensuite des ensembles $A_n \in \mathcal{E}$ tel que $B_n = \{X \in A_n\}$. On pose alors $\varphi_* = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{A_n}$. Il est clair que $\varphi_* : E \rightarrow [0, \infty]$ est \mathcal{E} -mesurable et on vérifie que pour tout $\omega \in \Omega$:

$$Y(\omega) = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}(\omega) = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{\{X \in A_n\}}(\omega) = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{A_n}(X(\omega)) = \varphi_*(X(\omega)) .$$

Comme $Y(\omega) < \infty$, pour tout $\omega \in \Omega$, on peut poser $\varphi := \varphi_* \mathbf{1}_{\varphi_*^{-1}([0, \infty[)}$, avec la convention que $0 \times \infty = 0$. On voit alors que $\varphi : E \rightarrow \mathbb{R}_+$ est \mathcal{E} -mesurable et telle que $Y = \varphi(X)$ partout sur Ω . Cela prouve l'implication directe du théorème dans le cas des v.a positives.

Le cas d'une v.a. réelle s'obtient simplement en se ramenant au cas positif comme suit : on pose $Y_1 = \max(0, Y)$ and $Y_2 = \max(0, -Y)$, qui sont respectivement les parties positives et négatives de Y . Clairement Y_1 et Y_2 sont des fonctions $\sigma(X)$ -mesurables positives ; le résultat que l'on vient de prouver implique l'existence de φ_1 et φ_2 , deux fonctions de E dans \mathbb{R}_+ qui sot \mathcal{E} -mesurables et telles que $Y_1 = \varphi_1(X)$ et $Y_2 = \varphi_2(X)$ partout sur Ω . On pose alors $\varphi = \varphi_1 - \varphi_2$ qui est clairement \mathcal{E} -mesurable et on vérifie que

$$Y = Y_1 - Y_2 = \varphi_1(X) - \varphi_2(X) = \varphi(X)$$

partout sur Ω . Cela complète la preuve. ■

Exemple II.1.3 (*Une application : description des tribus discrètes*) On applique les énoncés précédents à la description des tribus discrètes. Soit (Ω, \mathcal{G}) , un espace mesuré. On dit que \mathcal{G} est *discrète* (ou encore *atomique*) s'il existe une partition dénombrable $A_n, n \in \mathbb{N}$ de Ω engendrant \mathcal{G} . Explicitement, \mathcal{G} est discrète s'il existe $A_n \subset \Omega, n \in \mathbb{N}$, tels que

$$A_n \cap A_m = \emptyset, \text{ dès que } m \neq n, \quad \Omega = \bigcup_{n \in \mathbb{N}} A_n \quad \text{et} \quad \mathcal{G} = \sigma(\{A_n; n \in \mathbb{N}\}).$$

Il est possible de décrire tous les ensembles de \mathcal{G} de la manière suivante :

$$\mathcal{G} = \left\{ \bigcup_{n \in S} A_n; \quad S \subset \mathbb{N} \right\}, \quad (\text{II.10})$$

où on adopte la convention que $\emptyset =: \bigcup_{n \in \emptyset} A_n$.

Preuve de (II.10) : on pose $X = \sum_{n \in \mathbb{N}} n \mathbf{1}_{A_n}$, qui est clairement \mathcal{G} -mesurable, à valeurs dans \mathbb{N} . On remarque que $\{X \in n\} = A_n$. Cela montre que $\mathcal{G} = \sigma(X)$. Donc, le lemme II.1.10 implique que $\mathcal{G} = \{\{X \in S\}; S \subset \mathbb{N}\}$, ce qui est exactement (II.10). \square

Le théorème II.1.11 montre de plus que si $Y : \Omega \rightarrow \mathbb{R}$ est \mathcal{G} -mesurable, c'est-à-dire $\sigma(X)$ -mesurable par (II.10), il existe $\varphi : \mathbb{N} \rightarrow \mathbb{R}$, tel que $Y = \varphi(X)$. Si on pose $a_n = \varphi(n)$, alors cela implique que $Y = \sum_{n \in \mathbb{N}} a_n \mathbf{1}_{A_n}$. \square

II.2 Espérance.

II.2.a Rappels sur la construction et les propriétés de l'intégrale.

On expose la théorie de Lebesgue qui permet de définir l'intégrale des fonctions mesurables contre une mesure positive, notion étendant considérablement l'intégrale des fonctions réglées ou de Riemann. On procède en trois étapes : la première consiste à définir l'intégrale des fonctions étagées positives, la deuxième étape étend cette notion d'intégrale aux fonctions positives mesurables ; dans la troisième étape, on construit l'intégrale des fonctions mesurables réelles et complexes, avec une restriction essentielle.

Etape I.

On fixe (E, \mathcal{E}, μ) , un espace mesuré. On note \mathbf{S}_+ l'ensemble des fonctions étagées positives \mathcal{E} -mesurables. Il est clair que \mathbf{S}_+ est stable par produit et combinaison linéaire à coefficients positifs. Si $s \in \mathbf{S}_+$, alors il existe $A_1, \dots, A_n \in \mathcal{E}$ deux-à-deux disjoints et $c_1, \dots, c_n \in \mathbb{R}_+$ tels que

$$s = \sum_{k=0}^n c_k \mathbf{1}_{A_k}, \quad (\text{II.11})$$

On fixe $A \in \mathcal{E}$. Il est naturel de définir l'intégrale de s contre μ sur A par la quantité suivante

$$\int_A s d\mu := \sum_{k=0}^n c_k \mu(A \cap A_k), \quad (\text{II.12})$$

avec la convention $0 \times \infty = 0$.

Cohérence de la définition. Il est nécessaire de montrer que (II.12) ne dépend pas d'une écriture particulière de s sous la forme (II.11), car s en admet plusieurs. On indexe $s(E)$ sans répétition par $s(E) = \{\gamma_1, \dots, \gamma_p\}$ et on pose $B_\ell = s^{-1}(\{\gamma_\ell\})$, $1 \leq \ell \leq p$. On a donc $B_\ell = \bigcup_{k \in I_\ell} A_k$, où $I_\ell = \{k \in \{0, \dots, n\} : c_k = \gamma_\ell\}$. Comme les A_k sont disjoints deux-à-deux, on a $\mu(A \cap B_\ell) = \sum_{k \in I_\ell} \mu(A \cap A_k)$. Puisque les I_ℓ forment une partition de $\{0, \dots, n\}$, on a donc

$$\begin{aligned} \sum_{1 \leq \ell \leq p} \gamma_\ell \mu(A \cap B_\ell) &= \sum_{1 \leq \ell \leq p} \sum_{k \in I_\ell} \gamma_\ell \mu(A \cap A_k) = \sum_{1 \leq \ell \leq p} \sum_{k \in I_\ell} c_k \mu(A \cap A_k) \\ &= \sum_{0 \leq k \leq n} c_k \mu(A \cap A_k) = \int_A s \, d\mu. \end{aligned}$$

On observe que le membre de gauche ne dépend pas d'une écriture particulière de s sous la forme (II.11). ■

Proposition II.2.1 Soit (E, \mathcal{E}, μ) , un espace mesuré. On a les propriétés suivantes.

- (i) Pour tous $s_1, s_2 \in \mathbf{S}_+$, $c \in \mathbb{R}_+$ et $A \in \mathcal{E}$, on a $\int_A (s_1 + cs_2) \, d\mu = \int_A s_1 \, d\mu + c \int_A s_2 \, d\mu$.
- (ii) Soit $s \in \mathbf{S}_+$. Pour tout $A \in \mathcal{E}$, $\int_A s \, d\mu = \int_E \mathbf{1}_A s \, d\mu$.
- (iii) Pour tous $A, B \in \mathcal{E}$, $\int_A \mathbf{1}_B \, d\mu = \mu(A \cap B)$.
- (iv) Soit $A \in \mathcal{E}$ tel que $\mu(A) = 0$. Alors, $\int_A s \, d\mu = 0$, pour toute fonction $s \in \mathbf{S}_+$.
- (v) Pour tout $s \in \mathbf{S}_+$, on pose $\nu(A) = \int_A s \, d\mu$, $A \in \mathcal{E}$. Alors, ν est une mesure positive sur (E, \mathcal{E}) .

Preuve : commençons par le point (i). On suppose que

$$s_1 = \sum_{0 \leq k \leq m} a_k \mathbf{1}_{A_k^1} \quad \text{et} \quad s_2 = \sum_{0 \leq j \leq n} b_j \mathbf{1}_{A_j^2},$$

où $a_0, \dots, a_m, b_0, \dots, b_n \in \mathbb{R}_+$, où $A_0^1, \dots, A_m^1 \in \mathcal{E}$ sont deux-à-deux disjoints et $A_0^2, \dots, A_n^2 \in \mathcal{E}$. Par commodité, on pose

$$A_{m+1}^1 = E \setminus \bigcup_{0 \leq k \leq m} A_k^1, \quad a_{m+1} = 0 \quad \text{et} \quad A_{n+1}^2 = E \setminus \bigcup_{0 \leq j \leq n} A_j^2, \quad b_{n+1} = 0$$

si bien que $s_1 = \sum_{0 \leq k \leq m+1} a_k \mathbf{1}_{A_k^1}$ et $s_2 = \sum_{0 \leq j \leq n+1} b_j \mathbf{1}_{A_j^2}$. On a donc

$$s_1 = \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} a_k \mathbf{1}_{A_k^1 \cap A_j^2} \quad \text{et} \quad s_2 = \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} b_j \mathbf{1}_{A_k^1 \cap A_j^2}.$$

et pour tout $A \in \mathcal{E}$, on obtient

$$\int_A s_1 \, d\mu = \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} a_k \mu(A \cap A_k^1 \cap A_j^2) \quad \text{et} \quad \int_A s_2 \, d\mu = \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} b_j \mu(A \cap A_k^1 \cap A_j^2).$$

Donc

$$\begin{aligned}
\int_A (s_1 + c.s_2) d\mu &= \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} (a_k + cb_j) \mu(A \cap A_k^1 \cap A_j^2) \\
&= \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} a_k \mu(A \cap A_k^1 \cap A_j^2) + c \sum_{\substack{0 \leq k \leq m+1 \\ 0 \leq j \leq n+1}} b_j \mu(A \cap A_k^1 \cap A_j^2) \\
&= \int_A s_1 d\mu + c \int_A s_2 d\mu,
\end{aligned}$$

ce qui prouve le point (i). Les points (ii), (iii) et (iv) sont des conséquences immédiates de la définition.

Il reste à montrer (v). Clairement $\nu(\emptyset) = \int_{\emptyset} s d\mu = 0$. Soit $B_n \in \mathcal{E}$, $n \in \mathbb{N}$, une suite d'ensembles deux-à-deux disjoints. Soient $A_0, \dots, A_p \in \mathcal{E}$, des ensembles deux-à-deux disjoints tels que $s = \sum_{k=0}^p c_k \mathbf{1}_{A_k}$, où les $c_k \in \mathbb{R}_+$, $0 \leq k \leq p$. Pour simplifier les notations on pose $B = \bigcup_{n \in \mathbb{N}} B_n$. On a les égalités suivantes.

$$\begin{aligned}
\int_B s d\mu &= \sum_{k=0}^p c_k \mu(A_k \cap B) = \sum_{k=0}^p c_k \mu\left(\bigcup_{n \in \mathbb{N}} (A_k \cap B_n)\right) = \sum_{k=0}^p \sum_{n \in \mathbb{N}} c_k \mu(A_k \cap B_n) \\
&= \sum_{n \in \mathbb{N}} \sum_{k=0}^p c_k \mu(A_k \cap B_n) = \sum_{n \in \mathbb{N}} \int_{B_n} s d\mu.
\end{aligned}$$

En effet, on utilise la sigma additivité de μ à la première ligne et (I.3) à la deuxième. Cela montre bien que ν est sigma-additive. ■

Etape II.

Pour toute application $f : E \rightarrow [0, \infty]$, \mathcal{E} -mesurable et tout $A \in \mathcal{E}$ on pose

$$\int_A f d\mu = \sup \left\{ \int_A s d\mu ; s \in \mathbf{S}_+ : \forall x \in A, 0 \leq s(x) \leq f(x) \right\}. \quad (\text{II.13})$$

Remarquons d'abord que cette définition étend bien l'intégrale des fonctions étagées positives. Les propriétés suivantes sont des conséquences simples de la définition (II.13).

Proposition II.2.2 ◀► Soit (E, \mathcal{E}, μ) , un espace mesuré et soient $f, g : E \rightarrow [0, \infty]$, deux fonctions \mathcal{E} -mesurables.

- (i) Si $f \leq g$, alors $\int_A f d\mu \leq \int_A g d\mu$, pour tout $A \in \mathcal{E}$.
- (ii) Soit $A \in \mathcal{E}$. Si $f(x) = 0$ pour tout $x \in A$, alors $\int_A f d\mu = 0$.
- (iii) Soit $A \in \mathcal{E}$ tel que $\mu(A) = 0$. Alors, $\int_A f d\mu = 0$.
- (iv) Soit $A \in \mathcal{E}$. Alors, $\int_A f d\mu = \int_E \mathbf{1}_A f d\mu$.
- (v) Soit $c \in \mathbb{R}_+$ et $A \in \mathcal{E}$. Alors $\int_A cf d\mu = c \int_A f d\mu$.

Preuve : on remarque que $\{s \in \mathbf{S}_+ : s(x) \leq f(x), x \in A\} \subset \{s \in \mathbf{S}_+ : s(x) \leq g(x), x \in A\}$ si $f \leq g$, ce qui entraîne facilement (i). Soit $s \in \mathbf{S}_+ : s(x) = 0$ sur A , alors clairement $\int_A s d\mu = 0$, ce qui montre (ii). Soit $s \in \mathbf{S}_+$ de la forme (II.11) : $s = \sum_{0 \leq k \leq n} c_k \mathbf{1}_{A_k}$. On suppose que : $0 \leq s(x) \leq f(x)$, pour tout $x \in A$. Si $\mu(A) = 0$, puisque $A \cap A_k \subset A$, on a $0 \leq \mu(A \cap A_k) \leq \mu(A) = 0$ et donc $\mu(A \cap A_k) = 0$ pour tout $0 \leq k \leq n$. On a donc

$$\int_A s d\mu = \sum_{k=0}^n c_k \mu(A \cap A_k) = 0 .$$

Comme cela est vrai pour toute fonction $s \in \mathbf{S}_+$ telle que $s \leq f$ sur A , cela entraîne facilement (iii). Le point (iv) est une conséquence immédiate de la proposition II.2.1 (ii) et du fait que $s \leq f$ sur A ssi $\mathbf{1}_A s \leq \mathbf{1}_A f$ sur E . Le point (v) est une conséquence immédiate de la proposition II.2.1 (i) avec $s_1 = 0$ et du fait que $s \leq f$ sur A ssi $cs \leq cf$ sur A dès que $c > 0$. Le cas $c = 0$ est trivial. ■

On montre ensuite le théorème de convergence monotone, le résultat-clef de la théorie de l'intégration.

Théorème II.2.3 ◀▶ (Convergence monotone) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f_n : E \rightarrow [0, \infty]$, $n \in \mathbb{N}$, une suite d'applications \mathcal{E} -mesurables telles que $f_n \leq f_{n+1}$. On pose $f = \sup_{n \in \mathbb{N}} f_n$, qui est la limite ponctuelle des f_n . Alors,

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

Preuve : par le point (i) de la proposition II.2.2, il est clair que la suite $n \mapsto \int_E f_n d\mu$ est croissante. De plus $f_n \leq f$. En utilisant la même propriété de croissance de l'intégrale, on en déduit que $\int_E f_n d\mu \leq \int_E f d\mu$ et donc

$$\lim_{n \rightarrow \infty} \uparrow \int_E f_n d\mu = \sup_{n \in \mathbb{N}} \int_E f_n d\mu \leq \int_E f d\mu .$$

Il s'agit de montrer l'inégalité contraire. Pour cela on se fixe $a \in]0, 1[$ et $s \in \mathbf{S}_+$ telle que $s \leq f$. Pour tout entier $n \geq 1$, on pose

$$E_n = \{x \in E : f_n(x) \geq as(x)\} .$$

Il est facile de vérifier que les E_n sont \mathcal{E} -mesurables : comme s ne prend que des valeurs finies, $f_n - as$ est bien définie à valeurs dans $\mathbb{R} \cup \{\infty\}$ et \mathcal{E} -mesurable; on remarque que $E_n = (f_n - as)^{-1}([0, \infty]) \in \mathcal{E}$. On voit également que $E_n \subset E_{n+1}$. Montrons que

$$E = \bigcup_{n \in \mathbb{N}} E_n . \tag{II.14}$$

En effet, fixons $x \in E$; on a $0 \leq s(x) \leq f(x)$. Comme s est étagée, on a $s(x) < \infty$. Si $f(x) > 0$, alors $as(x) < f(x)$, et pour tout n assez grand, on a $as(x) < f_n(x)$, ce qui implique que $x \in \bigcup_{n \in \mathbb{N}} E_n$. Si $f(x) = 0$, alors $as(x) = f_n(x) = f(x) = 0$, et donc $x \in \bigcup_{n \in \mathbb{N}} E_n$. Cela prouve (II.14).

On remarque ensuite que pour tout $x \in E$,

$$f_n(x) \geq \mathbf{1}_{E_n}(x) f_n(x) \geq \mathbf{1}_{E_n}(x) as(x) .$$

Par les propriétés (i), (iv) et (v) de la proposition II.2.2,

$$\int_E f_n d\mu \geq a \int_{E_n} s d\mu . \quad (\text{II.15})$$

En utilisant le fait que $A \in \mathcal{E} \mapsto \int_A s d\mu$ est une mesure positive (proposition II.2.1 (iii)) et la propriété (i) de la proposition I.1.7 page 8, on remarque que

$$\lim_{n \rightarrow \infty} \uparrow \int_{E_n} s d\mu = \int_E s d\mu .$$

En passant à la limite dans (II.15) on a pour tout $s \in \mathbf{S}_+$ tel que $s \leq f$ et pour tout $a \in]0, 1[$:

$$\lim_{n \rightarrow \infty} \uparrow \int_E f_n d\mu \geq a \int_E s d\mu .$$

En passant tout d'abord au supremum sur toutes les fonctions étagées positives inférieures à f , on obtient $\lim_{n \rightarrow \infty} \uparrow \int_E f_n d\mu \geq a \int_E f d\mu$. Lorsque a tend vers 1 on obtient bien le résultat désiré. ■

On en déduit ensuite le lemme suivant.

Lemme II.2.4 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. Soient $f, g : E \rightarrow [0, \infty]$, deux applications \mathcal{E} -mesurables. Soit $c \in \mathbb{R}_+$. On a alors

$$\int_E (f + cg) d\mu = \int_E f d\mu + c \int_E g d\mu .$$

Preuve : on utilise l'approximation du lemme II.1.7 (page 33) qui permet d'affirmer l'existence de deux suites $s_n, s'_n \in \mathbf{S}_+$, $n \in \mathbb{N}$ telles que

$$s_n \leq s_{n+1} \leq f, \quad \lim_{n \rightarrow \infty} \uparrow s_n = f \quad \text{et} \quad s'_n \leq s'_{n+1} \leq g, \quad \lim_{n \rightarrow \infty} \uparrow s'_n = g .$$

D'après le point (i) de la proposition II.2.1 page 37,

$$\int_E (s_n + cs'_n) d\mu = \int_E s_n d\mu + c \int_E s'_n d\mu .$$

Le théorème de convergence monotone permet de passer à limite dans cette expression et implique le résultat voulu. ■

On déduit également du théorème de convergence monotone le résultat utile suivant.

Proposition II.2.5 ◀▶ (Interversion \sum / \int positive) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f_n : E \rightarrow [0, \infty]$, $n \in \mathbb{N}$, une suite d'applications \mathcal{E} -mesurables. On a

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

◀▶ **Preuve :** on pose $g_n = \sum_{k=0}^n f_k$. On a donc $0 \leq g_n \leq g_{n+1}$ et $\sup_n g_n = \sum_{n \in \mathbb{N}} f_n$. Par ailleurs le lemme II.2.4 montre que $\int_E g_n d\mu = \sum_{k=0}^n \int_E f_k$. Le théorème de convergence monotone permet de passer à la limite sur les g_n dans cette expression, ce qui implique le résultat voulu. ■

Théorème II.2.6 ◀▶ (Lemme de Fatou) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f_n : E \rightarrow [0, \infty]$, $n \in \mathbb{N}$, une suite d'applications \mathcal{E} -mesurables. On a

$$\int_E (\liminf_{n \rightarrow \infty} f_n) d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu .$$

Preuve : on pose $g_n = \inf_{p \geq n} f_p$, $n \in \mathbb{N}$. On remarque que $0 \leq g_n \leq g_{n+1}$ et que $g_n \leq f_n$. Par ailleurs, on a $\lim_n \uparrow g_n = \liminf_n f_n$. Le théorème de convergence monotone implique alors que

$$\int_E (\liminf_{n \rightarrow \infty} f_n) d\mu = \lim_{n \rightarrow \infty} \uparrow \int_E g_n d\mu .$$

Or on a

$$\lim_{n \rightarrow \infty} \uparrow \int_E g_n d\mu = \liminf_{n \rightarrow \infty} \int_E g_n d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu ,$$

ce qui montre bien le résultat voulu. ■

◀▶ **Notion de μ -presque partout.** Soit (E, \mathcal{E}, μ) , un espace mesuré. Supposons qu'une propriété, notée Prop, dépende de $x \in E$. On dit que la propriété Prop est vérifiée μ -presque partout, μ -p.p. en abrégé, (ou encore qu'elle est vérifiée pour μ -presque tout x) ssi elle est réalisée pour tout $x \in E \setminus N$, où $N \in \mathcal{E}$ est tel que $\mu(N) = 0$.

Ainsi, si $f, g : E \rightarrow [0, \infty]$ sont deux fonctions \mathcal{E} -mesurables, " $f = g$ μ -p.p." signifie qu'il existe $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et $f(x) = g(x)$ pour tout $x \in E \setminus N$. On voit dans ce cas que $\int_A f d\mu = \int_A g d\mu$, pour tout A .

En effet comme $A \cap N \subset N$, on a $\mu(A \cap N) = 0$ et donc $\int_{A \cap N} f d\mu = \int_{A \cap N} g d\mu = 0$ et comme $f \mathbf{1}_{A \cap (E \setminus N)} = g \mathbf{1}_{A \cap (E \setminus N)}$, on a

$$\int_A f d\mu = \int_E f \mathbf{1}_{A \cap N} d\mu + \int_E f \mathbf{1}_{A \cap (E \setminus N)} d\mu = \int_E g \mathbf{1}_{A \cap N} d\mu + \int_E g \mathbf{1}_{A \cap (E \setminus N)} d\mu = \int_A g d\mu .$$

Si $f_n : E \rightarrow [-\infty, \infty]$, $n \in \mathbb{N}$, est une suite de fonctions \mathcal{E} -mesurables, la phrase "la suite f_n converge μ -p.p." signifie simplement que $\limsup_{n \rightarrow \infty} f_n = \liminf_{n \rightarrow \infty} f_n$ μ -presque partout.

Proposition II.2.7 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f : E \rightarrow [0, \infty]$, une fonction \mathcal{E} -mesurable. Les assertions suivantes sont vraies.

(i) Si $\int_E f d\mu = 0$ alors $f = 0$ μ -presque partout.

(ii) Si $\int_E f d\mu < \infty$, alors $f < \infty$ μ -presque partout.

Preuve : on prouve d'abord (i). Pour tout $n \geq 1$, on pose $E_n = \{f \geq 1/n\} \in \mathcal{E}$. On a donc $\{f > 0\} = \bigcup_{n \in \mathbb{N}} E_n$. D'autre part, $f \geq \mathbf{1}_{E_n} f \geq \frac{1}{n} \mathbf{1}_{E_n}$. Par conséquent, $0 = \int_E f d\mu \geq \mu(E_n)/n$. Donc, $\mu(E_n) = 0$. Or par sigma sous-additivité de μ , $\mu(\{f > 0\}) \leq \sum_{n \in \mathbb{N}} \mu(E_n) = 0$, ce qui entraîne le résultat voulu.

On montre (ii) de manière similaire : pour tout $n \in \mathbb{N}$, on pose $B_n = \{f \geq n\} \in \mathcal{E}$. D'autre part $B_{n+1} \subset B_n$ et $\bigcap_{n \in \mathbb{N}} B_n = \{f = \infty\}$. Or $f \geq f \mathbf{1}_{B_n} \geq n \mathbf{1}_{B_n}$, donc

$$\infty > \int_E f d\mu \geq n \int_E \mathbf{1}_{B_n} d\mu = n \mu(B_n) ,$$

c'est-à-dire que $\mu(B_n) \leq \frac{1}{n} \int_E f d\mu \rightarrow 0$. Par la proposition I.1.7 (ii) page 8, on en déduit que $\mu(\{f = \infty\}) = \lim_n \mu(B_n) = 0$, ce qui implique (ii). ■

Etape III.

Définition II.2.1 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f : E \rightarrow \mathbb{R}$, une application \mathcal{E} -mesurable. On dit que f est μ -intégrable ssi

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

On note $\mathcal{L}^1(E, \mathcal{E}, \mu)$ l'ensemble des applications μ -intégrables. □

On se donne $f \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. On rappelle que f_+ et f_- sont les parties positive et négative de f , que $0 \leq f_{+/-} \leq |f|$ et que $f = f_+ - f_-$. Donc, pour tout $A \in \mathcal{E}$, $\int_A f_+ d\mu$ et $\int_A f_- d\mu$ sont deux quantités finies. Il est alors naturel de poser pour tout $A \in \mathcal{E}$:

$$\int_A f d\mu := \int_A f_+ d\mu - \int_A f_- d\mu . \quad (\text{II.16})$$

Proposition II.2.8 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. Soient $f, g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$. On a les propriétés suivantes.

- (i) $\int_A f d\mu = \int_E \mathbf{1}_A f d\mu$, pour tout $A \in \mathcal{E}$.
- (ii) $|\int_E f d\mu| \leq \int_E |f| d\mu$.
- (iii) $\mathcal{L}^1(E, \mathcal{E}, \mu)$ est un \mathbb{R} -espace vectoriel et pour tout $c \in \mathbb{R}$, $\int_E (f + cg) d\mu = \int_E f d\mu + c \int_E g d\mu$.
- (iv) Si $f \leq g$ alors $\int_E f d\mu \leq \int_E g d\mu$.
- (v) Si $A \in \mathcal{E}$ est tel que $\mu(A) = 0$, alors $\int_A f d\mu = 0$.

Preuve : (i) et (v) sont des conséquences directes des cas où f est positive et de la définition (II.16). Pour le second point on observe que

$$\left| \int_E f_+ d\mu - \int_E f_- d\mu \right| \leq \int_E f_+ d\mu + \int_E f_- d\mu = \int_E |f| d\mu .$$

Prouvons le troisième point. On observe d'abord que $|f + cg| \leq |f| + |c| |g|$. Le point (i) de la proposition II.2.2 combiné au lemme II.2.4 implique que $f + cg$ est μ -intégrable et donc que $\mathcal{L}^1(E, \mathcal{E}, \mu)$ est un \mathbb{R} -espace vectoriel.

On considère d'abord le cas $c = 1$. On pose $h = f + g$, ce qui se réécrit par $h_+ - h_- = f_+ - f_- + g_+ - g_-$. Donc on a : $h_+ + f_- + g_- = h_- + f_+ + g_+$ et par le lemme II.2.4 :

$$\int_E h_+ d\mu + \int_E f_- d\mu + \int_E g_- d\mu = \int_E h_- d\mu + \int_E f_+ d\mu + \int_E g_+ d\mu .$$

Chacune de ces intégrales étant finie, en réorganisant le calcul, on obtient :

$$\int_E h d\mu = \int_E h_+ d\mu - \int_E h_- d\mu = \int_E f_+ d\mu - \int_E f_- d\mu + \int_E g_+ d\mu - \int_E g_- d\mu ,$$

qui donne bien $\int_E h d\mu = \int_E f d\mu + \int_E g d\mu$. On termine la preuve du point (iii) en montrant que pour tout $c \in \mathbb{R}$, $\int_E cf d\mu = c \int_E f d\mu$. Pour cela on raisonne comme précédemment en distinguant le cas où c est positif du cas où c est négatif. Le point (iv) se montre en observant que $h := g - f$ est une fonction positive et donc que $\int_E h d\mu \geq 0$. ■

◀► **Intégrales des fonctions à valeurs complexes.** Soit $f : E \rightarrow \mathbb{C}$ qui est \mathcal{E} -mesurable. On note $|\cdot| : \mathbb{C} \rightarrow \mathbb{R}_+$, le module des complexes. C'est une application continue (c'est la norme Euclidienne), elle est donc mesurable. Par composition d'applications mesurables, $|f| : E \rightarrow \mathbb{R}_+$ est \mathcal{E} -mesurable. On dit que f est μ -intégrable ssi $\int_E |f| d\mu < \infty$.

On suppose que $\int_E |f| d\mu < \infty$ et on remarque ensuite que la partie réelle de f , $\text{Re}(f) : E \rightarrow \mathbb{R}$ est \mathcal{E} -mesurable, car $\text{Re} : \mathbb{C} \rightarrow \mathbb{R}$ est continue. De même pour la partie imaginaire de f : $\text{Im}(f)$ est \mathcal{E} -mesurable. Pour tout $x \in E$, on a

$$|\text{Re}(f(x))| \text{ et } |\text{Im}(f(x))| \leq |f(x)| = \sqrt{[\text{Re}(f(x))]^2 + [\text{Im}(f(x))]^2}.$$

Par conséquent $\int_E |\text{Re}(f)| d\mu$ et $\int_E |\text{Im}(f)| d\mu$ sont des quantités inférieures à $\int_E |f| d\mu$, donc des quantités finies. On en déduit donc que $\text{Re}(f)$ et $\text{Im}(f)$ sont des fonctions à valeurs réelles μ -intégrables. On définit l'intégrale de f contre μ par

$$\int_E f d\mu := \int_E \text{Re}(f) d\mu + i \int_E \text{Im}(f) d\mu,$$

où i est le nombre imaginaire. Il est alors facile de vérifier que pour toutes fonctions $f, g : E \rightarrow \mathbb{C}$, \mathcal{E} -mesurables μ -intégrables, on a

$$\forall A \in \mathcal{E}, \int_A f d\mu = \int_A \mathbf{1}_A f d\mu \quad \text{et} \quad \forall c \in \mathbb{C}, \int_E (f + cg) d\mu = \int_E f d\mu + c \int_E g d\mu.$$

Lemme II.2.9 ▶ Soit (E, \mathcal{E}, μ) , un espace mesurable. Soit $f : E \rightarrow \mathbb{C}$, une fonction \mathcal{E} -mesurable et μ -intégrable. Alors,

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Cette inégalité est appelée inégalité triangulaire. De plus, si $|\int_E f d\mu| = \int_E |f| d\mu$, alors il existe $z \in \mathbb{C}$ tel que $|z| = 1$ et tel que μ -p.p. $zf = |f|$.

Preuve : il existe $z \in \mathbb{C}$ tel que $|z| = 1$ et tel que $|\int_E f d\mu| = z \int_E f d\mu$. On a donc

$$\begin{aligned} \left| \int_E f d\mu \right| &= z \int_E f d\mu = \int_E zf d\mu \\ &= \int_E \text{Re}(zf) d\mu \leq \int_E |\text{Re}(zf)| d\mu \leq \int_E |zf| d\mu = \int_E |f| d\mu, \end{aligned}$$

où on utilise l'inégalité triangulaire pour les fonctions réelles et le fait que $|\text{Re}(zf)| \leq |zf|$. Traitons le cas d'égalité : on a donc $\int_E zf d\mu = \int_E |f| d\mu$. En considérant la partie réelle on a $\int_E (|f| - \text{Re}(zf)) d\mu = 0$. Or $|f| - \text{Re}(zf) = |zf| - \text{Re}(zf) \geq 0$, donc μ -p.p. $\text{Re}(zf) = |f|$ par la proposition II.2.7 page 41, ce qui implique que μ -p.p. $zf = |f|$. ■

On utilise également la notation suivante

$$\mathcal{L}^1(E, \mathcal{E}, \mu) = \left\{ f : E \rightarrow \mathbb{C}, \mathcal{E}\text{-mesurable} : \int_E |f| d\mu < \infty \right\}$$

pour l'espace des fonctions à valeurs complexes μ -intégrable et on montre facilement que c'est un \mathbb{C} -espace vectoriel. La notation $\mathcal{L}^1(E, \mathcal{E}, \mu)$ est la même que celle pour l'espace des fonctions réelles μ -intégrables : cette ambiguïté de notation est précisée toutes les fois où cela est nécessaire.

Montrons maintenant le théorème de convergence dominée de Lebesgue.

Théorème II.2.10 ◀▶ (Convergence dominée) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f_n : E \rightarrow \mathbb{C}$, $n \in \mathbb{N}$, une suite de fonctions \mathcal{E} -mesurables qui satisfont les hypothèses suivantes.

- (a) Il existe $f : E \rightarrow \mathbb{R}$, \mathcal{E} -mesurable telle que $\lim_n f_n = f$ μ -p.p.
- (b) Il existe $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ telle que $|f_n| \leq g$, μ -p.p. pour tout $n \in \mathbb{N}$.

Alors, les assertions suivantes sont vérifiées.

- (i) Les fonctions f_n , $n \in \mathbb{N}$, et f sont μ -intégrables.
- (ii) $\lim_{n \rightarrow \infty} \int_E |f - f_n| d\mu = 0$.
- (iii) $\lim_{n \rightarrow \infty} \int_E f_n d\mu = \int_E f d\mu$.

◀▶ **Preuve** : on suppose d'abord que (a) et (b) ont lieu pour tout x et non pour μ -presque tout x . En passant à la limite dans (b), on a $|f| \leq g$ et donc $\int_E |f_n| d\mu$ et $\int_E |f| d\mu \leq \int_E g d\mu < \infty$, ce qui prouve (i).

On remarque ensuite que $|f_n - f| \leq |f| + |f_n| \leq 2g$. Si on pose $h_n = 2g - |f - f_n|$, alors h_n est positive, \mathcal{E} -mesurable et μ -intégrable. De plus $\liminf_n h_n = \lim_n h_n = 2g$. En appliquant le lemme de Fatou à la suite h_n , on obtient

$$\begin{aligned} \int_E 2g d\mu &= \int_E (\liminf_n h_n) d\mu \leq \liminf_{n \rightarrow \infty} \int_E h_n d\mu = \liminf_{n \rightarrow \infty} \int_E (2g - |f - f_n|) d\mu \\ &\leq \int_E 2g d\mu - \limsup_{n \rightarrow \infty} \int_E |f - f_n| d\mu, \end{aligned}$$

ce qui implique que $\limsup_{n \rightarrow \infty} \int_E |f - f_n| d\mu \leq 0$. Cela entraîne (ii), qui implique (iii) car

$$\left| \int_E f d\mu - \int_E f_n d\mu \right| \leq \int_E |f - f_n| d\mu,$$

par l'inégalité de la proposition II.2.8 (ii).

Revenons à l'énoncé général : il existe $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et tel que pour tout $x \in E \setminus N$, on ait $\lim_n f_n(x) = f(x)$ et $|f_n(x)| \leq g(x)$, pour tout $n \in \mathbb{N}$. On pose alors $f' = f \mathbf{1}_{E \setminus N}$ et $f'_n = f_n \mathbf{1}_{E \setminus N}$, pour tout n . On a donc $f = f'$ et $f_n = f'_n$, pour tout $n \in \mathbb{N}$ μ -presque partout. On remarque alors que f' et les f'_n satisfont (a) et (b) pour tout $x \in E$ et on conclut en remarquant que $\int_E |f' - f'_n| d\mu = \int_E |f - f_n| d\mu$, $\int_E f' d\mu = \int_E f d\mu$ et $\int_E f'_n d\mu = \int_E f_n d\mu$, pour tout $n \in \mathbb{N}$. ■

Théorème II.2.11 ◀▶ (Interversion $\sum / \int \mathcal{L}^1$) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f_n : E \rightarrow \mathbb{C}$, $n \in \mathbb{N}$, une suite d'applications \mathcal{E} -mesurables. On suppose que $\sum_{n \in \mathbb{N}} \int_E |f_n| d\mu < \infty$. Alors, les assertions suivantes sont vraies.

- (i) Pour μ -presque tout x , la série $\sum_{n \in \mathbb{N}} f_n(x)$ est absolument convergente et il existe $h \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ tel que $h = \sum_{n \in \mathbb{N}} f_n$ μ -p.p.
- (ii) On a $\lim_{n \rightarrow \infty} \int_E |h - (f_0 + \dots + f_n)| d\mu = 0$.
- (iii) $\int_E h d\mu = \sum_{n \in \mathbb{N}} \int_E f_n d\mu$.

Preuve : on pose $g = \sum_{n \in \mathbb{N}} |f_n|$ qui est \mathcal{E} -mesurable à valeurs dans $[0, \infty]$. D'après les hypothèses et le théorème II.2.5, g est μ -intégrable. On pose $N = g^{-1}(\{\infty\})$. On a $N \in \mathcal{E}$ et par la proposition II.2.7 (ii) page 41, $\mu(N) = 0$. Pour tout $x \in E \setminus N$, $\sum_{n \in \mathbb{N}} f_n(x)$ est donc absolument convergente. On pose par exemple $h_0 = \limsup_n \sum_{0 \leq k \leq n} \operatorname{Re}(f_k)$ et $h_1 = \limsup_n \sum_{0 \leq k \leq n} \operatorname{Im}(f_k)$, qui sont à valeurs dans $[-\infty, \infty]$ et \mathcal{E} -mesurables. Si $x \in E \setminus N$, $h_0(x), h_1(x) \in \mathbb{R}$ et $h_0(x) + ih_1(x) = \sum_{n \in \mathbb{N}} f_n(x)$. On pose alors $h = (h_0 + ih_1) \mathbf{1}_{E \setminus N}$ qui est à valeurs dans \mathbb{C} et qui est \mathcal{E} -mesurable telle que μ -p.p. $h = \sum_{n \in \mathbb{N}} f_n$. On pose $h_n = \sum_{0 \leq k \leq n} f_k$. On a donc μ -p.p. : $|h_n| \leq g$ et $\lim_n h_n = h$, $n \in \mathbb{N}$. Le théorème de convergence dominée appliqué aux h_n entraîne le résultat voulu. ■

II.2.b L'espérance mathématique.

Il ne s'agit pas d'une vraie définition mais plutôt d'une notation probabiliste.

Définition II.2.2 ◀▶ (*Espérance d'une variable aléatoire*) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité.

(a) Soit $X : \Omega \rightarrow [0, \infty]$, une v.a. \mathcal{F} -mesurable. Son *espérance* est son intégrale par rapport à \mathbf{P} . On note cette intégrale $\mathbf{E}[X]$:

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) d\mathbf{P}(\omega) . \quad (\text{II.17})$$

(b) Soit $X : \Omega \rightarrow \mathbb{R}$ ou \mathbb{C} , v.a. \mathcal{F} -mesurable est dite intégrable si elle est \mathbf{P} -intégrable, c'est-à-dire que $\mathbf{E}[|X|] < \infty$. Son *espérance* est son intégrale par rapport à \mathbf{P} . On note cette intégrale $\mathbf{E}[X]$, c'est-à-dire que l'on a (II.17). \square

L'espérance, en tant qu'intégrale contre une mesure de probabilité satisfait les propriétés élémentaires suivantes qui sont simplement une réécriture de résultats sur l'intégrale.

◀▶ **Fait 1.** L'espérance est linéaire positive. Soient $X, Y : \Omega \rightarrow \mathbb{R}$, deux v.a. réelles \mathcal{F} -mesurables et \mathbf{P} -intégrables. Soit $c \in \mathbb{R}$. Alors $X + cY$ est une v.a. \mathcal{F} -mesurables et \mathbf{P} -intégrable et

$$\mathbf{E}[X + cY] = \mathbf{E}[X] + c\mathbf{E}[Y] .$$

◀▶ **Fait 2.** Soit $X : \Omega \rightarrow \mathbb{R}$, une v.a. réelle \mathcal{F} -mesurable et \mathbf{P} -intégrable. Alors, on a l'inégalité triangulaire :

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

◀▶ **Fait 3.** Soit $X : \Omega \rightarrow \mathbb{R}$, une v.a. réelle \mathcal{F} -mesurable et bornée : c'est-à-dire qu'il existe une constante $c \in \mathbb{R}$ telle que pour tout $\omega \in \Omega$, $|X(\omega)| \leq c$. Alors X est \mathbf{P} -intégrable et on a $|\mathbf{E}[X]| \leq c$. En effet sin on note simplement c la variable constante égale à c on a

$$\mathbf{E}[c] = \int_{\Omega} c d\mathbf{P}(\omega) = c \int_{\Omega} 1 d\mathbf{P}(\omega) = c\mathbf{P}(\Omega) = c.$$

◀▶ **Fait 4.** Pour tout $A \in \mathcal{F}$, on a

$$\forall A \in \mathcal{F}, \quad \mathbf{E}[1_A] = \mathbf{P}(A) .$$

◀▶ **Fait 5. Vocabulaire :** Dans les énoncés suivants *presque sûrement* signifie \mathbf{P} -presque partout. On abrège souvent *presque sûrement* par "p.s.". Soit $X : \Omega \rightarrow [0, \infty]$, une v.a. \mathcal{F} -mesurable. Alors les implications suivantes sont vérifiées.

- Si $\mathbf{E}[X] = 0$, alors \mathbf{P} -p.s. $X = 0$.
- Si $\mathbf{E}[X] < \infty$, alors \mathbf{P} -p.s. $X < \infty$.

Nous énonçons ensuite les principaux théorèmes de passage à la limite sous l'intégrale avec les notations probabilistes. Soit $(X_n)_{n \geq 0}$, une suite de v.a. \mathcal{F} -mesurables, à valeurs dans $[0, \infty]$, \mathbb{R} ou \mathbb{C} .

◀▶ **Convergence monotone.** On suppose les v.a. X_n à valeurs dans $[0, \infty]$ et telles que $X_n \leq X_{n+1}$, pour tout $n \in \mathbb{N}$. On pose $X_\infty = \sup_n X_n$. Alors, $\lim_n \mathbf{E}[X_n] = \mathbf{E}[X_\infty]$.

◀▶ **Interversion positive.** On suppose les v.a. X_n à valeurs dans $[0, \infty]$. Alors,

$$\mathbf{E}\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} \mathbf{E}[X_n].$$

◀▶ **Fatou.** On suppose les v.a. X_n à valeurs dans $[0, \infty]$. Alors,

$$\mathbf{E}[\liminf_n X_n] \leq \liminf_n \mathbf{E}[X_n].$$

◀▶ **Convergence dominée.** On suppose les v.a. X_n à valeurs dans \mathbb{C} . On fait les hypothèses suivantes.

(a) Il existe une v.a. réelle X_∞ telle que $\lim_n X_n = X_\infty$ p.s.

(b) Il existe une v.a. intégrable Z telle que p.s. $|X_n| \leq Z$, pour tout $n \geq 0$.

Alors, X_∞ est intégrable, $\lim_n \mathbf{E}[|X_n - X_\infty|] = 0$, et $\lim_n \mathbf{E}[X_n] = \mathbf{E}[X_\infty]$.

◀▶ **Interversion L^1 .** On suppose les v.a. X_n à valeurs dans \mathbb{C} . On suppose que $\sum_{n \geq 0} \mathbf{E}[|X_n|] < \infty$. Alors, $\sum_{n \geq 0} X_n$ est une v.a. bien définie p.s. et intégrable. On a de plus

$$\mathbf{E}\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} \mathbf{E}[X_n].$$

II.3 Loi d'une variable.

II.3.a Cas des variables discrètes.

Dans toute cette section $(\Omega, \mathcal{F}, \mathbf{P})$ désigne un espace de probabilité sur lequel sont définies toutes les variables aléatoires que l'on considère.

Définition II.3.1 ▶ (Variables discrètes) Soit (E, \mathcal{E}) , un espace mesurable et soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable. Elle est dite *discrète* si elle ne prend qu'un nombre dénombrable de valeurs, c'est-à-dire que l'ensemble image $X(\Omega)$ est dénombrable. \square

Lemme II.3.1 ▶ (Critère de mesurabilité des v.a. discrète) (E, \mathcal{E}) , un espace mesurable tel que pour tout $x \in E$, $\{x\} \in \mathcal{E}$. Soit $X : \Omega \rightarrow E$ une fonction dont l'ensemble image $X(\Omega)$ est dénombrable. On énumère $X(\Omega)$ par une suite $(y_n)_{n \in \mathbb{N}}$ (avec de possibles répétitions) :

$$X(\Omega) = \{y_n; n \in \mathbb{N}\}.$$

Alors, X est $(\mathcal{F}, \mathcal{E})$ -mesurable ssi

$$\forall n \in \mathbb{N}, \quad \{X = y_n\} \in \mathcal{F}. \quad (\text{II.18})$$

◀► **Preuve** : supposons d'abord que X est $(\mathcal{F}, \mathcal{E})$ -mesurable, c'est-à-dire que pour tout $B \in \mathcal{E}$, $\{X \in B\} \in \mathcal{F}$. En particulier, pour tout $n \in \mathbb{N}$, si on choisit $B = \{y_n\}$, on a $\{X = y_n\} = \{X \in \{y_n\}\} \in \mathcal{F}$, ce qui montre (II.18).

Réciproquement, on suppose (II.18). Soit $B \in \mathcal{E}$. On observe alors que

$$\{X \in B\} = \bigcup_{n \in \mathbb{N}: y_n \in B} \{X = y_n\}$$

qui est bien un événement de \mathcal{F} car chaque événement $\{X = y_n\}$ appartient à \mathcal{F} et que \mathcal{F} est stable par union dénombrable. Cela montre bien que pour tout $B \in \mathcal{E}$, $\{X \in B\} \in \mathcal{F}$ et donc que X est $(\mathcal{F}, \mathcal{E})$ -mesurable. ■

Soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable. Elle est supposée discrète. On note $S = X(\Omega)$ l'ensemble dénombrable de ses valeurs possibles. Par conséquent S est fini ou en bijection avec \mathbb{N} . Soit $f = E \rightarrow \mathbb{R}_+$, une fonction \mathcal{E} -mesurable. Alors $f(X) : \Omega \rightarrow \mathbb{R}_+$ est une v.a. \mathcal{F} -mesurable qui est discrète car l'ensemble de ses valeurs possibles est $f(S)$ qui est dénombrable puisque S l'est. Donnons une formule pour l'espérance $\mathbf{E}[f(X)]$.

On remarque tout d'abord que les événements $\{X = y\}$, $y \in S$, forment une *partition dénombrable* de Ω : en effet, soit $\omega \in \Omega$, on pose $z \in X(\omega)$, et on a bien $z \in S$ et $\omega \in \{X = z\}$. Donc pour tout $\omega \in \Omega$, on a $\omega \in \bigcup_{y \in S} \{X = y\}$ c'est-à-dire que $\Omega = \bigcup_{y \in S} \{X = y\}$. Ensuite si $y, z \in S$ sont distincts, on ne peut pas avoir $X(\omega) = y$ et $X(\omega) = z$ donc $\{X = y\} \cap \{X = z\} = \emptyset$. On a donc montré que

$$\Omega = \bigcup_{y \in S} \{X = y\} \quad \text{et} \quad \forall y, z \in S, \text{ distincts} \quad \{X = y\} \cap \{X = z\} = \emptyset.$$

On a donc l'égalité de fonctions suivante partout sur Ω

$$1 = \mathbf{1}_\Omega = \sum_{y \in S} \mathbf{1}_{\{X=y\}}.$$

En multipliant cette égalité par $f(X)$, on a l'égalité de fonctions suivante partout sur Ω :

$$f(X) = \sum_{y \in S} f(X) \mathbf{1}_{\{X=y\}}.$$

Or pour tout $\omega \in \Omega$, $f(X(\omega)) \mathbf{1}_{\{X=y\}}(\omega)$ vaut $f(y)$ si $X(\omega) = y$ (c'est-à-dire si $\omega \in \{X = y\}$) et vaut 0 si $X(\omega) \neq y$ (c'est-à-dire si $\omega \notin \{X = y\}$). Autrement dit, pour tout $y \in S$, on a l'égalité de fonctions suivante partout sur Ω :

$$f(X) \mathbf{1}_{\{X=y\}} = f(y) \mathbf{1}_{\{X=y\}}.$$

Donc

$$f(X) = \sum_{y \in S} f(y) \mathbf{1}_{\{X=y\}}.$$

En utilisant l'interversion série/espérance positive on obtient alors

$$\mathbf{E}[f(X)] = \mathbf{E}\left[\sum_{y \in S} f(y) \mathbf{1}_{\{X=y\}}\right] = \sum_{y \in S} \mathbf{E}[f(y) \mathbf{1}_{\{X=y\}}].$$

On fixe $y \in S$. Puisque $f(y)$ est une constante, la linéarité de l'espérance implique que

$$\mathbf{E}[f(y)\mathbf{1}_{\{X=y\}}] = f(y)\mathbf{E}[\mathbf{1}_{\{X=y\}}] = f(y)\mathbf{P}(X=y).$$

On en déduit donc

$$\mathbf{E}[f(X)] = \sum_{y \in S} f(y)\mathbf{P}(X=y).$$

On a donc montré dans le cas d'une fonction f positive la proposition suivante qui donne une formule permettant de calculer l'espérance de fonctions de variables discrètes.

Proposition II.3.2 ◀► Soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable discrète. On note $S = X(\Omega)$ l'ensemble dénombrable de ses valeurs possibles. Soit $f : E \rightarrow \mathbb{R}$, une fonction \mathcal{E} -mesurable qui est soit positive soit bornée. Alors, $f(X) : \Omega \rightarrow \mathbb{R}$ est une v.a. \mathcal{F} -mesurable qui est discrète, soit positive soit bornée, ce qui implique que son espérance est bien définie et on a

$$\mathbf{E}[f(X)] = \sum_{y \in S} f(y)\mathbf{P}(X=y) \quad (\text{II.19})$$

Preuve : il reste à prouver le cas où f est réelle bornée : nous laissons en exercice l'adaptation de la preuve du cas positif à ce cas. ■

Exemple II.3.1 On joue à « pile ou face » deux fois de suite. On note $R_1 = 1$ si le premier lancer est pile et $R_1 = 0$ si le premier lancer est face. On note $R_2 = 1$ si le second lancer est pile et $R_2 = 0$ si le second lancer est face. Ces deux lancers sont « indépendants ». On peut représenter cette expérience par l'espace de probabilité

$$\Omega = \{0, 1\} \times \{0, 1\}, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbf{P} = \text{équi-probabilité sur } \Omega.$$

On a donc

$$\forall \omega = (x, y) \in \Omega, \quad R_1(\omega) = x \quad \text{et} \quad R_2(\omega) = y.$$

Il est clair que R_1 et R_2 sont des variables aléatoires \mathcal{F} -mesurables. On pose $X = (R_1, R_2) : \Omega \rightarrow \mathbb{R}^2$ qui est une v.a. \mathcal{F} -mesurable discrète : on a effectivement $X(\Omega) = \{(0, 0), (0, 1), (1, 0), (1, 1)\} =: S$. On définit ensuite $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$, par $f(x, y) = 2^x 3^y$. C'est une fonction continue donc mesurable. On a donc

$$\begin{aligned} \mathbf{E}[f(X)] &= \sum_{(x,y) \in S} f(x,y)\mathbf{P}(X=(x,y)) \\ &= \mathbf{P}(X=(0,0)) + 3\mathbf{P}(X=(0,1)) + 2\mathbf{P}(X=(1,0)) + 6\mathbf{P}(X=(1,1)) \end{aligned}$$

Or $\mathbf{P}(X=(x,y)) = \frac{1}{\#\Omega} = \frac{1}{4}$ car \mathbf{P} est la probabilité uniforme. On en déduit donc que

$$\mathbf{E}[f(X)] = \frac{1}{4} + \frac{3}{4} + \frac{2}{4} + \frac{6}{4} = \frac{12}{4} = 3$$

ce qui n'a pas d'autre intérêt que de détailler un calcul. □

Dans la section suivante on cherche à généraliser la formule (II.19) à des variables non discrètes. Informellement, si on note $\mathbf{P}(X \in dy)$ la probabilité que X se trouve dans un voisinage de y , en remplaçant la somme par une intégrale on obtient la « formule » suivante :

$$\mathbf{E}[f(X)] = \int_E f(y) \mathbf{P}(X \in dy)$$

Tout le problème consiste à donner un sens clair à $\mathbf{P}(X \in dy)$, ce qui est l'objectif de la section suivante sur les mesures images.

II.3.b Variables de Bernoulli, binomiales, géométriques et de Poisson.

Variables de Bernoulli.

Définition II.3.2 ◀▶ (Loi de Bernoulli) On fixe $p \in [0, 1]$.

- (a) La loi $p\delta_1 + (1-p)\delta_0$ est appelée *loi de Bernoulli de paramètre p* .
- (b) Une v.a. $X : \Omega \rightarrow \{0, 1\}$ est appelée *variable de Bernoulli de paramètre $p = \mathbf{P}(X = 1)$* . □

Si $X : \Omega \rightarrow \{0, 1\}$ est une v.a. de Bernoulli de paramètre p , alors on montre facilement

$$\mathbf{E}[X] = p, \quad \mathbf{var}(X) = p(1-p) \quad \text{et} \quad \mathbf{E}[r^X] = 1 - p + rp, \quad r \in [0, 1]. \quad (\text{II.20})$$

Voir l'exercice II.12.1, page 94.

Variables binomiales.

Définition II.3.3 ◀▶ (Loi binomiale) On fixe $p \in [0, 1]$ et $n \in \mathbb{N}^*$. On rappelle la notation suivante

$$\forall 0 \leq k \leq n, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{k!}.$$

pour les coefficients binomiaux.

- (a) La loi $\sum_{0 \leq k \leq n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$ est appelée *loi binomiale de paramètres (n, p)* .
- (b) Une v.a. $X : \Omega \rightarrow \mathbb{N}$ est appelée *variable binomiale de paramètres (n, p)* si

$$\forall k \in \{0, \dots, n\}, \quad \mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{et} \quad \mathbf{P}(X = k) = 0 \quad \text{si} \quad k > n.$$

Autrement dit X est binomiale si sa loi est binomiale. □

La formule du binôme de Newton permet de calculer facilement la fonction génératrice d'une v.a. binomiale $X : \Omega \rightarrow \mathbb{N}$ de paramètres (n, p) , c'est-à-dire la fonction définie pour tout $r \in [0, 1]$, par

$$\mathbf{E}[r^X] = \sum_{0 \leq k \leq n} \binom{n}{k} p^k (1-p)^{n-k} r^k = (1-p+pr)^n. \quad (\text{II.21})$$

En dérivant une puis deux fois la fonction génératrice en $r = 1$, on obtient

$$\mathbf{E}[X] = np \quad \text{et} \quad \mathbf{var}(X) = np(1-p).$$

Voir l'exercice II.12.1, page 94 pour un calcul plus précis.

Variables géométriques.

Définition II.3.4 ◀▶ (Loi géométrique) Une v.a. $X : \Omega \rightarrow \mathbb{N}$ est (de loi) *géométrique de paramètre de succès $p \in]0, 1[$* si pour tout $k \in \mathbb{N}$, $\mathbf{P}(X = k) = p(1-p)^k$. □

Il est facile de montrer que si X est géométrique de paramètre de succès p , alors

$$\forall r \in [0, 1], \quad \mathbf{E}[r^X] = \sum_{k \in \mathbb{N}} p((1-p)r)^k = \frac{p}{1 - (1-p)r}.$$

En dérivant une puis deux fois la fonction génératrice en $r = 1$, on obtient

$$\mathbf{E}[X] = \frac{1-p}{p}, \quad \mathbf{E}[X^2] = \frac{2(1-p)^2}{p^2} \quad \text{et donc} \quad \mathbf{var}(X) = \frac{(1-p)^2}{p^2}.$$

Voir l'exercice II.12.1, page 94 pour un calcul plus précis.

La proposition suivante est une propriété remarquable des loi géométrique qui est parfois appelée *propriété d'oubli*.

Proposition II.3.3 (Propriété d'oubli) Soit X , une v.a. géométrique de paramètre de succès p . Alors, pour tout $n \in \mathbb{N}$, sous la probabilité conditionnelle $\mathbf{P}(\cdot | X \geq n)$, $X - n$ est également une v.a. géométrique de paramètre de succès p .

Preuve : on remarque d'abord que $\mathbf{P}(X \geq n) = \sum_{k \geq n} p(1-p)^k = (1-p)^n$. Donc

$$\mathbf{P}(X - n = k | X \geq n) = \frac{\mathbf{P}(X = n+k)}{\mathbf{P}(X \geq n)} = \frac{p(1-p)^{n+k}}{(1-p)^n} = p(1-p)^k,$$

ce qui démontre la proposition. ■

La propriété d'oubli permet d'expliquer pourquoi, par exemple, les lois géométriques sont utilisées comme modèles pour la fécondité naturelle : supposons qu'un animal soit fertile tout au long de sa vie (on néglige les périodes d'enfance et de vieillesse) et que les facteurs de fécondités sont constants au cours de sa vie et qu'il ne les contrôle pas. On note X le nombre total de petits de cet animal au cours de sa vie. Le fait que cet animal ait déjà eu au moins n enfants n'influe pas sur le nombre d'enfants supplémentaires qu'il va encore avoir et ce nombre supplémentaire suit exactement la même loi statistique. En formalisant cela on doit donc avoir

$$\forall k, n \in \mathbb{N}, \quad \mathbf{P}(X - n = k | X \geq n) = \mathbf{P}(X = k).$$

On montre facilement que X est nécessairement une v.a. géométrique de paramètre de succès $p = \mathbf{P}(X = 0)$.

Lois de Poisson. On rappelle la définition d'une loi de Poisson.

Définition II.3.5 ◀▶ (Loi de Poisson) Une v.a. $X : \Omega \rightarrow \mathbb{N}$ est (de loi) de Poisson de paramètre $\theta \in]0, \infty[$ si $\mathbf{P}(X = k) = e^{-\theta} \theta^k / k!$, pour tout $k \in \mathbb{N}$. □

On rappelle que si X est de Poisson de paramètre θ , alors

$$\mathbf{E}[r^X] = \exp(-\theta(1-r)), \quad r \in [0, 1]$$

et

$$\mathbf{E}[X] = \theta, \quad \mathbf{E}[X^2] = \theta + \theta^2 \quad \text{et donc} \quad \mathbf{var}(X) = \theta.$$

Voir l'exercice II.12.1, page 94 pour un calcul plus précis. Cela justifie la définition suivante des lois de Poisson dégénérées.

Définition II.3.6 (Loi de Poisson dégénérée) Une v.a. X suit une loi de Poisson de paramètre infini (resp. nul) si $\mathbf{P}(X = \infty) = 1$ (resp. si $\mathbf{P}(X = 0) = 1$). □

II.3.c Rappel sur les mesures images.

On se place dans le cadre général de l'intégration pour rappeler la définition d'une mesure image par une fonction mesurable et des principales propriétés de cette notion. Le lemme suivant en donne tout d'abord la définition.

Lemme II.3.4 ◀▶ (Définition de la mesure image) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit (E', \mathcal{E}') , un espace mesurable. Soit $f : E \rightarrow E'$, une application $(\mathcal{E}, \mathcal{E}')$ -mesurable. On définit $\nu : \mathcal{E}' \rightarrow [0, \infty]$ par

$$\forall B \in \mathcal{E}', \quad \nu(B) = \mu(f^{-1}(B)) .$$

Alors ν est une mesure positive sur (E', \mathcal{E}') qui est appelée la **mesure image de μ par f** . Elle est parfois notée $\nu = \mu \circ f^{-1}$.

◀▶ **Preuve :** soient $B_n \in \mathcal{E}'$, $n \in \mathbb{N}$, des ensembles deux-à-deux disjoints. On remarque que les ensembles $f^{-1}(B_n)$, $n \in \mathbb{N}$, sont également deux-à-deux disjoints et dans \mathcal{E} . D'autre part, $f^{-1}(\bigcup B_n) = \bigcup f^{-1}(B_n)$, par (II.1), page 27. La sigma additivité de μ entraîne alors

$$\nu\left(\bigcup B_n\right) = \mu\left(f^{-1}\left(\bigcup B_n\right)\right) = \mu\left(\bigcup f^{-1}(B_n)\right) = \sum \mu(f^{-1}(B_n)) = \sum \nu(B_n) .$$

Donc ν est sigma additive et il est clair que $\nu(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$. Cela montre que ν est une mesure positive. ■

Exemple II.3.2 Soient E et E' , deux ensembles non-vides et soit $x \in E$. On rappelle que δ_x est la masse de Dirac en x qui est une mesure positive définie sur $\mathcal{P}(E)$, la classe de tous les sous-ensembles de E . On munit E' de la tribu $\mathcal{P}(E')$. Soit $f : E \rightarrow E'$. Il est clair que f est $(\mathcal{P}(E), \mathcal{P}(E'))$ -mesurable. La mesure image de δ_x par f est $\nu = \delta_{f(x)}$, c'est-à-dire la masse de Dirac en $f(x) \in E'$. □

Exemple II.3.3 ◀▶ (Invariance de la mesure de Lebesgue par les translations) On rappelle que $\mathcal{B}(\mathbb{R})$ désigne les Boréliens de \mathbb{R} . Pour tout $x \in \mathbb{R}$ on note T_x l'application $T_x(y) = x + y$, $y \in \mathbb{R}$. Par conséquent T_x est continue donc mesurable. On note ℓ_x la mesure image de la mesure de Lebesgue ℓ par T_x . Par définition on a pour tous $a < b$, $\ell_x([a, b]) = \ell([a - x, b - x]) = b - a$. Par conséquent $\ell_x = \ell$, par unicité de la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Autrement dit, la mesure de Lebesgue est laissée invariante par les translations. □

Le théorème suivant est d'un usage constant en probabilité. Il est connu sous deux noms : le « théorème de changement de variable abstrait » ou le « théorème de transfert ».

Théorème II.3.5 ◀▶ (Changement de variable abstrait ou Transfert) Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit (E', \mathcal{E}') , un espace mesurable. Soit $f : E \rightarrow E'$, une application $(\mathcal{E}, \mathcal{E}')$ -mesurable. On note ν la mesure image de μ par f . Alors, pour toute fonction $h \in \mathcal{L}^1(E', \mathcal{E}', \nu)$,

$$\int_{E'} h(y) \nu(dy) = \int_E h(f(x)) \mu(dx) . \tag{II.22}$$

Ce résultat est également vrai pour toute fonction $h : E' \rightarrow [0, \infty]$ qui est \mathcal{E}' -mesurable.

◀► **Preuve** : on considère d'abord le cas où $h = \mathbf{1}_B$, avec $B \in \mathcal{E}'$. Alors, on remarque que $\mathbf{1}_B(f(x)) = \mathbf{1}_{f^{-1}(B)}(x)$, pour tout $x \in E$; cela entraîne bien (II.22) pour les fonctions indicatrices, par définition de la mesure image. Soit $h : E' \rightarrow [0, \infty]$, une application \mathcal{E}' -mesurable. Le lemma II.1.8 page 33 implique l'existence de $B_n \in \mathcal{E}'$ et $c_n \in \mathbb{R}_+$, $n \in \mathbb{N}$, tels que $h = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}$. La proposition II.2.5 d'interversion série/intégrale positive implique alors que

$$\begin{aligned} \int_{E'} h d\nu &= \int_{E'} \left(\sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n} \right) d\nu = \sum_{n \in \mathbb{N}} c_n \int_{E'} \mathbf{1}_{B_n} d\nu = \sum_{n \in \mathbb{N}} c_n \int_E \mathbf{1}_{B_n}(f(x)) d\mu(x) \\ &= \int_E \left(\sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}(f(x)) \right) d\mu(x) = \int_E h(f(x)) d\mu(x). \end{aligned}$$

On passe au cas des fonctions réelles ν -intégrables en considérant leur partie positive et leur partie négative : les détails sont laissés en exercice. ■

II.3.d Loi d'une v.a. : cas général ; transfert.

Dans toute cette section $(\Omega, \mathcal{F}, \mathbf{P})$ désigne un espace de probabilité sur lequel sont définies toutes les variables aléatoires que l'on considère. On définit la loi d'une variable aléatoire.

Définition II.3.7 ▶ (Loi d'une variable aléatoire) Soit (E, \mathcal{E}) , un espace mesurable. Soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable. La loi de X est la mesure image de \mathbf{P} par X . Dans ce cours, on la note μ_X .

La loi de μ_X est donc une mesure de probabilité $\mu_X : \mathcal{E} \rightarrow [0, 1]$ donnée par

$$\forall C \in \mathcal{E}, \quad \mu_X(C) = \mathbf{P}(X \in C).$$

Dans certains livres, au lieu de noter μ_X la loi de X elle est notée $\mathbf{P}(X \in dy)$. □

Le théorème II.3.5 (dit de *transfert*, page 51) se reformule comme suit.

Théorème II.3.6 ▶ (Transfert) Soit (E, \mathcal{E}) , un espace mesurable. Soit $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable dont on note μ_X la loi. Soit $f : E \rightarrow [0, \infty]$, une fonction qui est \mathcal{E} -mesurable. On a

$$\mathbf{E}[f(X)] = \int_E f(y) \mu_X(dy).$$

Cette égalité reste vraie lorsque f est à valeurs complexes si $f(X)$ est \mathcal{F} -mesurable intégrable, ce qui est équivalent à ce que f soit μ_X -intégrable.

Par exemple si X est une v.a. complexe \mathcal{F} -mesurable et \mathbf{P} -intégrable, on a

$$\mathbf{E}[X] = \int_{\mathbb{C}} y \mu_X(dy) \quad \text{et} \quad \mathbf{E}[|X|^2] = \int_{\mathbb{C}} |y|^2 \mu_X(dy).$$

Si on suppose que (E, \mathcal{E}) , un espace mesurable et que $X : \Omega \rightarrow E$ est une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable discrète, on note $S = X(\Omega)$ l'ensemble de ses valeurs possibles. Alors la loi de μ_X est la somme pondérée de masses de Dirac suivante :

$$\mu_X = \sum_{y \in S} a_y \delta_y \quad \text{où pour tout } y \in S \text{ on a posé } a_y := \mathbf{P}(X = y).$$

Exemple II.3.4 On reprend l'exemple II.3.1, page 48 dans lequel on joue à « pile ou face » deux fois de suite. On note $R_1 = 1$ si le premier lancer est pile et $R_1 = 0$ si le premier lancer est face. On note $R_2 = 1$ si le second lancer est pile et $R_2 = 0$ si le second lancer est face. Ces deux lancers sont « indépendants ». On peut représenter cette expérience par l'espace de probabilité

$$\Omega = \{0, 1\} \times \{0, 1\}, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad \mathbf{P} = \text{équi-probabilité sur } \Omega.$$

On a donc

$$\forall \omega = (x, y) \in \Omega, \quad R_1(\omega) = x \quad \text{et} \quad R_2(\omega) = y.$$

On a donc

$$\mu_{R_1}(\{0\}) = \mathbf{P}(R_1=0) = 1/2 \quad \text{et} \quad \mu_{R_1}(\{1\}) = \mathbf{P}(R_1=1) = 1/2.$$

On a donc $\mu_{R_1} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ qui est une loi sur $\{0, 1\}$ (ou encore sur \mathbb{R} qui contient $\{0, 1\}$). De même on vérifie que $\mu_{R_2} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Cela montre que $\mu_{R_1} = \mu_{R_2}$. Autrement dit R_1 et R_2 sont deux variables aléatoires différentes mais elles ont la même loi. On attire l'attention sur le fait que la loi d'une variable aléatoire ne capture qu'un aspect partiel de cette variable. \square

Variables uniformes.

Définition II.3.8 \blacktriangleleft (Variables de loi uniforme sur un intervalle) Soit I , un intervalle borné. On note ℓ désigne la mesure de Lebesgue sur la tribus des boréliens de \mathbb{R} . Comme I est borné $\ell(I) < \infty$. On suppose que $\ell(I) > 0$. Soit $U : \Omega \rightarrow \mathbb{R}$, une v.a. \mathcal{F} -mesurable. Si la loi de U est donnée par

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbf{P}(U \in B) = \mu_U(B) = \frac{\ell(B \cap I)}{\ell(I)}.$$

alors on dit que U suit une loi uniforme sur I (et par abus de langage on dit souvent que U est « uniforme sur I »). \square

Reprenons les notations de la définition précédente. On note a l'extrémité gauche de I et b son extrémité droite si bien que I est soit $[a, b]$, soit $]a, b]$, soit $[a, b[$, soit $]a, b[$. Soit $g : \mathbb{R} \rightarrow \mathbb{R}_+$, une fonction mesurable. Le théorème de transfert II.3.6, page 52, montre que

$$\mathbf{E}[g(U)] = \frac{1}{\ell(I)} \int_I g(y) \ell(dy), \quad \text{ce qui se note simplement aussi :} \quad \frac{1}{b-a} \int_a^b g(y) dy.$$

On vérifie que la variable U n'est pas discrète : en effet soit $y \in I$, on a

$$\mathbf{P}(U=y) = \mathbf{P}(\{U \in \{y\}\}) = \mu_U(\{y\}) \frac{\ell(\{y\})}{\ell(I)} = 0,$$

car $\{y\}$ est un intervalle de longueur zéro et donc, par définition de la mesure de Lebesgue $\ell(\{y\}) = 0$. La loi de U ne peut donc pas être une somme pondérée de masses de Dirac, comme cela serait le cas si U était discrète. Attention! car on a toujours les faits suivants :

$$\Omega = \bigcup_{y \in I} \{U=y\} \quad \text{et pour tous } x, y \in I, \quad \{U=x\} \cap \{U=y\} = \emptyset.$$

Autrement dit les événements $\{U=y\}$, $y \in I$ forment une partition de Ω mais elle n'est pas dénombrable car I n'est pas un ensemble dénombrable, ce qui explique que

$$1 = \mathbf{P}(\Omega) \neq \sum_{y \in I} \mathbf{P}(U=y) = 0.$$

II.4 Inégalités usuelles.

Dans cette section, $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace de probabilité sur lequel sont définies toutes les variables considérées. Nous récapitulons brièvement quelques inégalités de la théorie de l'intégration dans le contexte, plus simple, des probabilités et nous en établissons de nouvelles.

II.4.a Inégalité de Jensen, moments.

Le but de cette section est de démontrer les inégalités de Jensen, Hölder et de Minkowski, ce qui se fait par des arguments de convexité. Rappelons qu'une fonction $\varphi : I \rightarrow \mathbb{R}$, définie sur un intervalle I de \mathbb{R} est dite *convexe* ssi

$$\forall x, y \in I, \forall \theta \in [0, 1] \quad \varphi(\theta x + (1 - \theta)y) \leq \theta\varphi(x) + (1 - \theta)\varphi(y) .$$

On rappelle également que si $\varphi : I \rightarrow \mathbb{R}$ est dérivable sur l'intérieur de I et de dérivée croissante, alors φ est convexe : les fonctions \exp et $-\log$ sont donc des fonctions convexes. On utilise le fait qu'une fonction convexe est l'enveloppe supérieure des droites passant sous son graphe. Plus précisément, on pose

$$S_\varphi = \{(a, b) \in \mathbb{R}^2 : \forall x \in I, ax + b \leq \varphi(x)\}$$

et on vérifie que

$$\forall x \in I, \quad \varphi(x) = \sup_{(a,b) \in S_\varphi} ax + b . \quad (\text{II.23})$$

On déduit de cette propriété l'inégalité de Jensen.

Lemme II.4.1 ◀▶ (**Inégalité de Jensen**) Soit I un intervalle de \mathbb{R} . Soit $\varphi : I \rightarrow \mathbb{R}$, une fonction convexe. Soit $X : \Omega \rightarrow I$, une variable réelle \mathcal{F} -mesurable telle que X et $\varphi(X)$ soient intégrables. Alors $\mathbf{E}[X] \in I$ et Alors, on a

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

Preuve : nous laissons le lecteur vérifier que $\mathbf{E}[X] \in I$. Soit S_φ comme dans (II.23). Si $(a, b) \in S_\varphi$, alors pour tout $\omega \in \Omega$, $aX(\omega) + b \leq \varphi(X(\omega))$ et en intégrant,

$$a\mathbf{E}[X] + b = \mathbf{E}[aX + b] \leq \mathbf{E}[\varphi(X)] .$$

En passant au supremum pour tout $(a, b) \in S_\varphi$, (II.23) permet de conclure. ■

Définition II.4.1 ◀▶ (**Moments d'une v.a.**) Soit X , une v.a. complexe \mathcal{F} -mesurable. Soit un réel $p > 0$. On dit que X admet un moment d'ordre p si $\mathbf{E}[|X|^p] < \infty$. Si $p \in \mathbb{N}^*$ et si X admet un moment d'ordre p , alors son moment d'ordre p est le complexe $\mathbf{E}[X^p]$. □

Comme $x \in \mathbb{R} \mapsto |x|^p$ est convexe si $p \geq 1$, l'inégalité de Jensen implique que

$$\forall p_1, p_2 \in]0, \infty[\text{ tels que } p_1 > p_2, \quad (\mathbf{E}[|X|^{p_2}])^{p_1/p_2} \leq \mathbf{E}[|X|^{p_1}] . \quad (\text{II.24})$$

Donc, si la v.a. X admet un moment d'ordre p_1 , alors elle admet un moment d'ordre p_2 pour tout $p_2 \in]0, p_1]$. Notamment,

si X admet un moment d'ordre 2, elle est intégrable.

et dans ce cas

$$(\mathbf{E}[X])^2 \leq (\mathbf{E}[|X|])^2 \leq \mathbf{E}[X^2] .$$

II.4.b Inégalités de Hölder.

On montre tout d'abord l'inégalité de convexité suivante.

Lemme II.4.2 Soit $b \in]0, 1[$. Pour tous $s, t \in]0, \infty[$, on a $s^b t^{1-b} \leq bs + (1-b)t$,

Preuve : comme $-\log$ est convexe, pour tous $s, t \in]0, \infty[$, on a

$$-\log(bs + (1-b)t) \leq -b \log s - (1-b) \log t,$$

qui entraîne aisément le résultat voulu. ■

Définition II.4.2 (*Exposants conjugués*) On dit que deux nombres réels p, q sont des exposants conjugués si $p, q > 1$ et $\frac{1}{p} + \frac{1}{q} = 1$. □

Proposition II.4.3 ◀▶ (**Hölder**) Soit $p > 1$ et soit q son exposant conjugué. Soient X, Y , deux v.a. positives \mathcal{F} -mesurables. Alors,

$$\mathbf{E}[XY] \leq (\mathbf{E}[X^p])^{\frac{1}{p}} (\mathbf{E}[Y^q])^{\frac{1}{q}}.$$

avec les conventions $0 \times \infty = 0$ et $\infty^c = \infty$, pour tout $c \in]0, \infty[$. De plus, si X et Y sont des v.a. complexes \mathcal{F} -mesurables, on applique ce qui précède à $|X|$ et $|Y|$ pour obtenir

$$\mathbf{E}[|XY|] \leq (\mathbf{E}[|X|^p])^{\frac{1}{p}} (\mathbf{E}[|Y|^q])^{\frac{1}{q}}.$$

Preuve : si $\mathbf{E}[X^p]$ ou $\mathbf{E}[Y^q]$ sont des quantités infinies ou nulles, l'inégalité est trivialement vérifiée. On suppose que ce n'est pas le cas et on pose

$$c = (\mathbf{E}[X^p])^{\frac{1}{p}} \quad \text{et} \quad d = (\mathbf{E}[Y^q])^{\frac{1}{q}}.$$

On suppose donc que $0 < c, d < \infty$. On fixe $\omega \in \Omega$. En appliquant le lemme II.4.2 à $s = (X(\omega)/c)^p$, $t = (Y(\omega)/d)^q$ et $b = \frac{1}{p}$, on obtient $(cd)^{-1} X(\omega)Y(\omega) \leq \frac{1}{p} c^{-p} X(\omega)^p + \frac{1}{q} d^{-q} Y(\omega)^q$. En intégrant cette inégalité, on obtient alors $(cd)^{-1} \mathbf{E}[XY] d\mu \leq \frac{1}{p} + \frac{1}{q} = 1$, ce qui entraîne le résultat désiré. ■

II.4.c Inégalités de Minkowski.

Proposition II.4.4 ◀▶ (**Minkowski**) Soit $p \geq 1$. Soient X, Y , deux v.a. positives \mathcal{F} -mesurables. Alors,

$$(\mathbf{E}[(X+Y)^p])^{\frac{1}{p}} \leq (\mathbf{E}[X^p])^{\frac{1}{p}} + (\mathbf{E}[Y^p])^{\frac{1}{p}}.$$

avec les conventions : $0 \times \infty = 0$ et $\infty^c = \infty$, pour tout $c \in]0, \infty[$. De plus, si X et Y sont des v.a. complexes \mathcal{F} -mesurables, on applique ce qui précède à $|X|$ et $|Y|$ pour obtenir

$$(\mathbf{E}[(|X| + |Y|)^p])^{\frac{1}{p}} \leq (\mathbf{E}[|X|^p])^{\frac{1}{p}} + (\mathbf{E}[|Y|^p])^{\frac{1}{p}}.$$

Preuve : on remarque que cette inégalité est une égalité lorsque que $p = 1$. On suppose que $p > 1$ et on note q son conjugué. Par Hölder, on a donc

$$\mathbf{E}[X(X+Y)^{p-1}] \leq (\mathbf{E}[X^p])^{\frac{1}{p}} (\mathbf{E}[(X+Y)^{q(p-1)}])^{\frac{1}{q}}$$

et

$$\mathbf{E}[Y(X+Y)^{p-1}] \leq (\mathbf{E}[Y^p])^{\frac{1}{p}} (\mathbf{E}[(X+Y)^{q(p-1)}])^{\frac{1}{q}}.$$

En additionnant ces deux inégalités et en remarquant que $p = q(p-1)$, on obtient

$$\mathbf{E}[(X+Y)^p] \leq \left((\mathbf{E}[X^p])^{\frac{1}{p}} + (\mathbf{E}[Y^q])^{\frac{1}{p}} \right) (\mathbf{E}[(X+Y)^p])^{\frac{1}{q}},$$

qui implique le résultat désiré si $\mathbf{E}[(X+Y)^p]$ est une quantité finie. Supposons que $\mathbf{E}[(X+Y)^p] = \infty$. Comme $x \mapsto x^p$ est convexe, on a $2^{-p}(X+Y)^p \leq 2^{-1}(X^p + Y^p)$ partout sur Ω ; en intégrant cette inégalité, on voit que l'une des deux espérances $\mathbf{E}[X^p]$ ou $\mathbf{E}[Y^p]$ doit être infinie : dans ce cas l'inégalité recherchée est trivialement vraie. ■

Pour $p = q = 1/2$, l'inégalité de Hölder est l'inégalité de *Cauchy-Schwarz* : si X et Y ont un moment d'ordre 2, alors XY est intégrable et

$$|\mathbf{E}[XY]| \leq \sqrt{\mathbf{E}[X^2]} \sqrt{\mathbf{E}[Y^2]}.$$

On en déduit l'inégalité suivante.

Lemme II.4.5 ◀▶ Soit $Z : \Omega \rightarrow \mathbb{R}_+$, une variable \mathcal{F} mesurable positive qui admet un moment d'ordre 2 et qui n'est pas p.s. nulle. Alors,

$$\mathbf{P}(Z > 0) \geq \frac{(\mathbf{E}[Z])^2}{\mathbf{E}[Z^2]}.$$

◀▶ **Preuve** : par Cauchy-Schwarz, on a

$$\mathbf{E}[Z] = \mathbf{E}[Z \mathbf{1}_{\{Z>0\}}] \leq \sqrt{\mathbf{E}[Z^2]} \sqrt{\mathbf{E}[(\mathbf{1}_{\{Z>0\}})^2]} = \sqrt{\mathbf{E}[Z^2]} \sqrt{\mathbf{P}(Z > 0)},$$

ce qui implique le résultat. ■

II.4.d Inégalités de type Markov, variance et inégalité de Tchebychev.

Lemme II.4.6 ◀▶ (**Inégalités de Markov**) Soit $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, une fonction croissante telle que $\phi(a) > 0$ dès que $a > 0$. Soit X , une v.a. positive et \mathcal{F} -mesurable. Alors,

$$\mathbf{P}(X \geq a) \leq \frac{1}{\phi(a)} \mathbf{E}[\phi(X)].$$

◀▶ **Preuve** : on remarque que $\mathbf{1}_{\{X \geq a\}} \leq \frac{\phi(X)}{\phi(a)}$. En prenant l'espérance, on a donc

$$\mathbf{P}(X \geq a) = \mathbf{E}[\mathbf{1}_{\{X \geq a\}}] \leq \phi(a)^{-1} \mathbf{E}[\phi(X)],$$

ce qui montre l'inégalité désirée. ■

Par exemple, pour toute v.a. positive X qui est \mathcal{F} -mesurable, on a

$$\mathbf{P}(X \geq a) \leq a^{-1} \mathbf{E}[X], \quad \mathbf{P}(X \geq a) \leq \frac{1 - \mathbf{E}[e^{-X}]}{1 - e^{-a}}, \quad \mathbf{P}(X \leq a) \leq e^a \mathbf{E}[e^{-X}].$$

Définition II.4.3 ◀▶ (Variance et écart-type d'une v.a.) Soit X , une v.a. réelle \mathcal{F} -mesurable. On suppose qu'elle admet un moment d'ordre deux. On définit sa variance $\text{var}(X)$ et son écart-type $\sigma(X)$ par

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

Donc $\sigma(X) = \sqrt{\text{var}(X)}$. □

Lemme II.4.7 ◀▶ (Identité de König) Soit X , une v.a. réelle \mathcal{F} -mesurable admettant un moment d'ordre 2, c'est-à-dire que $\mathbf{E}[X^2]$. On a

$$\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 . \quad (\text{II.25})$$

◀▶ **Preuve** : comme X admet un moment d'ordre 2, elle admet un moment d'ordre 1 c'est-à-dire qu'elle est intégrable et pour simplifier les notations on note $m = \mathbf{E}[X]$, qui est l'espérance de X . Par définition de la variance on a

$$\begin{aligned} \text{var}(X) &= \mathbf{E}[(X - m)^2] = \mathbf{E}[X^2 - 2mX + m^2] \\ &= \mathbf{E}[X^2] - 2m\mathbf{E}[X] + m^2 \\ &= \mathbf{E}[X^2] - 2m \cdot m + m^2 = \mathbf{E}[X^2] - m^2, \end{aligned}$$

qui est ce qu'il fallait prouver. ■

Le lemme suivant est élémentaire mais d'usage très fréquent.

Lemme II.4.8 ◀▶ Soit X , une v.a. réelle \mathcal{F} -mesurable admettant un moment d'ordre 2, c'est-à-dire que $\mathbf{E}[X^2]$. Soient $a, b \in \mathbb{R}$, deux constantes. Alors

$$\text{var}(aX + b) = a^2 \text{var}(X) .$$

◀▶ **Preuve** : on a d'abord $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$, puisque d'une part l'espérance est linéaire et que l'espérance d'une constante est la constante elle-même. Donc

$$\begin{aligned} \text{var}(aX + b) &= \mathbf{E}[(aX + b - \mathbf{E}[aX + b])^2] \\ &= \mathbf{E}[(aX - a\mathbf{E}[X])^2] \\ &= \mathbf{E}[a(X - \mathbf{E}[X])^2] \\ &= a^2 \mathbf{E}[(X - \mathbf{E}[X])^2] \\ &= a^2 \text{var}(X), \end{aligned}$$

ce qui est le résultat voulu. ■

Lemme II.4.9 ◀▶ (Inégalité de Tchebychev) Soit X , une v.a. réelle \mathcal{F} -mesurable admettant un moment d'ordre 2. Pour tout $a \in \mathbb{R}_+$, on a

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

◀▶ **Preuve** : on applique Markov à $\phi(a) = a^2$ et à la variable $Y = |X - \mathbf{E}[X]|$. ■

Exemple II.4.1 ◀▶ Un fabricant d'ampoules indique sur le paquet « *Durée de vie moyenne de 180 heures, écart-type de 20 heures* ». Si on note X la durée de vie d'une ampoule cela signifie donc que $\mathbf{E}[X] = 180$ heures et que $\sigma(X) = 20$ heures. L'inégalité de Tchebychev permet d'interpréter ces données : pour tout $a \in]0, \infty[$ (en heures)

$$\mathbf{P}(|X - 180| \geq a) \leq \frac{20^2}{a^2}.$$

Si on veut être sûr à au moins 96% de ses prédictions, on obtient cela en prenant $(20/a)^2 \leq 4\%$, c'est-à-dire que $a = 100$ heures convient et

$$\mathbf{P}(|X - 180| \geq 100) \leq \frac{20^2}{100^2} = \frac{1}{25} = 4\%$$

c'est-à-dire que

$$\mathbf{P}(X \in [80, 280]) = \mathbf{P}(|X - 180| \leq 100) \geq 100\% - 4\% = 96\%$$

C'est-à-dire qu'avec une probabilité au moins égale à 96% l'ampoule va durée entre 80 heures et 280 heures. Ce n'est pas précis mais cela n'est pas complètement sans information. \square

II.5 Fonctions génératrices de v.a. entières.

Fonctions génératrices. Ce paragraphe est consacré à la caractérisation des lois sur \mathbb{N} . Sauf mention du contraire, $(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace de probabilité de référence pour cette section.

Définition II.5.1 ◀▶ (*Fonction génératrice d'une v.a. entière positive*) Soit $X : \Omega \rightarrow \mathbb{N}$, une v.a. \mathcal{F} -mesurable. On pose

$$\forall r \in [0, 1], \quad \varphi_X(r) = \mathbf{E}[r^X] = \sum_{n \in \mathbb{N}} r^n \mathbf{P}(X = n).$$

La fonction $\varphi_X : [0, 1] \rightarrow \mathbb{R}_+$ est appelé la *fonction génératrice de X* , ou de sa loi. \square

Proposition II.5.1 Soit $X : \Omega \rightarrow \mathbb{N}$, une v.a. \mathcal{F} -mesurable. On note $\varphi_X : [0, 1] \rightarrow \mathbb{R}_+$ sa fonction génératrice.

(i) Pour tous $r \in [0, 1[$ et $p \in \mathbb{N}$, on a $\sum_{n > p} n(n-1) \dots (n-p+1) r^{n-p} \mathbf{P}(X = n) < \infty$. Cela implique que φ_X est infiniment dérivable sur $[0, 1[$ et que

$$\forall r \in [0, 1[\quad \forall p \in \mathbb{N}, \quad \varphi_X^{(p)}(r) = \sum_{n \geq p} n(n-1) \dots (n-p+1) r^{n-p} \mathbf{P}(X = n). \quad (\text{II.26})$$

En particulier, cela implique

$$\forall p \in \mathbb{N}, \quad \mathbf{P}(X = p) = \frac{1}{p!} \varphi_X^{(p)}(0). \quad (\text{II.27})$$

(ii) La loi d'une variable aléatoire à valeurs dans \mathbb{N} est caractérisée par sa fonction génératrice. Plus précisément, soit $(\Omega', \mathcal{F}', \mathbf{P}')$, un espace de probabilité et soit $Y : \Omega' \rightarrow \mathbb{N}$ une v.a. \mathcal{F}' -mesurable. Alors,

$$\left(\forall r \in [0, 1], \quad \mathbf{E}[r^X] = \mathbf{E}'[r^Y] \right) \iff \left(\forall p \in \mathbb{N}, \quad \mathbf{P}(X=p) = \mathbf{P}'(Y=p) \right).$$

(iii) X admet un moment d'ordre p si et seulement si φ_X admet une dérivée $p^{\text{ième}}$ à gauche en 1. Dans ce cas on a

$$\varphi_X^{(p)}(1-) \lim_{\substack{r \rightarrow 1 \\ r < 1}} = \mathbf{E}[X(X-1)\dots(X-p+1)].$$

Preuve : on fixe $0 < r_0 < r_1 < 1$ et $p \geq 1$. On observe que $\lim_{n \rightarrow \infty} n^p (r_0/r_1)^{n-p} = 0$. On pose alors $C_p = \max_{n \geq p} n^p (r_0/r_1)^{n-p}$ et $g_n(r) = r^n \mathbf{P}(X=n)$, et on observe alors que

$$\forall n \geq p, \forall r \in [0, r_0], \quad 0 \leq g_n^{(p)}(r) = n(n-1)\dots(n-p+1)r^{n-p} \mathbf{P}(X=n) \leq C_p r_1^{n-p}.$$

On en déduit que

$$\sum_{n > p} n(n-1)\dots(n-p+1)r^{n-p} \mathbf{P}(X=n) < C_p \sum_{n \geq p} r_1^{n-p} = \frac{C_p}{1-r_1} < \infty.$$

En utilisant le théorème de dérivation sur l'intégrale (mais pour les séries, c'est-à-dire sous la mesure de comptage), on en déduit que $\varphi_X(r) = \sum_{n \in \mathbb{N}} g_n(r)$ est p fois dérivable sur $[0, r_0]$ et que

$$\varphi_X^{(p)}(r) = \sum_{n \geq p} g_n^{(p)}(r),$$

ce qui est bien (II.26). Lorsque l'on évalue (II.26) en $r=0$, on obtient (II.27), ce qui termine la preuve de (i).

Le point (ii) découle du (i). En effet il est clair que si pour tout $p \in \mathbb{N}$, on a $\mathbf{P}(X=p) = \mathbf{P}'(Y=p)$, alors $\varphi_X = \varphi_Y$. Montrons l'implication réciproque en supposant que $\varphi_X = \varphi_Y$. On donc par le (i) que

$$\mathbf{P}(X=p) = \frac{1}{p!} \varphi_X^{(p)}(0) = \frac{1}{p!} \varphi_Y^{(p)}(0) = \mathbf{P}'(Y=p),$$

ce qui termine la preuve du (ii).

Le point (iii) découle de (II.26) : en effet on remarque que pour tout $n \geq p \geq 0$, la fonction $r \in [0, 1[\mapsto n(n-1)\dots(n-p+1)r^{n-p} \mathbf{P}(X=n)$ est croissante positive. Par le théorème de convergence monotone

$$\begin{aligned} \varphi_X^{(p)}(1-) &= \lim_{\substack{r \rightarrow 1 \\ r < 1}} \varphi_X^{(p)}(r-) \\ &= \lim_{\substack{r \rightarrow 1 \\ r < 1}} \sum_{n \geq p} n(n-1)\dots(n-p+1)r^{n-p} \mathbf{P}(X=n) \\ &= \sum_{n \geq p} \lim_{\substack{r \rightarrow 1 \\ r < 1}} n(n-1)\dots(n-p+1)r^{n-p} \mathbf{P}(X=n) \\ &= \sum_{n \geq p} n(n-1)\dots(n-p+1) \mathbf{P}(X=n) \end{aligned}$$

$$= \mathbf{E}[X(X-1)\dots(X-p+1)]. \quad (\text{II.28})$$

Cette égalité a lieu dans $[0, \infty]$. Comme $X(X-1)\dots(X-p+1) \leq X^p$, si X admet un moment d'ordre p , alors les inégalité précédentes impliquent que $\varphi_X^{(p)}(1-) < \infty$ et donc φ_X admet une dérivée $p^{\text{ième}}$ à gauche en 1.

Réciproquement, puisque la fonction $x \in \mathbb{R}_+ \mapsto x^p$ est convexe, on a pour tous $a, b \in \mathbb{R}_+$, $(\frac{1}{2}(a+b))^p \leq \frac{1}{2}(a^p + b^p)$. On suppose que $X \geq p-1$ et on applique cette inégalité à $a = X-p+1$ et $b = p-1$ pour obtenir

$$\mathbf{1}_{\{X > p\}} X^p \leq 2^{p-1}((X-p+1)^p + (p-1)^p) \leq 2^{p-1}X(X-1)\dots(X-p+1) + 2^{p-1}(p-1)^p.$$

Donc si $\mathbf{E}[X(X-1)\dots(X-p+1)] < \infty$ alors $\mathbf{E}[\mathbf{1}_{\{X > p\}} X^p] < \infty$. Or

$$\mathbf{E}[X^p] = \mathbf{E}[\mathbf{1}_{\{X > p\}} X^p] + \mathbf{E}[\mathbf{1}_{\{X \leq p\}} X^p] \leq \mathbf{E}[\mathbf{1}_{\{X > p\}} X^p] + p^p$$

donc on a montré que si $\mathbf{E}[X(X-1)\dots(X-p+1)] < \infty$ alors $\mathbf{E}[X^p] < \infty$. L'inégalité (II.28) entraîne alors que si φ_X admet une dérivée $p^{\text{ième}}$ à gauche en 1, alors X admet un moment d'ordre p , ce qui prouve (iii). ■

La fonction génératrice d'une v.a. X à valeurs dans \mathbb{N} s'étend au disque complexe unité fermé

$$\forall z \in \mathbb{C} \text{ tel que } |z| \leq 1, \quad \varphi_X(z) = \sum_{n \in \mathbb{N}} z^n \mathbf{P}(X = n).$$

On montre facilement la proposition suivante.

Proposition II.5.2 Soit $X : \Omega \rightarrow \mathbb{N}$, une v.a. \mathcal{F} -mesurable. On note φ_X sa fonction génératrice sur le disque unité $\{z \in \mathbb{C} : |z| \leq 1\}$. Pour tout $p \in \mathbb{N}$,

$$\mathbf{P}(X = p) = \frac{1}{p!} \varphi_X^{(p)}(0) = \frac{1}{2\pi} \int_0^{2\pi} e^{-ipu} \varphi_X(e^{iu}) du. \quad (\text{II.29})$$

Cela montre que la loi de X est caractérisée par les valeurs de φ_X sur le cercle unité $\{z \in \mathbb{C} : |z| = 1\}$.

Preuve : on observe que pour tout $p \in \mathbb{N}$, on a $e^{ipu} \varphi_X(e^{iu}) = \sum_{n \in \mathbb{N}} e^{iu(n-p)} \mathbf{P}(X = n)$. Comme $\sum_{n \in \mathbb{N}} |e^{iu(n-p)} \mathbf{P}(X = n)| = 1$ qui est une fonction Lebesgue-intégrable sur $[0, 2\pi]$, on peut intervertir somme et intégrale dans l'égalité suivante

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-ipu} \varphi_X(e^{iu}) du = \sum_{n \in \mathbb{N}} \frac{1}{2\pi} \int_0^{2\pi} du e^{-iu(n-p)} \mathbf{P}(X = n).$$

Or $\frac{1}{2\pi} \int_0^{2\pi} du e^{-iu(n-p)} = 0$ si $p \neq n$ et 1 si $n = p$, ce qui implique (II.29). ■

II.6 Densité d'une variable aléatoire.

II.6.a Rappel de théorie de l'intégration : mesures à densité.

On se place dans le cadre général de la théorie de l'intégration et on rappelle tout d'abord le lemme suivant, dont la preuve peut être omise à première lecture.

Lemme II.6.1 Soit (E, \mathcal{E}, μ) , un espace mesuré sigma fini. Soient $f, g : E \rightarrow [0, \infty]$, deux fonctions \mathcal{E} -mesurables. On suppose que

$$\forall A \in \mathcal{E}, \quad \int_A f \, d\mu \leq \int_A g \, d\mu.$$

Alors $f \leq g$, μ -presque partout.

Preuve : on suppose d'abord que μ est finie. Soient $a, b \in \mathbb{Q}_+$ tels que $a < b$. On observe d'abord que

$$\{g \leq a < b \leq f\} = g^{-1}([0, a]) \cap f^{-1}([b, \infty]) \in \mathcal{E}.$$

On a ensuite

$$g \mathbf{1}_{\{g \leq a < b \leq f\}} \leq a \mathbf{1}_{\{g \leq a < b \leq f\}} \leq b \mathbf{1}_{\{g \leq a < b \leq f\}} \leq f \mathbf{1}_{\{g \leq a < b \leq f\}}.$$

En intégrant cette inégalité, on obtient

$$\int_{\{g \leq a < b \leq f\}} g \, d\mu \leq a \mu(\{g \leq a < b \leq f\}) \leq b \mu(\{g \leq a < b \leq f\}) \leq \int_{\{g \leq a < b \leq f\}} f \, d\mu.$$

Mais, par hypothèse, les membres extrêmes de ces inégalités sont égaux, donc

$$a \mu(\{g \leq a < b \leq f\}) = b \mu(\{g \leq a < b \leq f\}).$$

Comme μ est de masse finie et que $a \neq b$, on obtient $\mu(\{g \leq a < b \leq f\}) = 0$. On voit ensuite que

$$\{g < f\} = \bigcup_{a, b \in \mathbb{Q}_+} \{g \leq a < b \leq f\} \in \mathcal{E},$$

et par sigma sous-additivité, $\mu(\{g < f\}) \leq \sum_{a, b \in \mathbb{Q}_+} \mu(\{g \leq a < b \leq f\}) = 0$, ce termine la preuve dans le cas où μ est finie.

Il reste à montrer ce résultat lorsque μ est sigma-finie : dans ce cas il existe une suite $E_n \in \mathcal{E}$, $n \in \mathbb{N}$, tels que $E_n \subset E_{n+1}$, $\mu(E_n) < \infty$ et $\bigcup_{n \in \mathbb{N}} E_n = E$. Pour tout $n \in \mathbb{N}$, on note $\mu_n = \mu(\cdot \cap E_n)$. En remplaçant A par $A \cap E_n$ dans l'hypothèse, on voit que pour tout $A \in \mathcal{E}$, $\int_A f \, d\mu_n \leq \int_A g \, d\mu_n$. Comme les mesures μ_n sont finies, on en déduit que $\mu(E_n \cap \{g < f\}) = 0$. Par la proposition I.1.7 (i) page 8, $\mu(\{g < f\}) = \lim_n \uparrow \mu(E_n \cap \{g < f\}) = 0$, ce qui montre le résultat voulu. ■

La proposition suivante définit ce que sont les *mesures admettant une densité par rapport à une autre mesure*.

Proposition II.6.2 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f : E \rightarrow [0, \infty]$, une fonction \mathcal{E} -mesurable.

(i) Pour tout $A \in \mathcal{E}$ on pose

$$\nu(A) = \int_A f \, d\mu = \int_E \mathbf{1}_A f \, d\mu.$$

Alors ν est une mesure positive sur (E, \mathcal{E}) appelée **mesure de densité f par rapport à μ** , ce que l'on abrège en $\nu = f\mu$.

(ii) si $g : E \rightarrow [0, \infty]$ est \mathcal{E} -mesurable, alors pour tout $A \in \mathcal{E}$, on a

$$\int_A g d\nu = \int_A gf d\mu . \quad (\text{II.30})$$

Cette formule porte le nom de **formule de la densité**.

◀► **Preuve** : on prouve d'abord (i). Il est clair que $\nu(\emptyset) = 0$. Soit $A_n \in \mathcal{E}$, $n \in \mathbb{N}$, une suite d'ensembles mesurables deux-à-deux disjoints. Pour simplifier les notations, on pose $A = \bigcup A_n$. Comme les ensembles sont deux-à-deux disjoints on a $\mathbf{1}_A = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}$. Par la proposition II.2.5 d'interversion série/intégrale positive, on a

$$\nu(A) = \int_E f \mathbf{1}_A d\mu = \int_E \left(\sum_{n \in \mathbb{N}} f \mathbf{1}_{A_n} \right) d\mu = \sum_{n \in \mathbb{N}} \int_E f \mathbf{1}_{A_n} d\mu = \sum_{n \in \mathbb{N}} \nu(A_n),$$

ce qui montre bien que ν est sigma-additive et donc que ν est une mesure.

Montrons (ii) : la définition de ν entraîne facilement que (II.30) est vérifiée pour toute fonction g qui est une fonction indicatrice. Soit $g : E \rightarrow [0, \infty]$ qui est \mathcal{E} -mesurable. Par le lemme II.1.8, il existe $c_n \in \mathbb{R}_+$ et $B_n \in \mathcal{E}$, $n \in \mathbb{N}$, tels que $g = \sum_{n \in \mathbb{N}} c_n \mathbf{1}_{B_n}$. Par la proposition II.2.5 d'interversion série/intégrale positive appliquée deux fois, on a

$$\int_A g d\nu = \sum_{n \in \mathbb{N}} c_n \nu(A \cap B_n) = \sum_{n \in \mathbb{N}} c_n \int_A \mathbf{1}_{B_n} f d\mu = \int_A gf d\mu ,$$

ce qui termine la preuve de la proposition. ■

Remarque II.6.1 Il est clair que si ν et ν' admettent la même densité par rapport à μ , alors ces mesures sont égales. □

Le lemme II.6.1 implique immédiatement le corollaire suivant montrant l'unicité presque partout des densités.

Corollaire II.6.3 ▶ Soit (E, \mathcal{E}, μ) , un espace mesuré sigma-fini. Soient $f, g : E \rightarrow [0, \infty]$, deux fonctions \mathcal{E} -mesurables. Si $f\mu = g\mu$, alors $f = g$, μ -presque partout.

Autrement dit si la mesure ν admet la densité f par rapport à μ , l'unicité μ -presque sûre permet de parler de « la » densité. On utilise parfois la notation (symbolique)

$$f = \frac{d\nu}{d\mu} .$$

La mesure ν est une mesure de probabilité si et seulement si

$$\nu(E) = \int_E f d\mu = 1 .$$

On remarque en si ν admet la densité f par rapport à μ , alors

$$\forall A \in \mathcal{E}, \quad (\mu(A) = 0) \implies (\nu(A) = 0) . \quad (\text{II.31})$$

Il s'agit du point (iii) de la proposition II.2.2, page 38. Nous énonçons ci-dessous une version partielle du théorème de Radon-Nikodym, qui montre que réciproquement si deux mesures μ et ν sont telles que (II.31) soit vérifiée, alors ν admet une densité par rapport à μ .

Théorème II.6.4 (Radon-Nikodym : cas des mesures complexes) Soit (E, \mathcal{E}) , un espace mesurable. Soit μ , une mesure positive sigma finie sur \mathcal{E} . Soit ν , une mesure qui est une somme de mesures finies. On suppose que

$$\forall A \in \mathcal{E}, \quad (\mu(A) = 0) \implies (\nu(A) = 0) . \quad (\text{II.32})$$

Alors, ν admet une densité $f : E \rightarrow [0, \infty]$ par rapport à μ . Cette densité f est unique μ -p.p. De plus, si ν est sigma-finie, alors f peut être prise à valeurs dans \mathbb{R}_+ .

II.6.b Variables dont les lois admettent une densité.

Dans cette section, $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace de probabilité sur lequel sont définies toutes les variables considérées.

Définition II.6.1 \blacktriangleleft (Variable réelle à densité) Soit $X : \Omega \rightarrow \mathbb{R}$, une v.a. \mathcal{F} -mesurable réelle. On note ℓ la mesure de Lebesgue sur \mathbb{R} . On dit que X admet comme densité la fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}_+$ si la loi μ_X de X sous \mathbf{P} admet la densité f par rapport à la mesure de Lebesgue, c'est-à-dire si

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \mathbf{P}(X \in A) = \mu_X(A) = \int_A f(y) \ell(dy) .$$

On remarque que dire « X admet f pour densité » est abusif car cette assertion concerne la loi de X et non X elle-même mais nous continuerons à le dire ainsi et même à l'écrire, par commodité. \square

Remarque II.6.2 Il est clair que si deux variables réelles X et Y admettent la même densité alors X et Y ont la même loi (voir la remarque II.6.1, page 62). Autrement la densité (éventuelle) d'une variable aléatoire réelle caractérise sa loi. \square

La proposition II.30, page 62 et la formule de la densité au (ii) se réécrit alors de la manière suivante.

Proposition II.6.5 \blacktriangleleft (Formule de la densité : cas réel) Soit $X : \Omega \rightarrow \mathbb{R}$, une v.a. \mathcal{F} -mesurable réelle. On note ℓ la mesure de Lebesgue sur \mathbb{R} . On suppose que X admet la densité f . Alors pour toute fonction mesurable $g : \mathbb{R} \rightarrow [0, \infty]$, on a

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(y) f(y) \ell(dy) . \quad (\text{II.33})$$

Supposons maintenant que $g : \mathbb{R} \rightarrow \mathbb{R}$ soit mesurable. Alors, $g(X)$ est intégrable ssi gf est ℓ -intégrable (c'est-à-dire ssi $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$) Dans ce cas on a encore la formule (II.33).

\blacktriangleleft **Preuve :** par le théorème de transfert II.3.6, page 52, on a

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(z) \mu_X(dz) ,$$

où μ_X est la loi de X sous \mathbf{P} . Comme X admet pour densité f , signifie que la loi μ_X de X admet f comme densité par rapport à la mesure de Lebesgue sur \mathbb{R} , qui est notée on a $\mu_X = f \cdot \ell$ et la formule de la densité au point (ii) de la proposition II.30, page 62 appliquée à $\nu = \mu_X$ et $\mu = \ell$ implique que

$$\int_{\mathbb{R}} g(z) \mu_X(dz) = \int_{\mathbb{R}} g(y) f(y) \ell(dy) ,$$

ce qui complète la preuve de la proposition. ■

Soit X une variable réelle \mathcal{F} mesurable. On suppose d'une part qu'elle admet une densité f et on suppose qu'elle est intégrable sous \mathbf{P} (c'est-à-dire que $\mathbf{E}[|X|] < \infty$). Alors, la formule de la densité (II.33) appliquée à $g(y) = y$ donne

$$\mathbf{E}[X] = \int_{\mathbb{R}} y f(y) \ell(dy).$$

De même si X admet un moment d'ordre 2, alors

$$\mathbf{E}[X^2] = \int_{\mathbb{R}} y^2 f(y) \ell(dy).$$

Donnons deux exemples importants de variables réelles à densité : les variables exponentielles et les variables gaussiennes (aussi appelées variables normales).

Définition II.6.2 ◀▶ (*Loi exponentielle*) Une v.a. $X : \Omega \rightarrow \mathbb{R}_+$ est (de loi) exponentielle de paramètre $c \in]0, \infty[$ si sa loi admet la densité $t \in \mathbb{R} \mapsto \mathbf{1}_{\mathbb{R}_+}(t) c e^{-ct}$ par rapport à la mesure de Lebesgue. □

Si X est une exponentielle de paramètre c , alors pour toute fonction mesurable $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$,

$$\mathbf{E}[g(X)] = \int_0^\infty g(t) c e^{-ct} dt.$$

Par exemple, il est facile de vérifier pour tout $\lambda \in \mathbb{R}_+$ que

$$\mathbf{E}[e^{-\lambda X}] = \frac{c}{c + \lambda}. \quad (\text{II.34})$$

Un simple calcul intégral (avec tout de même une intégration par parties) implique les résultats suivants :

$$\mathbf{E}[X] = \int_0^\infty t c e^{-ct} dt = \frac{1}{c}, \quad \mathbf{E}[X^2] = \int_0^\infty t^2 c e^{-ct} dt = \frac{2}{c^2} \quad \text{et donc} \quad \text{var}(X) = \frac{1}{c^2}.$$

Définition II.6.3 ◀▶ (*Lois Gaussiennes (ou normales) en dimension 1*) Soient $m \in \mathbb{R}$ et $\sigma \in]0, \infty[$. Soit $X : \Omega \rightarrow \mathbb{R}$, une variable \mathcal{F} -mesurable. Cette variable est dite (de loi) Gaussienne, ou encore normale, de moyenne m et de variance σ^2 si sa loi admet la densité $g_{m,\sigma}$ donnée par

$$g_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (\text{II.35})$$

- On note cela $X \sim \mathcal{N}(m, \sigma^2)$.
- Si $m = 0$, X est dite *centrée* et une variable Gaussienne $\mathcal{N}(0, 1)$ est appelée *gaussienne standard* (ou loi *normale standard*)
- Par commodité, on dit également que X est une loi Gaussienne de moyenne m et de variance nulle si X est presque sûrement égal à m . On continue de noter cela $X \sim \mathcal{N}(m, 0)$. On parle alors de variable *Gaussienne dégénérée*. □

On vérifie par une intégration par parties que si $X \sim \mathcal{N}(m, \sigma^2)$,

$$\mathbf{E}[X] = \int_{\mathbb{R}} x g_{m,\sigma}(x) dx = m \quad \text{et} \quad \mathbf{var}(X) = \int_{\mathbb{R}} (x - m)^2 g_{m,\sigma}(x) dx = \sigma^2 .$$

Autrement dit la loi de X est entièrement déterminée par sa moyenne et sa variance. On montre le lemme suivant qui permet d'obtenir une variable gaussienne standard à partir d'une variable gaussienne générale.

Lemme II.6.6 *Soit $m \in \mathbb{R}$ et $\sigma \in]0, \infty[$. Soit $X \sim \mathcal{N}(m, \sigma^2)$. On pose $Y := \frac{1}{\sigma}(X - m)$. Alors Y est une variable gaussienne standard.*

Preuve : soit $f : \mathbb{R} \rightarrow \mathbb{R}$, une fonction mesurable bornée. La formule de la densité (II.33) combinée à un changement de variable linéaire implique

$$\begin{aligned} \mathbf{E}[f(Y)] &= \mathbf{E}\left[f\left(\frac{1}{\sigma}(X - m)\right)\right] = \int_{-\infty}^{\infty} g_{m,\sigma}(x) f\left(\frac{1}{\sigma}(x - m)\right) dx \\ &= \int_{-\infty}^{\infty} g_{m,\sigma}(y + m) f\left(\frac{1}{\sigma}y\right) dy = \int_{-\infty}^{\infty} \sigma g_{m,\sigma}(\sigma z + m) f(z) dz \\ &= \int_{-\infty}^{\infty} g_{0,1}(z) f(z) dz \end{aligned}$$

car $\sigma g_{m,\sigma}(\sigma z + m) = g_{0,1}(z)$. Donc Y admet la densité $g_{0,1}$, ce qui prouve le lemme car la densité d'une variable réelle caractérise sa loi. ■

Donnons quelques propriétés des v.a. réelles à densité.

• Soit $f : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction mesurable. Elle est la densité d'une variable aléatoire réelle ssi $\int_{\mathbb{R}} f(x) dx = 1$.

• Soit X une variable réelle admettant la densité f . On note ℓ la mesure de Lebesgue sur \mathbb{R} . Alors pour tout $y \in \mathbb{R}$,

$$\mathbf{P}(X = y) = \mu_X(\{y\}) = \int_{\{y\}} f d\ell = 0 ,$$

car $\ell(\{y\}) = 0$ (le singleton $\{y\}$ est en effet un intervalle de longueur 0) et par la propriété (iii) de la proposition II.2.2, page 38.

Plus généralement si $A \in \mathcal{B}(\mathbb{R})$ alors

$$\ell(A) = 0 \quad \implies \quad \mathbf{P}(X \in A) = 0 . \tag{II.36}$$

Définition II.6.4 ◀▶ (Vecteur aléatoire à densité) Soit $X : \Omega \rightarrow \mathbb{R}^d$, un vecteur aléatoire \mathcal{F} -mesurable. On note ℓ_d la mesure de Lebesgue sur \mathbb{R}^d . On dit que X admet comme densité la fonction mesurable $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ si la loi μ_X de X sous \mathbf{P} admet la densité f par rapport à la mesure de Lebesgue ℓ_d sur \mathbb{R}^d , c'est-à-dire si pour tout borélien $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mathbf{P}(X \in A) = \mu_X(A) = \int_A f(y) \ell_d(dy) = \int_{\mathbb{R}} dy_1 \dots \int_{\mathbb{R}} dy_d \mathbf{1}_A(y_1, \dots, y_d) f(y_1, \dots, y_d) .$$

Autrement dit, $\mu_X = f \cdot \ell_d$. □

En raisonnant comme dans la preuve de la proposition II.6.5, page 63, on obtient la formule de la densité pour les vecteurs aléatoires.

Proposition II.6.7 ◀▶ (Formule de la densité) Soit $X : \Omega \rightarrow \mathbb{R}^d$, une v.a. \mathcal{F} -mesurable. On note ℓ_d la mesure de Lebesgue sur \mathbb{R}^d . On suppose que X admet la densité f . Alors pour toute fonction mesurable $g : \mathbb{R}^d \rightarrow \mathbb{R}_+$, on a

$$\mathbf{E}[g(X)] = \int_{\mathbb{R}^d} g(y)f(y) \ell_d(dy) = \int_{\mathbb{R}} dy_1 \dots \int_{\mathbb{R}} dy_d g(y_1, \dots, y_d) f(y_1, \dots, y_d). \quad (\text{II.37})$$

Supposons maintenant que $g : \mathbb{R}^d \rightarrow \mathbb{R}$ soit mesurable. Alors, $g(X)$ est intégrable ssi gf est ℓ_d -intégrable et dans ce cas on a encore (II.37).

II.7 Fonctions de répartition.

Avant de discuter de la fonction de répartition d'une mesure, il convient tout d'abord de rappeler quelques résultats concernant les atomes des mesures en général et de rappeler également quelques résultats d'analyse élémentaire sur la continuité à droite et à gauche des fonctions d'une variable réelle.

II.7.a Rappel sur les atomes d'une mesure.

Définition II.7.1 ◀▶ (Atome d'une mesure, mesures diffuses) Soit (E, \mathcal{E}, μ) , un espace mesuré. On suppose que pour tout $x \in E$, on $\{x\} \in \mathcal{E}$. Un point $x \in E$ est un *atome* de μ si $\mu(\{x\}) > 0$. La quantité $\mu(\{x\})$ est la masse de l'atome x . On commence tout d'abord par prouver le résultat suivant. Une mesure μ qui n'a pas d'atome est dite *diffuse*. ◻

Exemple II.7.1 Soit $E = \mathbb{R}$, soit $\mathcal{E} = \mathcal{P}(\mathbb{R})$ qui est la classe de tous les sous-ensembles de \mathbb{R} , et soit :

$$\mu = \frac{1}{3}\delta_0 + \frac{1}{8}\delta_7 + \frac{1}{12}\delta_{37}.$$

Alors,

$$\mu(\{0\}) = \frac{1}{3}, \quad \mu(\{7\}) = \frac{1}{8} \quad \text{et} \quad \mu(\{37\}) = \frac{1}{12}.$$

La mesure μ admet donc 3 atomes. ◻

Exemple II.7.2 Soit $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$, et on considère la mesure de Lebesgue ℓ . On remarque que pour tout $x \in \mathbb{R}$ le singleton $\{x\}$ est un fermé : il est donc dans la tribu des Boréliens $\mathcal{B}(\mathbb{R})$, il est le complémentaire d'un ouvert. D'autre part $\{x\}$ est un cas particulier d'intervalle de longueur nulle : $\ell(\{x\}) = \ell([x, x]) = x - x = 0$. On en déduit que ℓ , la mesure de Lebesgue sur \mathbb{R} , est diffuse. ◻

Exemple II.7.3 Soit $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$; on considère la mesure $\mu = \pi^2\ell + \frac{1}{3}\delta_0 + \frac{1}{8}\delta_7$. Cette mesure n'est pas diffuse car elle a deux atomes en 0 et 7. ◻

Proposition II.7.1 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré. On suppose que pour tout $x \in E$, $\{x\} \in \mathcal{E}$. On suppose également que $\mu(E) < \infty$. Alors,

$$A_\mu := \{x \in E : \mu(\{x\}) > 0\} \text{ est dénombrable.}$$

Preuve : on pose $B_n = \{x \in E : \mu(\{x\}) \geq 2^{-n}\}$. Si $x_1, \dots, x_p \in B_n$ sont distincts, alors

$$p 2^{-n} \leq \mu(\{x_1\}) + \dots + \mu(\{x_p\}) = \mu(\{x_1, \dots, x_p\}) \leq \mu(E).$$

On voit donc que B_n est un ensemble fini qui a moins de $2^n \mu(E)$ éléments. Par ailleurs, on remarque que $A_\mu = \bigcup_{n \in \mathbb{N}} B_n$, donc A_μ est dénombrable. ■

Définition II.7.2 ◀▶ (Mesures purement atomiques) Soit (E, \mathcal{E}, μ) , un espace mesuré. On suppose \mathcal{E} contient les singletons. On note

$$A_\mu = \{x \in E : \mu(\{x\}) > 0\}$$

l'ensemble de ses atomes. Alors, μ est dite *purement atomique* s'il existe $N \in \mathcal{E}$ tel que $E \setminus A_\mu \subset N$ et $\mu(N) = 0$. □

On voit facilement que toute somme de masses de Dirac est une mesure purement atomique.

Proposition II.7.2 ◀▶ Soit (E, \mathcal{E}, μ) , un espace mesuré tel que \mathcal{E} contienne les singletons et tel que $\mu(E) < \infty$. On note $A_\mu = \{x \in E : \mu(\{x\}) > 0\}$ l'ensemble de ses atomes et on pose

$$\mu^{\text{di}} = \mu(\cdot \cap (E \setminus A_\mu)) \quad \text{et} \quad \mu^{\text{at}} = \sum_{x \in A_\mu} \mu(\{x\}) \delta_x.$$

Alors, $\mu = \mu^{\text{di}} + \mu^{\text{at}}$, μ^{di} est diffuse, μ^{at} est purement atomique. Si de plus $\mu = \nu_1 + \nu_2$ avec ν_1 diffuse et ν_2 purement atomique, alors $\nu_1 = \mu^{\text{di}}$ et $\nu_2 = \mu^{\text{at}}$.

Preuve : il est clair que $\mu = \mu^{\text{di}} + \mu^{\text{at}}$, μ^{di} est diffuse, μ^{at} est purement atomique. Il est clair que $\mu(\{x\}) = \nu_2(\{x\})$, pour tout $x \in E$. Donc $A_\mu = A_{\nu_2}$ qui est donc dénombrable et dans \mathcal{E} . Comme ν_2 est purement atomique, pour tout $B \in \mathcal{E}$, on a

$$\nu_2(B) = \nu_2(B \cap A_\mu) = \sum_{x \in B \cap A_\mu} \mu(\{x\}) = \mu^{\text{at}}(B).$$

Donc $\nu_2 = \mu^{\text{at}}$. On constate aussi que pour tout $B \in \mathcal{E}$, $\mu(B \cap (E \setminus A_\mu)) = \nu_1(B \cap (E \setminus A_\mu)) = \nu_1(B)$ car A_μ est dénombrable et ν_1 diffuse. ■

Exemple II.7.4 Comme à l'exemple II.7.3, page 66, on considère $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}(\mathbb{R})$ et la mesure $\mu = \pi^2 \ell + \frac{1}{3} \delta_0 + \frac{1}{8} \delta_7$. Cette mesure n'est pas diffuse car elle a deux atomes en 0 et 7 et elle n'est pas purement atomique parce que par exemple $\mu([1, 2]) = \pi^2 > 0$. On vérifie que $\pi^2 \ell = \mu^{\text{di}}$ et $\frac{1}{3} \delta_0 + \frac{1}{8} \delta_7 = \mu^{\text{at}}$. □

II.7.b Continuité, limites à droite et à gauche.

On rappelle la définition suivante.

Définition II.7.3 Soit $g = \mathbb{R} \rightarrow \mathbb{R}$.

- On dit que g admet $y \in \mathbb{R}$ comme limite à droite (resp. à gauche) en $x \in \mathbb{R}$ si pour toute suite $(u_n)_{n \in \mathbb{N}}$ qui décroît strictement (resp. croît strictement) vers x , on a $\lim_{n \rightarrow \infty} g(u_n) = y$. On utilise la notation $g(x+)$ (resp. $g(x-)$) pour désigner la limite à droite y (resp. à gauche) lorsqu'elle existe. \square
- On dit que g est continue à droite (resp. à gauche) en x si $g(x+) = g(x)$ (resp. $g(x-) = g(x)$). On abrège cela par « **càd** » pour *continue à droite* et par « **càg** » pour *continue à gauche*.
- Une fonction est dite continue à droite, ou **càd**, ssi elle est continue à droite en tout point de \mathbb{R} . Elle est dite continue à gauche, ou **càg**, si elle est continue à gauche en tout point de \mathbb{R} .
- Une fonction g est dite *continue à droite avec des limites à gauche*, ce que l'on abrège en « **càdlàg** », si pour tout $x \in \mathbb{R}$, $g(x+) = g(x)$ et si $g(x-)$ existe. \square

On rappelle le fait suivant :

g est continue en x ssi $g(x+)$, $g(x-)$ existent et sont telles que $g(x+) = g(x-) = g(x)$.

Exemple II.7.5 Il existe des fonctions qui n'ont de limite à droite ou à gauche. Par exemple, on peut considérer $g : \mathbb{R} \rightarrow \mathbb{R}$ définie comme suit : si $x \notin \mathbb{Q}$, on pose $g(x) = 0$; soit $x \in \mathbb{Q}$: alors il existe une unique paire d'entiers (p, q) tels que p est un entier relatif, q un entier strictement positif, p et q n'ont aucun diviseur commun et $x = p/q$; si p est pair on pose $g(x) = 1$ et si p est impair on pose $g(x) = -1$. On vérifie que pour tous $a, b \in \mathbb{R}$ tels que $a < b$, la fonction g sur $]a, b[$ prend une infinité de fois chacune des trois valeurs 0, -1 et 1. Cela implique facilement qu'en tout point g n'admet ni limite à gauche ni limite à droite. \square

Nous allons concentrer notre attention sur le cas des fonctions croissantes et montrer les propriétés élémentaires suivantes.

Proposition II.7.3 \blacktriangleleft Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante c'est-à-dire que pour tous $x, y \in \mathbb{R}$,

$$x \leq y \implies g(x) \leq g(y).$$

Alors les propriétés suivantes sont vérifiées.

\blacktriangleleft (i) En tout point $x \in \mathbb{R}$, $g(x+)$ et $g(x-)$ existent et on a

$$g(x+) = \inf_{]x, \infty[} g(y) \quad \text{et} \quad g(x-) = \sup_{]-\infty, x[} g(y). \quad (\text{II.38})$$

On en déduit que

$$g(x-) \leq g(x) \leq g(x+). \quad (\text{II.39})$$

(ii) La fonction g est **càdlàg** ssi elle est **càd**, c'est-à-dire ssi pour tout $x \in \mathbb{R}$, $g(x+) = g(x)$.

(iii) Pour tout $x \in \mathbb{R}$, on pose $\Delta g(x) = g(x+) - g(x-)$. On remarque que $\Delta g(x) \geq 0$ et la fonction g est continue en x ssi $\Delta g(x) = 0$.

(iv) On note $D := \{x \in \mathbb{R} : \Delta g(x) > 0\}$ qui est l'ensemble des points de discontinuité de g . Alors D est dénombrable (c'est-à-dire, soit vide, soit fini, soit en bijection avec \mathbb{N}).

Preuve : on montre la première égalité de (II.38). Pour cela on pose $a = \inf_{]x, \infty[} g(y)$. On se donne aussi une suite $(u_n)_{n \in \mathbb{N}}$ qui décroît strictement vers x . On a donc pour tout $n \in \mathbb{N}$, $u_n > x$ et donc $g(u_n) \geq a$. Comme $u_n > u_{n+1}$, on a $g(u_n) \geq g(u_{n+1})$: la suite $(g(u_n))_{n \in \mathbb{N}}$ est donc décroissante et minorée par a ; elle doit donc converger vers une limite a' et on a $a' \geq a$. Soit $y \in]x, \infty[$. Alors pour tout n assez grand on a $u_n < y$ et donc on a pour tout n assez grand, $g(u_n) \leq g(y)$. En passant à la limite dans cette inégalité, on voit que $a' \leq g(y)$, pour tout $y \in]x, \infty[$. Donc, en passant à l'infimum en y dans cette inégalité on a $a' \leq \inf_{]x, \infty[} g(y) = a$. On a donc $a = a'$. On a donc montré que pour toute suite $(u_n)_{n \in \mathbb{N}}$ qui décroît strictement vers x , $\lim_{n \rightarrow \infty} g(u_n)$ existe et vaut $\inf_{]x, \infty[} g(y)$, ce qui prouve la première égalité de (II.38) par définition de la limite à droite de g en x .

On raisonne de la même manière pour la limite à gauche dans (II.38) : les détails sont laissés au lecteur. On remarque ensuite que pour tout $y \in]x, \infty[$, on a $g(x) \leq g(y)$. En passant à l'infimum en y dans cette inégalité on a $g(x) \leq \inf_{]x, \infty[} g(y) = g(x+)$. On raisonne de même pour montrer que $g(x-) \leq g(x)$, ce qui complète la preuve de (II.39).

Les points (ii) et (iii) sont des conséquences directes de (II.39) : nous laissons les détails au lecteur. Montrons (iv) : on remarque que les intervalles ouverts $]g(x-), g(x-)[$, $x \in D$ sont disjoints deux-à-deux. Pour chaque $x \in D$, on peut trouver un rationnel $q(x) \in]g(x-), g(x+)[$. Alors $x \in D \mapsto q(x) \in \mathbb{Q}$ est une injection. Comme \mathbb{Q} est dénombrable, D l'est également. ■

II.7.c Fonctions de répartition.

Dans cette section, $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace de probabilité sur lequel sont définies toutes les variables considérées. Un moyen efficace de coder la loi d'une variable aléatoire réelle consiste à introduire les fonctions dites de répartition qui se définissent comme suit.

Définition II.7.4 ◀▶ (Fonction de répartition d'une mesure de probabilité réelle) Soit $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$, une mesure de probabilité sur les boréliens de \mathbb{R} . Sa fonction de répartition $F_\mu : \mathbb{R} \rightarrow [0, 1]$ est donnée par

$$\forall y \in \mathbb{R}, \quad F_\mu(y) = \mu(]-\infty, y]) .$$

On définit également la fonction de répartition d'une v.a. réelle comme suit : Soit X , une v.a. réelle \mathcal{F} -mesurable. Sa fonction de répartition $F_X : \mathbb{R} \rightarrow [0, 1]$ est donnée par

$$\forall y \in \mathbb{R}, \quad F_X(y) = \mathbf{P}(X \leq y) .$$

En fait, la fonction de répartition d'une variable réelle est la fonction de répartition de sa loi.

$$F_X(y) = \mathbf{P}(X \leq y) = \mathbf{P}(X \in]-\infty, y]) = \mu_X(]-\infty, y]), \quad y \in \mathbb{R},$$

par définition de la loi μ_X de X . □

La proposition suivante détaille les propriétés des fonctions de répartition.

Proposition II.7.4 ◀▶ Soit X , une v.a. réelle \mathcal{F} -mesurable. Les propriétés suivantes sont vérifiées.

(i) F_X est continue à droite, croissante et

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

(ii) Pour tout $x \in \mathbb{R}$, on a

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

Donc la loi de X est diffuse ssi sa fonction de répartition est continue.

◀► **Preuve :** on fixe $x \in \mathbb{R}$. Les événements $\{X \leq x + 2^{-n}\}$ sont décroissants pour l'inclusion en n et leur intersection est $\{X \leq x\}$. La décroissance séquentielle pour \mathbf{P} implique que

$$\lim_{n \rightarrow \infty} \mathbf{P}(X \leq x + 2^{-n}) = \mathbf{P}(X \leq x) .$$

Comme l'application $x \mapsto \mathbf{P}(X \leq x)$ est croissante la limite précédente est la limite à droite de la fonction de répartition de X , ce qui montre que F_X est continue à droite.

On remarque ensuite que les événements $\{X \leq -n\}$, $n \in \mathbb{N}$, sont décroissants en n et d'intersection vide. La décroissance séquentielle pour \mathbf{P} implique que $\lim_n \mathbf{P}(X \leq -n) = 0$. Comme l'application $x \mapsto \mathbf{P}(X \leq x)$ est croissante, la limite précédente est aussi $\lim_{x \rightarrow -\infty} F_X(x)$. Enfin, on observe que les événements $\{X \leq n\}$, $n \in \mathbb{N}$, sont croissants en n et leur réunion est Ω . La croissance séquentielle pour \mathbf{P} implique que $\lim_n \mathbf{P}(X \leq n) = 1$. Comme l'application $x \mapsto \mathbf{P}(X \leq x)$ est croissante, la limite précédente est aussi $\lim_{x \rightarrow \infty} F_X(x)$. cela termine la preuve du (i).

Montrons (ii) : on a

$$F_X(x) - F_X(x-) = \lim_{n \rightarrow \infty} F_X(x) - F_X(x - 2^{-n}) .$$

Or $F_X(x) - F(x - 2^{-n}) = \mathbf{P}(x - 2^{-n} < X \leq x)$. Les événements $\{x - 2^{-n} < X \leq x\}$ sont décroissants pour l'inclusion en n et leur intersection est $\{X = x\}$. La décroissance séquentielle pour \mathbf{P} implique alors $\lim_n \mathbf{P}(x - 2^{-n} < X \leq x) = \mathbf{P}(X = x)$, ce qui montre (ii). ■

La proposition suivante, qui est une conséquence de résultats antérieurs, montre que la loi d'une v.a. réelle est caractérisée par sa fonction de répartition.

Proposition II.7.5 ▶ Soient $(\Omega, \mathcal{F}, \mathbf{P})$ et $(\Omega', \mathcal{F}', \mathbf{P}')$, des espaces de probabilité. Soit $X : \Omega \rightarrow \mathbb{R}$, une v.a. \mathcal{F} -mesurable et soit $Y : \Omega' \rightarrow \mathbb{R}$, une v.a. \mathcal{F}' -mesurable. Alors X et Y ont même loi ssi elles ont même fonction de répartition, c'est-à-dire

$$F_X = F_Y \iff \mu_X = \mu_Y .$$

▶ **Preuve :** si $\mu_X = \mu_Y$, alors $F_X(x) = \mu_X(]-\infty, x]) = \mu_Y(]-\infty, x]) = F_Y(x)$, pour tout $x \in \mathbb{R}$. Réciproquement si $F_X = F_Y$, alors μ_X et μ_Y coïncident sur le pi-système

$$\mathcal{P} = \{\mathbb{R}\} \cup \{]-\infty, x]; x \in \mathbb{R}\} .$$

Comme $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{P})$ (voir le lemme I.1.2, page 4), le théorème I.1.12 d'unicité du prolongement des mesures, page 11, implique donc que $\mu_X = \mu_Y$, ce qui termine la preuve. ■

Remarque II.7.1 Soit X , une v.a. réelle. S'il existe $y_0 \in \mathbb{R}$ tel que $F_X(y_0) = 1$, alors $\mathbf{P}(X \leq y_0) = 1$ et donc \mathbf{P} -p.s. $X \leq y_0$. De même s'il existe $y_0 \in \mathbb{R}$ tel que $F_X(y_0) = 0$, alors $\mathbf{P}(X \leq y_0) = 0$, c'est-à-dire que $\mathbf{P}(X > y_0) = 1$ et donc \mathbf{P} -p.s. $X > y_0$. □

La proposition suivante montre les liens entre la densité (éventuelle) d'une variable réelle et sa fonction de répartition.

Proposition II.7.6 ◀► Soit X , une v.a. réelle \mathcal{F} -mesurable. On suppose qu'elle admet une densité f . Alors, on a :

$$\forall y \in \mathbb{R}, \quad F_X(y) = \mathbf{P}(X \leq y) = \int_{-\infty}^y f(z) dz . \quad (\text{II.40})$$

De plus on a l'équivalence suivante.

- (a) La fonction de répartition est dérivable partout, de dérivée continue.
- (b) La variable admet une densité qui est continue.

De plus, si les conditions équivalentes (a) ou (b) sont satisfaites, alors (II.40) implique que $F'_X = f$.

Preuve : on a

$$\begin{aligned} \mathbf{P}(X \leq y) &= \mathbf{P}(X \in]-\infty, y]) \\ &= \mathbf{E}[\mathbf{1}_{\{X \in]-\infty, y]\}}] \\ &= \mathbf{E}[\mathbf{1}_{]-\infty, y]}(X)] \\ &= \int_{\mathbb{R}} \mathbf{1}_{]-\infty, y]}(z) f(z) dz = \int_{-\infty}^y f(z) dz , \end{aligned}$$

l'avant-dernière égalité étant une application directe de la formule de la densité (II.33), page 63, avec $g(z) = \mathbf{1}_{]-\infty, y]}(z)$. Le reste du lemme est une conséquence de résultats bien connus sur la dérivation des fonctions de classe C^1 . ■

La propriété d'oubli des v.a. de loi exponentielle. En application, on montre que les variables de loi exponentielle (voir la définition II.6.2, page 64) sont caractérisées par ce que l'on appelle « la propriété d'oubli » comme définie dans la proposition suivante.

Proposition II.7.7 (Propriété d'oubli) Soit $X : \Omega \rightarrow \mathbb{R}_+$, une v.a. \mathcal{F} -mesurable telle que

$$\forall t \in]0, \infty[, \quad \mathbf{P}(X > t) > 0 .$$

On suppose qu'elle possède la **propriété d'oubli**, c'est-à-dire :

$$\forall s, t \in \mathbb{R}_+, \quad \mathbf{P}(X - t > s \mid X > t) = \mathbf{P}(X > s) .$$

Alors X est une variable exponentielle pour un certain paramètre $c \in]0, \infty[$.

Preuve : on pose $f(t) = -\log \mathbf{P}(X > t)$. La fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est bien définie. Par ailleurs f est continue à droite car F_X est continue à droite et $f(t) = -\log(1 - F_X(t))$. La propriété d'oubli implique que $f(t) + f(s) = f(t + s)$, pour tous $s, t \in \mathbb{R}_+$. On pose $c = f(1)$. Pour tout entier $n \in \mathbb{N}^*$, $f(n) = cn$ et $nf(1/n) = c$. Donc pour tout $t \in \mathbb{Q}_+$, $f(t) = ct$. La continuité à droite implique que $f(t) = ct$ pour tout $t \in \mathbb{R}_+$. On a donc $\mathbf{P}(X > t) = e^{-ct}$, pour tout $t \in \mathbb{R}_+$ et c ne peut être nul, ce qui termine la preuve de la proposition. ■

Les lois exponentielles modélisent de nombreux phénomènes : intervalles de temps entre deux accidents, intervalles de temps entre l'émission d'électrons par une plaque de fer chauffée etc. Les

loi exponentielles sont couramment utilisées en théorie de la fiabilité pour modéliser les durées de vie d'appareils (ou d'animaux) : on néglige les périodes de rodage ou d'usure ; les raisons de la défaillance de l'appareil sont uniquement dues à des facteurs externes qui sont supposés statistiquement constants : sachant que l'appareil a fonctionné sans défaillance jusqu'au temps t , la probabilité qu'il fonctionne encore un temps s doit être la même que la probabilité qu'il fonctionne sans défaillance pendant un temps s . Sa loi est donc nécessairement exponentielle.

◀► **Pseudo-réciproque d'une fonction de répartition.** Il est naturel de se demander si une fonction satisfaisant les propriétés du (i) de la proposition II.7.4, page 69, est toujours la fonction de répartition d'une variable aléatoire. La proposition suivante répond par l'affirmative à cette question.

Définition II.7.5 ▶ (Pseudo-réciproque) Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante, continue à droite et telle que

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

On définit la fonction $G :]0, 1[\rightarrow \mathbb{R}$ par

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

Elle est appelée le *pseudo-réciproque* de F . □

Proposition II.7.8 ▶ (Pseudo-réciproque) Soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante, continue à droite et telle que

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

On note $G :]0, 1[\rightarrow \mathbb{R}$ sa pseudo-réciproque. La fonction G satisfait les propriétés suivantes.

- (i) La fonction G est continue à droite et croissante sur $]0, 1[$.
- (ii) Si F est strictement croissante et continue, alors elle réalise une bijection de \mathbb{R} sur $]0, 1[$ et G est sa fonction réciproque.
- (iii) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité et soit $U : \Omega \rightarrow]0, 1[$, une v.a. \mathcal{F} -mesurable uniforme sur $]0, 1[$. On pose

$$\forall \omega \in \Omega , \quad X(\omega) = G(U(\omega))$$

Alors, la fonction de répartition de X est F :

$$\forall x \in \mathbb{R}, \quad F(x) = \mathbf{P}(X \leq x) = \mu_X(]-\infty, x]) .$$

où μ_X désigne la loi de X .

Preuve : on montre d'abord (i). Le fait que G soit croissante est clair à partir de la définition. Soit $y \in]0, 1[$. Par (II.38) au point (i) de la proposition II.7.3, page 68, on sait que $G(y+)$ est la borne inférieure $\inf_{z \in]y, 1[} G(z)$. Supposons que $G(y+) > x > G(y)$. Alors, $F(x) > y$. D'autre part $x < G(z)$, et donc $F(x) \leq z$, pour tout $z \in]y, 1[$, ce qui implique que $F(x) \leq y$, entraînant une contradiction. On doit donc avoir $G(y+) = G(y)$ et G est continue à droite.

Le point (ii) est clair : nous laissons les détails au lecteur. Montrons (iii) : on voit facilement que $G(y) \leq x$ implique que $F(x) \geq y$ et que si $y < F(x)$, alors $G(y) \leq x$. Par conséquent $[0, F(x)[\subset G^{-1}(]-\infty, x]) \subset [0, F(x)]$. Donc

$$\{U < F(x)\} \subset \{G(U) \leq x\} \subset \{U \leq F(x)\} .$$

Comme la loi de U est diffuse (c'est la mesure de Lebesgue), $\mathbf{P}(U = F(x)) = 0$ et donc

$$\mathbf{P}(G(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x),$$

ce qui implique le résultat voulu. ■

La proposition précédente implique immédiatement le résultats suivant.

Corollaire II.7.9 ◀▶ *La fonction $F: \mathbb{R} \rightarrow \mathbb{R}$ est la fonction de répartition d'une mesure de probabilité sur \mathbb{R} ssi elle est croissante, continue à droite et telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow \infty} F(x) = 1$.*

Exemple II.7.6 Soit c , un réel strictement positif. On note μ la mesure de probabilité sur les boréliens de \mathbb{R} ayant pour densité $t \in \mathbb{R} \mapsto \mathbf{1}_{\mathbb{R}_+}(t)ce^{-ct}$ par rapport à la mesure de Lebesgue, c'est-à-dire :

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mu(B) = \int_{B \cap \mathbb{R}_+} ce^{-ct} dt.$$

Sa fonction de répartition est donc $F(x) = (1 - e^{-cx})\mathbf{1}_{\mathbb{R}_+}(x)$. La pseudo-réciproque de F est donc

$$\forall y \in]0, 1[, \quad G(y) = -\frac{1}{c} \log(1 - y).$$

Si U est une v.a. uniforme sur $[0, 1]$, alors $-c^{-1} \log(1 - U)$ est une variable de loi μ . □

II.8 Exercices.

Exercice II.8.1 Soit X une variable aléatoire à valeurs dans \mathbb{N} . Montrer qu'on a l'égalité

$$\mathbf{E}[X] = \sum_{n \geq 1} \mathbf{P}(X \geq n).$$

Exercice II.8.2 Soient $X, Y : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{N}$ deux variables à valeurs dans \mathbb{N} . Montrer de deux façons différentes que

$$\sum_{n \geq 0} n\mathbf{P}(X = n) + \sum_{n \geq 0} n\mathbf{P}(Y = n) = \sum_{n \geq 0} n\mathbf{P}(X + Y = n).$$

Exercice II.8.3 Soit Ω un ensemble. Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction. Montrer qu'il existe une unique paire de fonctions (g, h) de Ω dans \mathbb{R}_+ telles que $f = g - h$ et qu'on ait à la fois $g \leq |f|$ et $h \leq |f|$. Montrer que si (\tilde{g}, \tilde{h}) est une autre paire de fonctions positives telles que $f = \tilde{g} - \tilde{h}$, alors on a $g \leq \tilde{g}$ et $h \leq \tilde{h}$.

Exercice II.8.4 Soit X une variable aléatoire positive qui admet un moment d'ordre 2. On suppose $\mathbf{P}(X = 0) < 1$. On rappelle que

$$\mathbf{P}(X > 0) \geq \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]}.$$

Soit Y une variable aléatoire réelle telle que $\mathbf{P}(Y = 0) < 1$ et qui admet un moment d'ordre 4. Montrer que

$$\mathbf{P}(Y \neq 0) \geq \frac{\mathbf{E}[Y^2]^2}{\mathbf{E}[Y^4]}.$$

Exercice II.8.5 On dit qu'une variable aléatoire est centrée et réduite si elle admet un moment d'ordre 2, si elle est d'espérance nulle et de variance 1. Montrer que pour toute variable aléatoire X qui admet un moment d'ordre 2 et dont la variance n'est pas nulle, il existe une unique paire (a, b) de réels telle qu'on ait $a > 0$ et que la variable aléatoire $aX + b$ soit centrée et réduite.

Exercice II.8.6 Soit X une variable aléatoire réelle. Soit $M > 0$ un réel. Montrer que les deux propositions suivantes sont équivalentes.

1. On a l'inégalité $|X| \leq M$ presque sûrement.
2. La variable X admet des moments de tous les ordres et, pour tout $n \geq 1$, on a l'égalité $\mathbf{E}[X^{2n}] \leq M^{2n}$.

Exercice II.8.7 Soient X et Y deux variables aléatoires bornées. On rappelle que cela signifie que les deux fonctions X et Y de Ω dans \mathbb{R} sont des fonctions bornées, c'est-à-dire qu'il existe une constante c telle que pour tout $\omega \in \Omega$, $|X(\omega)| \leq c$ et $|Y(\omega)| \leq c$. On suppose que pour tout $n \geq 1$, on a $\mathbf{E}[X^n] = \mathbf{E}[Y^n]$. Montrer que pour toute fonction polynomiale $p : \mathbb{R} \rightarrow \mathbb{R}$, on a $\mathbf{E}[p(X)] = \mathbf{E}[p(Y)]$. Montrer que ce résultat est encore vrai si on suppose seulement que p est une fonction continue. En déduire que pour toute fonction continue $f : \mathbb{R} \rightarrow \mathbb{R}$, on a $\int_{\mathbb{R}} f d\mu_X = \int_{\mathbb{R}} f d\mu_Y$. En déduire que X et Y ont même loi.

L'hypothèse que X et Y sont bornées est essentielle. Il est possible de construire deux variables aléatoires qui admettent des moments de tous les ordres, dont les moments sont égaux, mais dont les lois sont différentes.

Exercice II.8.8 Que peut-on dire de la loi d'une variable aléatoire X à valeurs dans \mathbb{N} qui satisfait, pour tous $n, m \geq 0$, l'égalité

$$\mathbf{P}(X > n + m) = \mathbf{P}(X > n)\mathbf{P}(X > m) ?$$

Exercice II.8.9 Soient $X, Y, Z : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ trois variables aléatoires réelles. Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable. On suppose que X et Y ont même loi. Est-il vrai que $g(X)$ et $g(Y)$ ont même loi? Est-il vrai que $X + Z$ et $Y + Z$ ont même loi?

II.9 Vecteurs aléatoires.

II.9.a Rappels sur les mesures produit et le théorème de Fubini.

Définition II.9.1 \blacktriangleleft Soient (E_k, \mathcal{E}_k) , $1 \leq k \leq n$, des espaces mesurables. On introduit les notions suivantes.

- (a) On dit que $Q \subset E_1 \times \dots \times E_n$ est un *pavé* s'il est de la forme $Q = A_1 \times \dots \times A_n$, avec $A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n$. On note \mathcal{P} la classe des *pavés* de $E_1 \times \dots \times E_n$. On remarque que

$$(A_1 \times \dots \times A_n) \cap (B_1 \times \dots \times B_n) = (A_1 \cap B_1) \times \dots \times (A_n \cap B_n).$$

Cela implique que \mathcal{P} est stable par intersection simple et donc que c'est un pi-système.

- (b) On définit la *tribu produit* obtenue à partir des tribus $\mathcal{E}_1, \dots, \mathcal{E}_n$, comme la tribu sur l'espace produit $E_1 \times \dots \times E_n$ engendrée par \mathcal{P} . On utilise la notation

$$\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n := \sigma(\mathcal{P}).$$

qui est standard. □

Le théorème suivant, que nous rappelons sans preuve, montre l'existence et caractérise les mesures produit.

Théorème II.9.1 ◀► Soient $(E_k, \mathcal{E}_k, \mu_k)$, $1 \leq k \leq n$, des espaces mesurés. On suppose que les mesures μ_1, \dots, μ_n sont **sigma finies**. Alors il existe une unique mesure $\nu : \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n \rightarrow [0, \infty]$ telle que

$$\nu(A_1 \times \dots \times A_n) = \mu_1(A_1) \dots \mu_n(A_n), \quad \text{pour tous } A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n.$$

La mesure ν est appelée **la mesure produit des mesures** μ_1, \dots, μ_n et elle est notée habituellement par

$$\nu = \mu_1 \otimes \dots \otimes \mu_n.$$

Faisons un commentaire : $\mu_1 \otimes \mu_2$ généralise la notion d'aire et $\mu_1 \otimes \mu_2 \otimes \mu_3$ généralise la notion de volume acr

$$\text{Aire}([a, b] \times [c, d]) = (b-a)(d-c) \quad \text{et} \quad \text{Volume}([a, b] \times [c, d] \times [e, f]) = (b-a)(d-c)(f-e).$$

Dans le cas du produit de deux mesures, on rappelle, sans preuve, les théorèmes de Fubini-Tonelli et de Fubini-Lebesgue.

Théorème II.9.2 ◀► (**Fubini positif**) Soient $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$, deux espaces mesurés. On suppose que μ_1 et μ_2 sont sigma finies. Soit $f : E_1 \times E_2 \rightarrow [0, \infty]$, une application $\mathcal{E}_1 \otimes \mathcal{E}_2$ -mesurable. Alors, les assertions suivantes sont vraies.

(i) La fonction $y \in E_2 \mapsto \int_{E_1} f(x, y) \mu_1(dx) \in [0, \infty]$ est bien définie et elle est \mathcal{E}_2 -mesurable. De même, la fonction $x \in E_1 \mapsto \int_{E_2} f(x, y) \mu_2(dy) \in [0, \infty]$ est bien définie et elle est \mathcal{E}_1 -mesurable.

(ii) Les égalités suivantes sont vérifiées :

$$\int_{E_1 \times E_2} f d\mu_1 \otimes \mu_2 = \int_{E_2} \left(\int_{E_1} f(x, y) \mu_1(dx) \right) \mu_2(dy) = \int_{E_1} \left(\int_{E_2} f(x, y) \mu_2(dy) \right) \mu_1(dx). \quad (\text{II.41})$$

Théorème II.9.3 ◀► (**Fubini intégrable**) Soient $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$, deux espaces mesurés. On suppose que μ_1 et μ_2 sont **sigma finies**. Soit $f : E_1 \times E_2 \rightarrow \mathbb{C}$, une application $\mathcal{E}_1 \otimes \mathcal{E}_2$ -mesurable. On suppose que

$$\int_{E_1 \times E_2} |f| d\mu_1 \otimes \mu_2 < \infty.$$

Alors, les assertions suivantes sont vraies.

(i) Pour μ_2 -presque tout y , $f(\cdot, y)$ est μ_1 -intégrable. De même, pour μ_1 -presque tout x , $f(x, \cdot)$ est μ_2 -intégrable.

(ii) La fonction $y \mapsto \int_{E_1} f(x, y) \mu_1(dx)$, bien définie μ_2 -p.p. et est μ_2 -intégrable. De même, la fonction $x \mapsto \int_{E_2} f(x, y) \mu_2(dy)$ est définie μ_1 -p.p. et est μ_1 -intégrable.

(iii) Les égalités suivantes sont vérifiées :

$$\int_{E_1 \times E_2} f d\mu_1 \otimes \mu_2 = \int_{E_2} \left(\int_{E_1} f(x, y) \mu_1(dx) \right) \mu_2(dy) = \int_{E_1} \left(\int_{E_2} f(x, y) \mu_2(dy) \right) \mu_1(dx). \quad (\text{II.42})$$

Fubini dans le cas du produit de 3 espaces. On se donne $(E_1, \mathcal{E}_1, \mu_1)$, $(E_2, \mathcal{E}_2, \mu_2)$ et $(E_3, \mathcal{E}_3, \mu_3)$, trois espaces mesurés supposés sigma finis. Soit $f: E_1 \times E_2 \times E_3 \rightarrow [0, \infty]$, une fonction $\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3$ -mesurable. Alors appliquant Fubini-Tonelli à $E_1 \times (E_2 \times E_3)$ puis ensuite à $E_2 \times E_3$, on montre que

$$\int_{E_1 \times E_2 \times E_3} f d\mu_1 \otimes \mu_2 \otimes \mu_3 = \int_{E_1} \mu_1(dx_1) \left(\int_{E_2} \mu_2(dx_2) \left(\int_{E_3} \mu_3(dx_3) f(x_1, x_2, x_3) \right) \right).$$

Mais on aurait très bien pu faire les intégration dans un ordre différent et on aurait aussi obtenu :

$$\begin{aligned} \int_{E_1 \times E_2 \times E_3} f d\mu_1 \otimes \mu_2 \otimes \mu_3 &= \int_{E_1} \mu_1(dx_1) \left(\int_{E_3} \mu_3(dx_3) \left(\int_{E_2} \mu_2(dx_2) f(x_1, x_2, x_3) \right) \right) \\ &= \int_{E_2} \mu_2(dx_2) \left(\int_{E_1} \mu_1(dx_1) \left(\int_{E_3} \mu_3(dx_3) f(x_1, x_2, x_3) \right) \right) \\ &= \int_{E_2} \mu_2(dx_2) \left(\int_{E_3} \mu_3(dx_3) \left(\int_{E_1} \mu_1(dx_1) f(x_1, x_2, x_3) \right) \right) \\ &= \int_{E_3} \mu_3(dx_3) \left(\int_{E_1} \mu_1(dx_1) \left(\int_{E_2} \mu_2(dx_2) f(x_1, x_2, x_3) \right) \right) \\ &= \int_{E_3} \mu_3(dx_3) \left(\int_{E_2} \mu_2(dx_2) \left(\int_{E_1} \mu_1(dx_1) f(x_1, x_2, x_3) \right) \right) \end{aligned}$$

En tout, il y a $3! = 6$ façons d'intégrer.

◀► **Fubini dans le cas du produit de n espaces.** Plus généralement, soient $(E_k, \mathcal{E}_k, \mu_k)$, $1 \leq k \leq n$, des espaces mesurés supposés sigma-finis. Soit $f: E_1 \times \dots \times E_n \rightarrow [0, \infty]$, une fonction $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ -mesurable. En appliquant récursivement le théorème de Fubini-Tonelli, on montre que pour toute permutation $\gamma: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$, on a

$$\begin{aligned} &\int_{E_1 \times \dots \times E_n} f d\mu_1 \otimes \dots \otimes \mu_n \\ &= \int_{E_{\gamma(1)}} \mu_{\gamma(1)}(dx_{\gamma(1)}) \left(\int_{E_{\gamma(2)}} \mu_{\gamma(2)}(dx_{\gamma(2)}) \left(\dots \int_{E_{\gamma(n)}} \mu_{\gamma(n)}(dx_{\gamma(n)}) f(x_1, x_2, \dots, x_n) \dots \right) \right). \end{aligned}$$

Il y a en tout donc $n!$ façons d'intégrer.

◀► **Notations et mesures de Lebesgue.** Lorsque $(E, \mathcal{E}, \mu) = (E_k, \mathcal{E}_k, \mu_k)$, pour tout $1 \leq k \leq n$, on utilise la notation

$$(E^n, \mathcal{E}^{\otimes n}, \mu^{\otimes n}) := (E_1 \times \dots \times E_n, \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n, \mu_1 \otimes \dots \otimes \mu_n).$$

Lorsque l'on considère l'espace \mathbb{R}^n muni de la topologie de la norme et des boréliens associés, on rappelle, sans preuve, le théorème suivant.

Théorème II.9.4 ◀▶ Pour tous $m, n \in \mathbb{N}$, $\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$.

On rappelle que ℓ désigne la mesure de Lebesgue sur les Boréliens de \mathbb{R} . Le lemme précédent montre que $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R})^{\otimes n}$. On pose alors

$$\ell_n = \ell \otimes \dots \otimes \ell = \ell^{\otimes n},$$

C'est la mesure de Lebesgue en dimension n . C'est l'unique mesure ℓ_n telle que pour tous réels $a_1 \leq b_1, \dots, a_n \leq b_n$,

$$\ell_n([a_1, b_1[\times \dots \times [a_n, b_n[) = (b_1 - a_1) \dots (b_n - a_n). \quad (\text{II.43})$$

On a donc

$$\forall n, m \geq 1, \quad \ell_n \otimes \ell_m = \ell_{n+m}.$$

II.9.b Outil de calcul intégral en dimension n .

Changement de variable linéaire. On note $\{e_1, \dots, e_n\}$ la base canonique de \mathbb{R}^n . On note $M_n(\mathbb{R})$, l'espace des matrices carrées de taille $n \times n$. On note Id_n la matrice identité. Si $M = (m_{i,j})_{1 \leq i, j \leq n}$ est une matrice, alors on lui fait correspondre une application linéaire de \mathbb{R}^n dans \mathbb{R}^n (un endomorphisme) que l'on note M également et tel que les coordonnées du vecteur Me_j dans la base canonique soient données par la j -ème colonne $(m_{1,j}, \dots, m_{n,j})$ de la matrice M . On note $\det : M_n(\mathbb{R}) \rightarrow \mathbb{R}$, le déterminant, qui est l'application multilinéaire alternée des colonnes des matrices telle que $\det(\text{Id}_n) = 1$. On rappelle que $GL_n(\mathbb{R}) = \{M \in M_n(\mathbb{R}) : \det(M) \neq 0\}$ est le groupe des endomorphismes de \mathbb{R}^n qui sont inversibles. On rappelle que $\det(MN) = \det(M) \det(N)$. Soit $M \in GL_n(\mathbb{R})$ et $A \in \mathcal{B}(\mathbb{R}^n)$. Comme M et M^{-1} sont continues bijectives, $M(A)$ se voit comme l'image réciproque de A par M^{-1} et on a donc $M(A) \in \mathcal{B}(\mathbb{R}^n)$. On rappelle, sans preuve, le résultat suivant.

Théorème II.9.5 ◀▶ (Changement de variable linéaire) Pour toute matrice $M \in GL(\mathbb{R}^n)$,

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \quad \ell_n(M(A)) = |\det(M)| \ell_n(A).$$

De plus, pour toute fonction mesurable $f : \mathbb{R}^n \rightarrow [0, \infty]$ et tout $A \in \mathcal{B}(\mathbb{R}^n)$, on a

$$|\det(M)| \int_A f(M \cdot \mathbf{x}) d\ell_n(\mathbf{x}) = \int_{M(A)} f(\mathbf{x}) d\ell_n(\mathbf{x}),$$

avec une assertion analogue pour les fonctions ℓ_n -intégrables sur $M(A)$.

Mentionnons deux conséquences du théorème de changement de variable linéaire.

- Soit F un sous-espace affine de dimension $m < n$. Alors $\ell_n(F) = 0$. En effet, il existe $\mathbf{x} \in \mathbb{R}^n$ et $M \in GL_n(\mathbb{R})$ telle que $\mathbf{x} + M(F)$ soit inclus dans l'hyperplan H des vecteurs donc la dernière coordonnée est nulle. On a donc $|\det(M)| \ell_n(F) \leq \ell_n(H)$. En identifiant \mathbb{R}^n à $\mathbb{R}^{n-1} \times \mathbb{R}$ et ℓ_n à $\ell_{n-1} \otimes \ell$, le théorème de Fubini implique que $\ell_n(H) = \int_{\mathbb{R}^{n-1}} (\int_{\mathbb{R}} \mathbf{1}_{\{0\}}(y) \ell(dy)) \ell_{n-1}(dx) = 0$. □
- Les seules transformations linéaires qui préservent la mesure de Lebesgue sont les applications linéaires M telles que $|\det(M)| = 1$. □

Exemple II.9.1 (*Volume de la boule Euclidienne*) Pour tout $n \in \mathbb{N}$ et tout $r \in \mathbb{R}_+$, on pose

$$B_n(0, r) = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n : x_1^2 + \dots + x_n^2 \leq r^2\} \quad \text{et} \quad b_n = \ell_n(B_n(0, 1))$$

$B_n(0, r)$ est la boule de centre 0 et de rayon r pour la norme Euclidienne et b_n est le volume de la boule unité. Le changement de variable linéaire implique que

$$\ell_n(B_n(0, r)) = b_n r^n$$

Combiné avec Fubini, cela implique

$$\begin{aligned} b_n &= \int_{\mathbb{R}^n} dx_1 \dots dx_n \mathbf{1}_{\{x_1^2 + \dots + x_n^2 \leq 1\}} = \int_{[-1, 1]} dx_n \int_{\mathbb{R}^{n-1}} dx_1 \dots dx_{n-1} \mathbf{1}_{\{x_1^2 + \dots + x_{n-1}^2 \leq 1 - x_n^2\}} \\ &= \int_{[-1, 1]} dx_n b_{n-1} (1 - x_n^2)^{\frac{n-1}{2}}. \end{aligned}$$

On pose $I_n = \int_{-1}^1 (1 - y^2)^{n/2} dy$. Ce qui précède montre que $b_n = b_{n-1} I_{n-1}$. En effectuant une intégration par parties, on trouve que $I_n = \frac{n}{n+1} I_{n-2}$ et donc pour tout $n \geq 2$, on a $n I_{n-1} I_{n-2} = 2 I_1 I_0$. Or $I_0 = 2$ et $I_1 = \pi/2$. Donc $b_n = I_{n-1} I_{n-2} b_{n-2} = 2\pi b_{n-2}/n$. Or $b_1 = 2$ et $b_2 = \pi$. On en déduit

$$\forall p \in \mathbb{N}^*, \quad b_{2p} = \frac{\pi^p}{p!} \quad \text{et} \quad b_{2p-1} = \frac{\pi^{p-1}}{(p - \frac{1}{2})(p - \frac{3}{2}) \dots \frac{5}{2} \frac{3}{2} \frac{1}{2}}, \quad (\text{II.44})$$

ce qui se résume par

$$b_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)}, \quad (\text{II.45})$$

où Γ est la fonction gamma d'Euler. □

Changement de variable géométrique. Chacun connaît le théorème de changement de variable pour l'intégrale de Riemann qui s'énonce comme suit : soit $\varphi : [a, b] \rightarrow \mathbb{R}$, une application de classe C^1 telle que φ' ne s'annule pas. Elle est donc strictement monotone. On note $[c, d] = \varphi([a, b])$. Alors pour toute fonction continue $f : [c, d] \rightarrow \mathbb{R}$, on a

$$\int_a^b f(\varphi(t)) |\varphi'(t)| dt = \int_c^d f(t) dt. \quad (\text{II.46})$$

Preuve : on suppose sans restriction que φ' reste strictement positive et on se ramène à un calcul de primitive. Pour tout $x \in [a, b]$, on pose $G(x) = \int_a^x f(\varphi(t)) \varphi'(t) dt$. On a donc $G' = \varphi' \cdot f \circ \varphi$. On pose aussi $F(y) := \int_c^y f(t) dt$, pour tout $y \in [c, d]$ et on a $F' = f$. On a donc $(F \circ \varphi)' = \varphi' \cdot f \circ \varphi = G'$, donc $F \circ \varphi$ et G ne diffèrent que par une constante. Or elles sont nulles en a , donc $F \circ \varphi = G$ qui implique (II.46). □

Le théorème de changement de variable géométrique, que nous énonçons ci-dessous, est une généralisation pour la mesure de Lebesgue en dimension n de (II.46). Avant d'établir ce résultat, rappelons quelques définitions de calcul différentiel.

Définition II.9.2 Soient U et V , deux ouverts non-vides de \mathbb{R}^n . Une fonction $\varphi : U \rightarrow V$ est un *homéomorphisme de U sur V* si φ est bijective, continue et si sa réciproque φ^{-1} est continue. □

Si $\varphi : U \rightarrow V$ est un homéomorphisme de l'ouvert U sur V , pour tout ouvert $O \subset U$, $\varphi(O)$ est un ouvert : en effet, c'est la pré-image de O par la réciproque de φ . On voit que φ est une bijection transportant la topologie de U sur celle de V et que c'est aussi une bijection de $\mathcal{B}(U)$ sur $\mathcal{B}(V)$.

Définition II.9.3 Soit U un ouvert de \mathbb{R}^n et $\varphi : U \rightarrow \mathbb{R}^n$.

- (a) Soit $\mathbf{x} \in U$. On dit que φ est différentiable en \mathbf{x} s'il existe une application linéaire $M : \mathbb{R}^n \rightarrow \mathbb{R}^n$ telle que

$$\lim_{\|\mathbf{h}\| \rightarrow 0} \frac{1}{\|\mathbf{h}\|} \cdot \|\varphi(\mathbf{x} + \mathbf{h}) - \varphi(\mathbf{x}) - M\mathbf{h}\| = 0.$$

On note alors $M = D_{\mathbf{x}}(\varphi)$: c'est la *différentielle de φ en \mathbf{x}* . La fonction φ est dite *différentiable sur U* si elle est différentiable en tout point de U .

- (b) φ est *continûment différentiable sur U* si $\mathbf{x} \in U \mapsto D_{\mathbf{x}}(\varphi) \in M_n(\mathbb{R})$ est continue, pour la topologie de la norme sur l'espace vectoriel $M_n(\mathbb{R})$. \square

On observe que la différentiabilité implique la continuité, la réciproque étant évidemment fausse.

Supposons que $\varphi : U \rightarrow \mathbb{R}^n$ soit continûment différentiable sur U . Soit $\mathbf{x} = (x_1, \dots, x_n) \in U$. On note les coordonnées du vecteur $\varphi(\mathbf{x})$ par $(\varphi_1(\mathbf{x}), \dots, \varphi_n(\mathbf{x}))$: les φ_j sont donc des applications de U dans \mathbb{R} . Le fait que φ soit continûment différentiable est équivalent à ce que pour tous $1 \leq i, j \leq n$, et pour tout $\mathbf{x} = (x_1, \dots, x_n) \in U$, la limite suivante existe dans \mathbb{R} ,

$$\frac{\partial \varphi_j}{\partial x_i}(\mathbf{x}) = \lim_{r \rightarrow 0} \frac{\varphi_j(x_1, \dots, x_i + r, \dots, x_n) - \varphi_j(\mathbf{x})}{r}$$

et que $\frac{\partial \varphi_j}{\partial x_i} : U \rightarrow \mathbb{R}$ soit continue. Il est alors facile de vérifier que l'application linéaire $D_{\mathbf{x}}(\varphi)$ est donnée par la matrice

$$D_{\mathbf{x}}(\varphi) = \left(\frac{\partial \varphi_i}{\partial x_j}(\mathbf{x}) \right)_{1 \leq i, j \leq n}.$$

On définit ensuite le *Jacobien de φ* , noté $J_{\varphi} : U \rightarrow \mathbb{R}$, par $J_{\varphi}(\mathbf{x}) := \det(D_{\mathbf{x}}(\varphi))$. Si φ est continûment différentiable, alors son Jacobien est continu sur U . Rappelons, sans preuve, le théorème d'inversion globale.

Théorème II.9.6 (Théorème d'inversion globale) Soient U et V , deux ouverts de \mathbb{R}^n que l'on suppose non-vides. Soit $\varphi : U \rightarrow V$, une bijection continue. On suppose également que φ est continûment différentiable et que $J_{\varphi}(\mathbf{x}) \neq 0$, pour tout $\mathbf{x} \in U$. Alors $\varphi^{-1} : V \rightarrow U$ est également continûment différentiable. On dit que φ est un **difféomorphisme de U sur V** .

Sous les mêmes hypothèses que le théorème d'inversion globale, il est facile de montrer que

$$\forall \mathbf{y} \in V, \quad D_{\mathbf{y}}(\varphi^{-1}) = (D_{\varphi^{-1}(\mathbf{y})}(\varphi))^{-1} \quad \text{et} \quad J_{\varphi^{-1}}(\mathbf{y}) = \frac{1}{J_{\varphi}(\varphi^{-1}(\mathbf{y}))}. \quad (\text{II.47})$$

Nous rappelons, sans preuve, le théorème de changement de variable géométrique.

Théorème II.9.7 $\blacktriangleleft \blacktriangleright$ (Changement de variable géométrique) Soient U et V , deux ouverts de \mathbb{R}^n non-vides. Soit $\varphi : U \rightarrow V$, une bijection continûment différentiable sur U telle que $J_{\varphi}(\mathbf{x}) \neq 0$, pour tout $\mathbf{x} \in U$. Alors pour toute fonction mesurable $f : V \rightarrow [0, \infty]$, on a

$$\int_U f(\varphi(\mathbf{x})) |J_{\varphi}(\mathbf{x})| d\ell_n(\mathbf{x}) = \int_V f(\mathbf{y}) d\ell_n(\mathbf{y}). \quad (\text{II.48})$$

Autrement dit, la mesure image par φ^{-1} de ℓ_n sur V est la mesure $|J_{\varphi}| \ell_n(\cdot \cap U)$. Par ailleurs, si $f : V \rightarrow \mathbb{C}$ est ℓ_n -intégrable, alors $f \circ \varphi$ est $|J_{\varphi}| \ell_n$ intégrable sur U et l'égalité (II.48) a lieu.

Coordonnées polaires Pour tout point $(\theta, r) \in \mathbb{R}^2$, on pose $\varphi(\theta, r) = (|r| \cos \theta, |r| \sin \theta) \in \mathbb{R}^2$. Il est facile de voir que φ est continûment différentiable sur l'ouvert $U =]0, 2\pi[\times]0, \infty[$ et on a

$$\frac{\partial \varphi_1}{\partial \theta} = -r \sin \theta, \quad \frac{\partial \varphi_1}{\partial r} = \cos \theta, \quad \frac{\partial \varphi_2}{\partial \theta} = r \cos \theta, \quad \frac{\partial \varphi_2}{\partial r} = \sin \theta, \quad J_\varphi(\theta, r) = r.$$

On note D , le demi-axe des abscisses positives : $D = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0, x_1 \geq 0\}$. On voit que φ est bijective de U sur $V = \mathbb{R}^2 \setminus D$. Par le théorème de changement de variable géométrique, pour toute fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ qui est mesurable, on a

$$\int_{\mathbb{R}^2 \setminus D} f \, d\ell_2 = \int_{]0, 2\pi[\times]0, \infty[} f(r \cos \theta, r \sin \theta) r \, d\ell_2(\theta, r).$$

On observe ensuite que $N := [0, 2\pi] \times [0, \infty[\setminus]0, 2\pi[\times]0, \infty[$ est la réunion des ensembles $\{0\} \times [0, \infty[$, $\{2\pi\} \times [0, \infty[$ et $[0, 2\pi] \times \{0\}$, tous ℓ_2 -négligeables. De même, D est ℓ_2 -négligeable. On en déduit la proposition suivante qui est la formule de changement de variable en coordonnées polaires.

Proposition II.9.8 ◀▶ (Changement de variable polaire) Pour toute fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ mesurable, on a

$$\int_{\mathbb{R}^2} f \, d\ell_2 = \int_{]0, 2\pi[\times \mathbb{R}_+} f(r \cos \theta, r \sin \theta) r \, d\ell_2(\theta, r), \quad (\text{II.49})$$

avec un énoncé analogue pour les fonctions ℓ_2 -intégrables sur \mathbb{R}^2 .

Exemple II.9.2 ◀▶ En application, montrons que

$$I = \int_{\mathbb{R}} \exp(-x^2) \, d\ell(x) = \sqrt{\pi}. \quad (\text{II.50})$$

Preuve de (II.50) : pour tout $(x_1, x_2) \in \mathbb{R}^2$, on pose $g(x_1, x_2) = \exp(-x_1^2 - x_2^2)$, qui est continue donc mesurable et strictement positive. Il est facile de montrer que $I = \int_{\mathbb{R}} \exp(-x^2) \, d\ell(x) < \infty$. Le théorème de Fubini positif implique que

$$\int_{\mathbb{R}^2} g(x_1, x_2) \, d\ell_2(x_1, x_2) = \int_{\mathbb{R}} \exp(-x_1^2) \left(\int_{\mathbb{R}} \exp(-x_2^2) \, d\ell(x_2) \right) \, d\ell(x_1) = I^2.$$

Or $g(r \cos \theta, r \sin \theta) = \exp(-r^2)$. Donc $I^2 = \int_{]0, 2\pi[\times]0, \infty[} r \exp(-r^2) \, d\ell_2(\theta, r)$. Par le théorème de Fubini,

$$\int_{]0, 2\pi[\times]0, \infty[} r \exp(-r^2) \, d\ell_2(\theta, r) = \left(\int_{]0, 2\pi[} d\ell(\theta) \right) \left(\int_{]0, \infty[} r \exp(-r^2) \, d\ell(r) \right) = 2\pi \int_{]0, \infty[} r \exp(-r^2) \, d\ell(r).$$

Comme $x \mapsto \int_{]0, x]} r \exp(-r^2) \, d\ell(r)$ est la primitive de $x \mapsto x \exp(-x^2)$ qui est nulle en 0, c'est donc la fonction $x \mapsto (1 - \exp(-x^2))/2$ et on a $\int_{]0, \infty[} r \exp(-r^2) \, d\ell(r) = 1/2$, ce qui implique que $I^2 = \pi$. ◻

II.9.c Vecteurs aléatoires ; covariance.

On munit \mathbb{R}^n du produit scalaire canonique $\langle \cdot, \cdot \rangle$ et on note $\|\cdot\|$ la norme Euclidienne associée. C'est-à-dire que si $\mathbf{x} = (x_1, \dots, x_n)$ et $\mathbf{y} = (y_1, \dots, y_n)$ sont deux vecteurs de \mathbb{R}^n on a

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_n y_n \quad \text{et} \quad \|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_n^2}.$$

Sauf mention du contraire, $(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace de probabilité de référence pour cette section.

Définition II.9.4 ◀▶ (Vecteur aléatoire) Une v.a. $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ \mathcal{F} -mesurable est appelée *vecteur aléatoire*. □

Proposition II.9.9 ◀▶ Soient $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$, des fonctions. On définit $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ en posant

$$\forall \omega \in \Omega, \quad \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n.$$

Alors, les deux assertions suivantes sont équivalentes.

- (i) Pour tout $1 \leq j \leq n$, X_j est \mathcal{F} -mesurable.
- (ii) \mathbf{X} est un vecteur aléatoire \mathcal{F} -mesurable.

◀▶ **Preuve :** montrons d'abord que (ii) \Rightarrow (i). On suppose donc que $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ est \mathcal{F} -mesurable. Pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ et tout $1 \leq j \leq n$, on pose $\pi_j(\mathbf{x}) = x_j$. Alors, π_j est j^{me} -projection canonique et elle est mesurable (car elle est continue). Or on remarque que alors pour tout $\omega \in \Omega$, $X_j(\omega) = \pi_j(\mathbf{X}(\omega))$. Donc X_j est également \mathcal{F} -mesurable, par composition d'application mesurable. Cela montre (i).

Réciproquement, montrons que (i) \Rightarrow (ii) : on suppose donc que pour tout $1 \leq j \leq n$, X_j est \mathcal{F} -mesurable. Soient $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, alors on remarque que

$$\begin{aligned} \{\mathbf{X} \in A_1 \times \dots \times A_n\} &= \{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in A_1 \times \dots \times A_n\} \\ &= \{\omega \in \Omega : X_1(\omega) \in A_1 \text{ et } \dots \text{ et } X_n(\omega) \in A_n\} \\ &= \bigcap_{1 \leq j \leq n} \{X_j \in A_j\} \in \mathcal{F}. \end{aligned}$$

Puisque les pavés engendrent par définition la tribu produit $\mathcal{B}(\mathbb{R})^{\otimes n}$ et puisque $\mathcal{B}(\mathbb{R})^{\otimes n} = \mathcal{B}(\mathbb{R}^n)$, alors la proposition II.1.2, page 28 (cette proposition permet de simplifier les vérifications à faire lorsque l'on veut montrer qu'une application est mesurable) implique que \mathbf{X} est un vecteur aléatoire $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$ -mesurable, ce qui est (ii). Cela termine la preuve de la proposition. ■

Covariance.

Définition II.9.5 ◀▶ On introduit les définitions suivantes.

- (a) (Covariance de deux v.a. réelles) Soient X, Y , deux v.a. réelles \mathcal{F} -mesurables. On suppose qu'elles admettent un moment d'ordre 2. L'inégalité de Hölder implique que XY est intégrable et on pose

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

Cette quantité est appelée la *covariance de X et Y*. On vérifie facilement que

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

(b) (*Matrice de covariance d'un vecteur aléatoire*) Soit $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$, un vecteur aléatoire \mathcal{F} -mesurable. On dit qu'il admet un moment d'ordre 2 si $\mathbf{E}[\|\mathbf{X}\|^2] < \infty$. Comme

$$\mathbf{E}[\|\mathbf{X}\|^2] = \mathbf{E}[X_1^2 + \dots + X_n^2] = \mathbf{E}[X_1^2] + \dots + \mathbf{E}[X_n^2],$$

Le fait que \mathbf{X} admette un moment d'ordre 2 est *équivalent* à ce que les v.a. réelles X_k admettent toutes des moments d'ordre 2. Cela permet de poser

$$\Gamma_{\mathbf{X}} = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq n}.$$

Cette matrice carrée $n \times n$ est appelée la *matrice de covariance de \mathbf{X}* . □

La proposition suivante donne quelques propriétés des matrices de covariance.

Proposition II.9.10 ◀▶ $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$, un vecteur aléatoire \mathcal{F} -mesurable admettant un moment d'ordre 2. On note Γ sa matrice de covariance et on note

$$\mathbf{v} = \mathbf{E}[\mathbf{X}] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]) \in \mathbb{R}^n,$$

que l'on appelle son **vecteur moyen**. Alors, les propriétés suivantes sont vérifiées.

(i) Pour tout $\mathbf{u} \in \mathbb{R}^n$,

$$\mathbf{E}[\langle \mathbf{u}, \mathbf{X} \rangle] = \langle \mathbf{u}, \mathbf{v} \rangle \quad \text{et} \quad \text{var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \langle \mathbf{u}, \Gamma \mathbf{u} \rangle.$$

Cela implique que $\langle \mathbf{u}, \Gamma \mathbf{u} \rangle \geq 0$, pour tout $\mathbf{u} \in \mathbb{R}^n$ et donc que Γ est une matrice symétrique positive.

(ii) Soit M , une matrice réelle $n \times n$. On note M^* sa transposée. Alors

$$\Gamma_{M\mathbf{X}} = M\Gamma M^*.$$

◀▶ **Preuve** : il est facile de vérifier que $\mathbf{E}[\langle \mathbf{u}, \mathbf{X} \rangle] = \langle \mathbf{u}, \mathbf{v} \rangle$ et que par le lemme II.4.8 (page 57) on a

$$\text{var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \text{var}(\langle \mathbf{u}, \mathbf{X} - \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle) = \text{var}(\langle \mathbf{u}, \mathbf{X} - \mathbf{v} \rangle).$$

On peut donc se ramener au cas où $\mathbf{E}[\mathbf{X}] = 0$. Dans ce cas $\text{cov}(X_i, X_j) = \mathbf{E}[X_i X_j]$. Pour tout $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, on a donc

$$\text{var}(\langle \mathbf{u}, \mathbf{X} \rangle) = \mathbf{E}[\langle \mathbf{u}, \mathbf{X} \rangle^2] = \sum_{1 \leq j, k \leq n} u_j u_k \mathbf{E}[X_j X_k] = \langle \mathbf{u}, \Gamma \mathbf{u} \rangle,$$

en développant le carré. Cela montre (i).

Montrons (ii) : par définition de la transposée, $\langle \mathbf{x}, M\mathbf{y} \rangle = \langle M^*\mathbf{x}, \mathbf{y} \rangle$, pour tous $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Le (i) implique donc pour tout $\mathbf{u} \in \mathbb{R}^n$,

$$\langle \mathbf{u}, \Gamma_{M\mathbf{X}} \mathbf{u} \rangle = \text{var}(\langle \mathbf{u}, M\mathbf{X} \rangle) = \text{var}(\langle M^*\mathbf{u}, \mathbf{X} \rangle) = \langle M^*\mathbf{u}, \Gamma M^*\mathbf{u} \rangle = \langle \mathbf{u}, M\Gamma M^*\mathbf{u} \rangle$$

et il est facile de voir que cela implique (ii). ■

II.10 Fonction caractéristique des vecteurs aléatoires.

II.10.a Rappel sur les intégrales à paramètres.

Voici un théorème sur la continuité des intégrales à paramètre où l'espace sur lequel la continuité a lieu est un espace métrique.

Théorème II.10.1 ◀▶ (Continuité des intégrales à paramètre) Soit (Y, d) , un espace métrique. Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f : E \times Y \rightarrow \mathbb{C}$. On fait les hypothèses suivantes.

- (a) Pour tout $y \in Y$, $f(\cdot, y)$ est \mathcal{E} -mesurable.
- (b) Pour μ -presque tout $x \in E$, $y \in Y \mapsto f(x, y) \in \mathbb{C}$ est continue.
- (c) Il existe une fonction $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ telle que pour tout $y \in Y$,

$$|f(x, y)| \leq g(x) \quad \text{pour } \mu\text{-presque tout } x \in E.$$

Alors la fonction $F : Y \rightarrow \mathbb{C}$ donnée par $F(y) = \int_E f(x, y) \mu(dx)$ est bien définie et continue.

Preuve : (c) implique que $F(y)$ est bien définie pour tout $y \in Y$. Soient $y_n \in Y$, $n \in \mathbb{N}$, tels que $\lim_n d(y_n, y) = 0$. Pour tout $n \in \mathbb{N}$, on pose $h_n(x) = f(x, y_n)$, $x \in E$. L'hypothèse (a) entraîne que les h_n sont des applications \mathcal{E} -mesurables. De plus, l'hypothèse (b) montre que μ -presque partout on a $\lim_n h_n = h$, où $h(x) = f(x, y)$, $x \in E$. Enfin, on a $|h_n| \leq g$, μ -p.p. pour tout $n \in \mathbb{N}$, d'après l'hypothèse (c). Le théorème de convergence dominée implique alors que

$$\lim_{n \rightarrow \infty} F(y_n) = \lim_{n \rightarrow \infty} \int_E h_n d\mu = \int_E h d\mu = F(y),$$

ce qui entraîne bien que F est continue. ■

Le théorème suivant concerne la dérivabilité des intégrales à paramètres.

Théorème II.10.2 ◀▶ (Dérivabilité des intégrales à paramètre) Soit I un intervalle ouvert de \mathbb{R} . Soit (E, \mathcal{E}, μ) , un espace mesuré. Soit $f : E \times I \rightarrow \mathbb{C}$. On fait les hypothèses suivantes.

- (a) On suppose que pour tout $y \in I$, $f(\cdot, y) \in \mathcal{L}^1(E, \mathcal{E}, \mu)$.
- (b) Pour μ -presque tout $x \in E$ $y \in I \mapsto f(x, y)$ est dérivable sur I . On note $\frac{\partial f}{\partial y}(x, y)$ cette dérivée.
- (c) Il existe une fonction $g \in \mathcal{L}^1(E, \mathcal{E}, \mu)$ telle que pour tout $y \in I$,

$$\left| \frac{\partial f}{\partial y}(x, y) \right| \leq g(x) \quad \text{pour } \mu\text{-presque tout } x \in E.$$

Alors la fonction $F : I \rightarrow \mathbb{C}$ donnée par $F(y) = \int_E f(x, y) \mu(dx)$ est bien définie, continue et dérivable sur I . De plus on a

$$F'(y) = \int_E \frac{\partial f}{\partial y}(x, y) d\mu(x), \quad y \in I.$$

Preuve : soit $N \in \mathcal{E}$ tel que $\mu(N) = 0$ et tel que (b) et (c) soient vérifiées pour tout $x \in E \setminus N$. On fixe $y_n \in I \setminus \{y\}$, $n \in \mathbb{N}$, tels que $\lim_n |y_n - y| = 0$. Pour tout $n \in \mathbb{N}$, on pose $h_n(x) = \frac{f(x, y_n) - f(x, y)}{y_n - y}$. C'est une fonction \mathcal{E} -mesurable et le théorème des accroissements finis et (c) impliquent que $|h_n(x)| \leq g(x)$, pour tout $x \in E \setminus N$ et pour tout $n \in \mathbb{N}$. Par ailleurs, $\lim_n h_n(x) = \frac{\partial f}{\partial y}(x, y)$, pour tout $x \in E \setminus N$. Le théorème de convergence dominée s'applique donc à la suite des h_n et montre que

$$\lim_{n \rightarrow \infty} \frac{F(y_n) - F(y)}{y_n - y} = \lim_{n \rightarrow \infty} \int_E h_n d\mu = \int_E \frac{\partial f}{\partial y}(x, y) \mu(dx),$$

ce qui implique le résultat désiré. ■

Exemple II.10.1 On montre plus loin (voir l'exemple II.9.2, page 80) que $\int_{\mathbb{R}} e^{-x^2} d\ell(x) = \sqrt{\pi}$, c'est-à-dire $\int_{\mathbb{R}} e^{-x^2/2} d\ell(x) = \sqrt{2\pi}$. On pose

$$\forall u \in \mathbb{R}, \quad F(u) = \int_{\mathbb{R}} e^{iux} e^{-x^2/2} d\ell(x).$$

qui est bien défini car $|e^{iux} e^{-x^2/2}| = e^{-x^2/2}$ est Lebesgue intégrable, par un critère usuel. On a aussi $F(0) = \sqrt{2\pi}$. On pose $f(u, x) = e^{iux} e^{-x^2/2}$. On a donc

$$\forall u, x \in \mathbb{R}, \quad \left| \frac{\partial f}{\partial u}(u, x) \right| = |ix e^{iux} e^{-x^2/2}| \leq |x| e^{-x^2/2}.$$

Comme $x \mapsto |x| e^{-x^2/2}$ est Lebesgue intégrable, f satisfait les hypothèses du théorème II.10.2. Cela implique que F est dérivable sur \mathbb{R} et que

$$\forall u \in \mathbb{R}, \quad F'(u) = \int_{\mathbb{R}} ix e^{iux} e^{-x^2/2} d\ell(x).$$

Par un calcul de primitive évident, pour tout a ,

$$\int_{[-a, a]} (iu - x) e^{iux} e^{-x^2/2} d\ell(x) = e^{-a^2/2} (e^{iua} - e^{-iua}) \xrightarrow{a \rightarrow \infty} 0.$$

Donc par convergence dominée, pour tout $u \in \mathbb{R}$, on a

$$iuF(u) + iF'(u) = \lim_{a \rightarrow \infty} \int_{\mathbb{R}} \mathbf{1}_{[-a, a]}(x) (iu - x) e^{iux} e^{-x^2/2} d\ell(x) = 0,$$

c'est-à-dire que F satisfait l'équation différentielle :

$$\forall u \in \mathbb{R}, \quad F'(u) + uF(u) = 0.$$

On a donc $(e^{u^2/2} F(u))' = 0$. Cela entraîne $F(u) = F(0) e^{-u^2/2}$ et on a donc montré que

$$\forall u \in \mathbb{R}, \quad F(u) = \int_{\mathbb{R}} e^{iux} e^{-x^2/2} d\ell(x) = \sqrt{2\pi} e^{-u^2/2}.$$

Ce résultat est utilisé plus loin dans le cours. □

II.10.b Injectivité de la fonction caractéristique, formule d'inversion.

On munit \mathbb{R}^n du produit scalaire canonique qui est noté $\langle \cdot, \cdot \rangle$ et de la norme Euclidienne correspondante qui est notée $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Sauf mention du contraire, $(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace de probabilité de référence pour cette section. On introduit la transformée de Fourier des mesures finies sur \mathbb{R}^n . Cet outil d'analyse trouve de nombreuses applications en probabilité.

Définition II.10.1 ◀▶ (Transformée de Fourier des mesures de probabilité sur \mathbb{R}^n) Soit $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, une mesure de probabilité. Pour tout $\mathbf{u} \in \mathbb{R}^n$ on pose

$$\widehat{\mu}(\mathbf{u}) := \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} \mu(d\mathbf{x}) .$$

$\widehat{\mu} : \mathbb{R}^n \rightarrow \mathbb{C}$ est la transformée de Fourier de la mesure μ . □

On donne immédiatement la définition de la fonction caractéristique d'une variables aléatoire.

Définition II.10.2 ◀▶ (Fonction caractéristique d'une v.a.) Soit $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$, un vecteur aléatoire \mathcal{F} -mesurable. Sa fonction caractéristique est la transformée de Fourier de sa loi, c'est-à-dire

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{\mu}_{\mathbf{X}}(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} \mu_{\mathbf{X}}(d\mathbf{x}) = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] .$$

par le théorème de transfert. On voit que la fonction caractéristique d'une variable aléatoire ne dépend que de la loi de cette variable. Donc Montrer des propriétés des transformées de Fourier de mesures de probabilité revient donc exactement à montrer des propriétés sur les fonctions caractéristiques des variables aléatoire. □

Proposition II.10.3 ◀▶ Soit $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, une mesure de probabilité. Pour tout $\mathbf{u} \in \mathbb{R}^n$, on a

$$|\widehat{\mu}(\mathbf{u})| \leq 1 = \widehat{\mu}(0) . \tag{II.51}$$

De plus $\widehat{\mu}$ est une fonction continue.

Preuve : puisque $|e^{i\langle \mathbf{u}, \mathbf{x} \rangle}| = 1$, l'inégalité triangulaire implique que

$$\left| \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} \mu(d\mathbf{x}) \right| \leq \int_{\mathbb{R}^n} |e^{i\langle \mathbf{u}, \mathbf{x} \rangle}| \mu(d\mathbf{x}) = \int_{\mathbb{R}^n} \mu(d\mathbf{x}) = \mu(\mathbb{R}^n) = 1 ,$$

ce qui montre (II.51).

La continuité de $\widehat{\mu}$ est une conséquence du théorème II.10.1 des intégrales à paramètres, page 83 : pour tout $\mathbf{u}, \mathbf{x} \in \mathbb{R}^n$, on pose $f(\mathbf{u}, \mathbf{x}) = e^{i\langle \mathbf{u}, \mathbf{x} \rangle}$. Il est clair que pour tout $\mathbf{u} \in \mathbb{R}^n$, on a $|f(\mathbf{u}, \mathbf{x})| = 1$. Or la fonction constante égale à 1 est μ -intégrable. Le théorème s'applique donc et montre que

$$\mathbf{u} \in \mathbb{R}^n \longmapsto \widehat{\mu}(\mathbf{u}) = \int_{\mathbb{R}^n} f(\mathbf{u}, \mathbf{x}) \mu(d\mathbf{x})$$

est continue. ■

Remarque II.10.1 Soit $X : \Omega \rightarrow \mathbb{N}$, une v.a. \mathcal{F} -mesurable dont la fonction génératrice φ_X s'étend au disque complexe unité fermé

$$\forall z \in \mathbb{C} \text{ tel que } |z| \leq 1, \quad \varphi_X(z) = \sum_{n \in \mathbb{N}} z^n \mathbf{P}(X = n).$$

On a en particulier

$$\forall u \in \mathbb{R}, \quad \widehat{\mu}_X(u) = \mathbf{E}[e^{iuX}] = \varphi_X(e^{iu}).$$

Donc la fonction génératrice étend la fonction caractéristique d'une variable entière. \square

Exemple II.10.2 (Densités gaussiennes) On rappelle la définition II.6.3, page 64, de la densité gaussienne de moyenne $m \in \mathbb{R}$ et d'écart-type $\sigma \in]0, \infty[$

$$g_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

On rappelle qu'en application du théorème de la dérivation des intégrales à paramètre, on a montré à l'exemple II.10.1, page 84, que

$$\forall u \in \mathbb{R}, \quad \int_{\mathbb{R}} e^{iux} e^{-x^2/2} d\ell(x) = \sqrt{2\pi} e^{-u^2/2}.$$

C'est-à-dire que

$$\forall u \in \mathbb{R}, \quad \int_{\mathbb{R}} e^{iux} g_{0,1}(x) d\ell(x) = e^{-u^2/2}.$$

Par un changement de variable linéaire on montre facilement que pour tous $\sigma \in]0, \infty[$ et tous $u \in \mathbb{R}$, on a

$$\int_{\mathbb{R}} e^{iux} g_{0,\sigma}(x) d\ell(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2\sigma^2}} d\ell(x) = e^{-\frac{1}{2}\sigma^2 u^2}. \quad (\text{II.52})$$

Pour tout $\sigma > 0$, on pose ensuite

$$\forall \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad g_{\sigma}(\mathbf{x}) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \|\mathbf{x}\|^2}.$$

On remarque tout d'abord que pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$\begin{aligned} g_{\sigma}(x_1, \dots, x_n) &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2}(x_1^2 + \dots + x_n^2)\right) \\ &= \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_j^2}{2\sigma^2}} \\ &= g_{0,\sigma}(x_1)g_{0,\sigma}(x_2) \dots g_{0,\sigma}(x_n). \end{aligned}$$

En appliquant Fubini, on voit que

$$\int_{\mathbb{R}^n} g_{\sigma} d\ell_n = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} g_{\sigma}(x_1, \dots, x_n) dx_1 \dots dx_n$$

$$\begin{aligned}
&= \left(\int_{\mathbb{R}} g_{0,\sigma}(x_1) dx_1 \right) \cdots \left(\int_{\mathbb{R}} g_{0,\sigma}(x_n) dx_n \right) \\
&= 1
\end{aligned}$$

car $\int_{\mathbb{R}} g_{0,\sigma}(x) dx = 1$, puis que c'est la densité d'une variable aléatoire. On remarque également que

$$g_{\sigma}(\mathbf{x}) = \sigma^{-n} g_1(\sigma^{-1}\mathbf{x}).$$

On pose ensuite la notation suivante

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{g}_{\sigma}(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} g_{\sigma}(\mathbf{x}) \ell_n(d\mathbf{x}). \quad (\text{II.53})$$

Cela est bien défini car d'une part g_{σ} est positive donc $|g_{\sigma}| = g_{\sigma}$ et g_{σ} est Lebesgue-intégrable car $\int_{\mathbb{R}^n} |g_{\sigma}| d\ell_n = \int_{\mathbb{R}^n} g_{\sigma} d\ell_n = 1$. On remarque tout d'abord que pour tout $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ et tout $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$,

$$\begin{aligned}
e^{i\langle \mathbf{u}, \mathbf{x} \rangle} g_{\sigma}(x_1, \dots, x_n) &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(iu_1x_1 + \dots + iu_nx_n - \frac{1}{2\sigma^2}(x_1^2 + \dots + x_n^2)\right) \\
&= \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{iu_jx_j - \frac{x_j^2}{2\sigma^2}} \\
&= (e^{iu_1x_1} g_{0,\sigma}(x_1)) (e^{iu_2x_2} g_{0,\sigma}(x_2)) \cdots (e^{iu_nx_n} g_{0,\sigma}(x_n)).
\end{aligned}$$

En appliquant Fubini, on voit que

$$\begin{aligned}
\widehat{g}_{\sigma}(\mathbf{u}) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{iu_1x_1 + \dots + iu_nx_n} g_{\sigma}(x_1, \dots, x_n) dx_1 \cdots dx_n \\
&= \left(\int_{\mathbb{R}} e^{iu_1x_1} g_{0,\sigma}(x_1) dx_1 \right) \cdots \left(\int_{\mathbb{R}} e^{iu_nx_n} g_{0,\sigma}(x_n) dx_n \right) \\
&= (e^{-\frac{\sigma^2 u_1^2}{2}}) \cdots (e^{-\frac{\sigma^2 u_n^2}{2}})
\end{aligned}$$

par II.52. On a donc montré

$$\widehat{g}_{\sigma}(\mathbf{u}) = \exp\left(-\frac{\sigma^2}{2} \|\mathbf{u}\|^2\right) = \left(\frac{\sqrt{2\pi}}{\sigma}\right)^n g_{\frac{1}{\sigma}}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^n. \quad (\text{II.54})$$

Nous utilisons ce résultat plus loin. □

Le théorème suivant montre que la transformée de Fourier est injective.

Théorème II.10.4 ◀▶ (Injectivité de la transformée de Fourier) Soient $\mu, \nu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, des mesures de probabilité. Alors

$$\widehat{\mu} = \widehat{\nu} \iff \mu = \nu.$$

Ce résultat se traduit immédiatement pour les variables aléatoires et montre que la fonction caractéristique d'un vecteur aléatoire caractérise entièrement sa loi.

Théorème II.10.5 ◀▶ (Injectivité de la fonction caractéristique) Soient $(\Omega, \mathcal{F}, \mathbf{P})$ et $(\Omega', \mathcal{F}', \mathbf{P}')$, des espaces de probabilité. Soit $X : \Omega \rightarrow \mathbb{R}^d$, une v.a. \mathcal{F} -mesurable et soit $Y : \Omega' \rightarrow \mathbb{R}^d$, une v.a. \mathcal{F}' -mesurable. Alors X et Y ont même loi ssi elles ont même fonction caractéristique, c'est-à-dire

$$\widehat{\mu}_X = \widehat{\mu}_Y \iff \mu_X = \mu_Y .$$

La preuve du théorème II.10.4 est longue et occupe le reste de la section. Les arguments utilisés permettent également de montrer le théorème suivant.

Théorème II.10.6 ◀▶ (Inversion de Fourier L^1) Soit $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, une mesure de probabilité. On note ℓ_n la mesure de Lebesgue sur \mathbb{R}^n . On suppose

$$\int_{\mathbb{R}^n} |\widehat{\mu}| d\ell_n < \infty . \quad (\text{II.55})$$

Alors, μ admet une densité h par rapport à ℓ_n , c'est-à-dire que $\mu = h \ell_n$. La densité h est continue bornée et elle est donnée par la formule d'inversion suivante :

$$\forall \mathbf{x} \in \mathbb{R}^n , \quad h(\mathbf{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-i\langle \mathbf{u}, \mathbf{x} \rangle} \widehat{\mu}(\mathbf{u}) d\ell_n(\mathbf{u}) . \quad (\text{II.56})$$

Ce résultat se traduit immédiatement pour les variables aléatoires.

Théorème II.10.7 ◀▶ (Inversion de Fourier L^1) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $X : \Omega \rightarrow \mathbb{R}^d$, une v.a. \mathcal{F} -mesurable. On note $\widehat{\mu}_X$ la fonction caractéristique de X . On note ℓ_n la mesure de Lebesgue sur \mathbb{R}^n . On suppose

$$\int_{\mathbb{R}^n} |\widehat{\mu}_X| d\ell_n < \infty . \quad (\text{II.57})$$

Alors, X est un vecteur aléatoire admettant une densité h par rapport à ℓ_n , c'est-à-dire que $\mu_X = h \ell_n$. La densité h est continue bornée et elle est donnée par la formule d'inversion suivante :

$$\forall \mathbf{x} \in \mathbb{R}^n , \quad h(\mathbf{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-i\langle \mathbf{u}, \mathbf{x} \rangle} \widehat{\mu}_X(\mathbf{u}) d\ell_n(\mathbf{u}) . \quad (\text{II.58})$$

Avant de prouver les théorèmes II.10.4 et II.10.6, donnons l'application suivante du théorème d'inversion de Fourier.

Exemple II.10.3 (Loi de Cauchy) On se place en dimension 1. On remarque que $x \mapsto e^{-|x|}$ est Lebesgue-intégrable. Pour tout $a > 0$, et tout $u \in \mathbb{R}$, on a

$$\begin{aligned} \int_{-a}^a dx e^{iux} e^{-|x|} &= \int_0^a dx e^{(iu-1)x} + \int_0^a dx e^{-(iu+1)x} \\ &= \frac{e^{(iu-1)a} - 1}{iu - 1} - \frac{e^{-(iu+1)a} - 1}{iu + 1} \xrightarrow{a \rightarrow \infty} \frac{1}{iu + 1} - \frac{1}{iu - 1} = \frac{2}{u^2 + 1} . \end{aligned}$$

Par convergence dominée $\lim_{a \rightarrow \infty} \int_{-a}^a dx e^{iux} e^{-|x|} = \int_{\mathbb{R}} dx e^{iux} e^{-|x|}$. Donc

$$\forall u \in \mathbb{R}, \quad \int_{\mathbb{R}} dx e^{iux} e^{-|x|} = \frac{2}{u^2 + 1} .$$

Comme le membre de droite est Lebesgue intégrable, la formule d'inversion de Fourier L^1 implique que

$$\forall x \in \mathbb{R}, \quad e^{-|x|} = \frac{1}{\pi} \int_{\mathbb{R}} \frac{e^{-iux} du}{u^2 + 1},$$

qui est une formule pénible à obtenir par des moyens élémentaires. La fonction $u \in \mathbb{R} \mapsto \frac{1}{\pi(u^2+1)}$ est appelée *densité de Cauchy*. \square

Preuve du théorème II.10.4.

On rappelle la définition de g_σ donnée dans l'exemple II.10.2 page 86. Pour toute mesure de probabilité $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$ et tout $\sigma > 0$, on pose

$$(g_\sigma * \mu)(\mathbf{x}) = \int_{\mathbb{R}^n} g_\sigma(\mathbf{x} - \mathbf{y}) d\mu(\mathbf{y}). \quad (\text{II.59})$$

On vérifie facilement que par le théorème II.10.1 de continuité des intégrales à paramètre $g_\sigma * \mu$ est continue. Par ailleurs on vérifie immédiatement que

$$\|g_\sigma * \mu\|_\infty \leq (\sigma\sqrt{2\pi})^{-n} \mu(\mathbb{R}^n) < \infty. \quad (\text{II.60})$$

où pour toute fonction $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $\|h\|_\infty := \sup_{\mathbf{x} \in \mathbb{R}^n} |h(\mathbf{x})|$ désigne la norme uniforme de h .

Lemme II.10.8 Soit $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, une mesure de probabilité. Soit $\sigma \in]0, \infty[$. On a

$$(g_\sigma * \mu)(\mathbf{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-\frac{\sigma^2}{2}\|\mathbf{u}\|^2} e^{i\langle \mathbf{x}, \mathbf{u} \rangle} \widehat{\mu}(-\mathbf{u}) d\ell_n(\mathbf{u}). \quad (\text{II.61})$$

Preuve : on rappelle qu'en conséquence de (II.54), on a $(\sigma\sqrt{2\pi})^n g_\sigma = \widehat{g}_{1/\sigma}$. Donc, par le théorème de Fubini,

$$\begin{aligned} (\sigma\sqrt{2\pi})^n (g_\sigma * \mu)(\mathbf{x}) &= \int_{\mathbb{R}^n} \widehat{g}_{\frac{1}{\sigma}}(\mathbf{x} - \mathbf{y}) d\mu(\mathbf{y}) = \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{u}) e^{i\langle \mathbf{x} - \mathbf{y}, \mathbf{u} \rangle} g_{\frac{1}{\sigma}}(\mathbf{u}) \\ &= \int_{\mathbb{R}^n} d\ell_n(\mathbf{u}) e^{i\langle \mathbf{x}, \mathbf{u} \rangle} g_{\frac{1}{\sigma}}(\mathbf{u}) \int_{\mathbb{R}^n} d\mu(\mathbf{y}) e^{-i\langle \mathbf{y}, \mathbf{u} \rangle}, \end{aligned}$$

ce qui permet de conclure. \blacksquare

Soit $\psi \in C_c(\mathbb{R}^n)$. Elle est uniformément continue ; on note son *module d'uniforme continuité* par

$$\forall \delta \in \mathbb{R}_+, \quad \omega_\psi(\delta) := \sup \{ |\psi(\mathbf{x}) - \psi(\mathbf{y})| ; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} - \mathbf{y}\| \leq \delta \}.$$

Clairement $\delta \mapsto \omega_\psi(\delta)$ est une application croissante donc mesurable et on vérifie que

$$\forall \psi \in C_c(\mathbb{R}^n), \quad \forall \delta \in \mathbb{R}_+, \quad \omega_\psi(\delta) \leq 2\|\psi\|_\infty \quad \text{et} \quad \lim_{\delta \rightarrow 0} \omega_\psi(\delta) = 0. \quad (\text{II.62})$$

Lemme II.10.9 Soit $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, une mesure de probabilité. Soit $\psi \in C_c(\mathbb{R}^n)$ et $\sigma > 0$. Alors,

$$\left| \int_{\mathbb{R}^n} \psi d\mu - \int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu) d\ell_n \right| \leq \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}) \xrightarrow{\sigma \rightarrow 0} 0.$$

Preuve : l'invariance de ℓ_n par translation et le théorème Fubini impliquent

$$\int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu) d\ell_n = \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z} - \mathbf{y}) \psi(\mathbf{z}) = \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) \psi(\mathbf{z} + \mathbf{y}).$$

Comme $\int g_\sigma d\ell_n = 1$, on a d'une part $\int \psi d\mu = \int d\mu(\mathbf{y}) \int g_\sigma(\mathbf{z}) \psi(\mathbf{y}) d\ell_n(\mathbf{z})$ donc

$$\begin{aligned} a &:= \left| \int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu) d\ell_n - \int_{\mathbb{R}^n} \psi d\mu \right| \\ &= \left| \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) \psi(\mathbf{z} + \mathbf{y}) - \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) \psi(\mathbf{y}) \right| \\ &= \left| \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) (\psi(\mathbf{z} + \mathbf{y}) - \psi(\mathbf{y})) \right| \\ &\leq \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) |\psi(\mathbf{z} + \mathbf{y}) - \psi(\mathbf{y})| \\ &\leq \int_{\mathbb{R}^n} d\mu(\mathbf{y}) \int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) \omega_\psi(\|\mathbf{z}\|) = \left(\int_{\mathbb{R}^n} d\mu(\mathbf{y}) \right) \cdot \left(\int_{\mathbb{R}^n} d\ell_n(\mathbf{z}) g_\sigma(\mathbf{z}) \omega_\psi(\|\mathbf{z}\|) \right) \\ &\leq \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}), \end{aligned}$$

et on conclut par (II.62) et par convergence dominée. ■

Fin de la preuve du théorème II.10.4.

On suppose que $\hat{\mu} = \hat{\nu}$. Le lemme II.10.8 implique ensuite que pour tout $\sigma > 0$, $g_\sigma * \mu = g_\sigma * \nu$. Ceci combiné avec le lemme II.10.9 implique que pour toute fonction $\psi \in C_c(\mathbb{R}^n)$ et tout $\sigma > 0$,

$$\left| \int_{\mathbb{R}^n} \psi d\mu - \int_{\mathbb{R}^n} \psi d\nu \right| \leq 2 \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}) \xrightarrow{\sigma \rightarrow 0} 0,$$

et donc $\int_{\mathbb{R}^n} \psi d\mu = \int_{\mathbb{R}^n} \psi d\nu$, pour tout $\psi \in C_c(\mathbb{R}^n)$, ce qui implique que $\mu = \nu$ grâce au lemme suivant.

Lemme II.10.10 Soient $\mu, \nu : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$, deux mesures de probabilité. Alors

$$\mu = \nu \iff \forall \psi \in C_c(\mathbb{R}^n), \int_{\mathbb{R}^n} \psi d\mu = \int_{\mathbb{R}^n} \psi d\nu.$$

Preuve : pour tout $A \subset \mathbb{R}^n$, on note $\overset{\circ}{A}$ l'intérieur de A qui est le plus grand ouvert contenu dans A et \bar{A} , l'adhérence de A , qui est le plus petit fermé contenant A :

$$\overset{\circ}{A} = \bigcup \{U \subset A : U \text{ ouvert}\} \quad \text{et} \quad \bar{A} = \bigcap \{F \subset E : A \subset F \text{ et } F \text{ fermé}\}.$$

Pour tout $\mathbf{x} \in \mathbb{R}^n$ et tout $r \in \mathbb{R}_+$, on pose aussi

$$B(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| < r\} \quad \text{et} \quad \bar{B}(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{y}\| \leq r\},$$

qui sont les boules ouverte et fermée de centre \mathbf{x} et de rayon r . Pour tout $A \subset \mathbb{R}^n$, non-vidé, on définit

$$\forall \mathbf{x} \in E, \quad d(\mathbf{x}, A) = \inf_{\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|,$$

qui est fonction *distance* à A . On voit facilement que $d(\cdot, A)$ est 1-Lipschitzienne, c'est-à-dire que

$$\forall \mathbf{x}, \mathbf{x}' \in E, \quad |d(\mathbf{x}, A) - d(\mathbf{x}', A)| \leq \|\mathbf{x} - \mathbf{x}'\|.$$

Cela implique en particulier que $d(\cdot, A)$ est continue. Il est facile de voir que l'ensemble de ses zéros est l'adhérence de A :

$$\{\mathbf{x} \in \mathbb{R}^n : d(\mathbf{x}, A) = 0\} = \bar{A}.$$

Donc, si F est un fermé $d(\mathbf{x}, F) = 0$ ssi $\mathbf{x} \in F$.

Soit $U \subset \mathbb{R}^n$ un ouvert **borné**. Pour tout entier $p \geq 1$, on pose

$$\mathbf{x} \in \mathbb{R}^n, \quad \psi_p(\mathbf{x}) = \min(1, p \cdot d(\mathbf{x}, \mathbb{R}^n \setminus U)).$$

Ce qui précède montre que ψ_p est une fonction continue. De plus Si $\mathbf{x} \notin U$, alors $d(\mathbf{x}, \mathbb{R}^n \setminus U) = 0$ et donc $\psi_p(\mathbf{x}) = 0$. Comme U est borné, on en déduit que ψ_p est une fonction continue à support compact donc $\psi_p \in C_c(\mathbb{R}^n)$. Par conséquent

$$\forall p \geq 1, \quad \int_{\mathbb{R}^n} \psi_p d\mu = \int_{\mathbb{R}^n} \psi_p d\nu \tag{II.63}$$

On vérifie ensuite immédiatement que pour tout $\mathbf{x} \in \mathbb{R}^n$, que $\psi_p(\mathbf{x}) \leq \psi_{p+1}(\mathbf{x})$. Enfin si $\mathbf{x} \in U$, comme il n'est pas dans le fermé $\mathbb{R}^n \setminus U$, on a $d(\mathbf{x}, \mathbb{R}^n \setminus U) > 0$; donc pour tout p assez grand on a $p \cdot d(\mathbf{x}, \mathbb{R}^n \setminus U) > 1$ et donc $\psi_p(\mathbf{x}) = 1$. Par conséquent,

$$\forall p \geq 1, \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad 0 \leq \psi_p(\mathbf{x}) \leq \psi_{p+1}(\mathbf{x}) \leq \mathbf{1}_U(\mathbf{x}) \quad \text{et} \quad \lim_{p \rightarrow \infty} \psi_p(\mathbf{x}) = \mathbf{1}_U(\mathbf{x}).$$

Par convergence monotone, on a donc

$$\lim_{p \rightarrow \infty} \int_{\mathbb{R}^n} \psi_p d\mu = \int_{\mathbb{R}^n} \mathbf{1}_U d\mu = \mu(U) \quad \text{et} \quad \lim_{p \rightarrow \infty} \int_{\mathbb{R}^n} \psi_p d\nu = \int_{\mathbb{R}^n} \mathbf{1}_U d\nu = \nu(U).$$

Par (II.63), cela implique que pour tout ouvert borné U , $\mu(U) = \nu(U)$. On note ensuite $\mathcal{P} := \{\mathbb{R}^n\} \cup \{U \subset \mathbb{R}^n : U \text{ ouvert borné}\}$. On vérifie facilement que \mathcal{P} est un pi-système engendrant $\mathcal{B}(\mathbb{R}^n)$. On a montré que μ et ν coïncident sur \mathcal{P} . Le Théorème I.1.12 d'unicité du prolongement des mesures page 11 s'applique et permet de montrer que $\mu = \nu$. ■

Preuve du théorème II.10.6.

On rappelle la définition (II.59) de $g_\sigma * \mu$ et la formule (II.61) du lemme II.10.8. On observe tout d'abord que par convergence dominée pour tout $\mathbf{x} \in \mathbb{R}^n$, $\lim_{\sigma \rightarrow 0} g_\sigma * \mu(\mathbf{x}) = h(\mathbf{x})$, où la fonction h est donnée par (II.56). On fixe ensuite $\psi \in C_c(\mathbb{R}^n)$ et on observe ensuite que

$$|\psi(\mathbf{x})(g_\sigma * \mu)(\mathbf{x})| \leq C|\psi(\mathbf{x})|$$

où $C = (2\pi)^{-n} \int |\widehat{\mu}| d\ell_n$. Comme $C|\psi|$ est Lebesgue intégrable, on a, par convergence dominée,

$$\lim_{\sigma \rightarrow 0} \int \psi(g_\sigma * \mu) d\ell_n = \int \psi h d\ell_n .$$

Le lemme II.10.9 implique ensuite que $\int \psi d\mu = \int \psi h d\ell_n$, pour toute fonction $\psi \in C_c(\mathbb{R}^n)$. Le lemme II.10.10 page 90 implique facilement que $\mu = h \ell_n$, ce qui termine la preuve. ■

II.11 Transformée de Laplace; fonctions génératrices.

Sauf mention du contraire, $(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace de probabilité de référence pour cette section.

Définition II.11.1 (*Transformée de Laplace d'une v.a. positive*) Soit $X : \Omega \rightarrow \mathbb{R}_+$, une v.a. \mathcal{F} -mesurable **positive**. Pour tout $\lambda \in \mathbb{R}_+$, on remarque que $0 \leq e^{-\lambda X} \leq 1$. On peut donc définir la fonction suivante

$$\forall \lambda \in \mathbb{R}_+, \quad L_X(\lambda) = \mathbf{E}[e^{-\lambda X}] = \int_{\mathbb{R}} e^{-\lambda x} d\mu_X(x) .$$

La fonction $L_X : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est appelé la *transformée de Laplace de X* , ou de sa loi. □

Remarque II.11.1 Soit $X : \Omega \rightarrow \mathbb{N}$, une v.a. \mathcal{F} -mesurable dont on note la fonction génératrice φ_X . On remarque que

$$\forall \lambda \in \mathbb{R}_+, \quad \varphi_X(e^{-\lambda}) = L_X(\lambda) . \quad (\text{II.64})$$

Autrement dit, la fonction génératrice d'une v.a. X à valeurs dans \mathbb{N} est une version de la transformée de Laplace. □

La transformée de Laplace est proche de la transformée de Fourier. Elle est propre aux variables positives et elle est d'usage plus commode que la transformée de Fourier car elle est positive. La proposition suivante détaille quelques propriétés de cette transformée.

Proposition II.11.1 Soit $X : \Omega \rightarrow \mathbb{R}_+$, une v.a. \mathcal{F} -mesurable. Alors, les assertions suivantes sont vérifiées.

- (i) La transformée de Laplace de X est positive, décroissante, continue sur \mathbb{R}_+ , analytique sur $]0, \infty[$ et $L_X(0) = 1$. De plus pour tous réels $\lambda, a > 0$,

$$\mathbf{P}(X \geq a) \leq \frac{1 - L_X(\lambda)}{1 - e^{-a\lambda}} \quad \text{et} \quad \mathbf{P}(X \leq a) \leq e^{a\lambda} L_X(\lambda) .$$

- (ii) L_X admet une dérivée à droite finie en $0+$ ssi $\mathbf{E}[X] < \infty$. Dans ce cas on a

$$L'_X(0+) = -\mathbf{E}[X] .$$

(iii) Plus généralement, L_X admet une dérivée $p^{\text{ième}}$ à droite finie en $0+$ ssi X admet un moment d'ordre p . Dans ce cas

$$L_X^{(p)}(0+) = (-1)^p \mathbf{E}[X^p] .$$

Preuve : montrons le point (i). Le fait que L_X soit décroissante et positive est évident ainsi que $L_X(0) = 1$. La continuité sur \mathbb{R}_+ est une conséquence immédiate du théorème de continuité des intégrales à paramètre et des inégalités de Markov établies à la section II.4.

Il reste à montrer qu'elle est analytique sur $]0, \infty[$. On fixe $\lambda_0 > 0$. Pour tout $\lambda \in]-\lambda_0, \lambda_0[$, on a

$$\begin{aligned} \sum_{n \in \mathbb{N}} |\mathbf{E}[e^{-\lambda_0 X} (-\lambda X)^n / n!]| &\leq \sum_{n \in \mathbb{N}} \mathbf{E}[e^{-\lambda_0 X} (|\lambda| X)^n / n!] \\ &\leq \mathbf{E}[e^{-(\lambda_0 - |\lambda|)X}] < \infty , \end{aligned}$$

la seconde ligne étant obtenue par interversion positive série/espérance. Les hypothèses du théorème d'interversion L^1 sont donc vérifiées et on a

$$L_X(\lambda_0 + \lambda) = \sum_{n \in \mathbb{N}} \lambda^n \mathbf{E}[e^{-\lambda_0 X} (-X)^n / n!] ,$$

ce qui montre bien que L_X est développable en série entière au voisinage de chaque point réel $\lambda_0 > 0$, c'est-à-dire que L_X est analytique sur $]0, \infty[$.

Montrons (ii) : on suppose que $\mathbf{E}[X] < \infty$. Par convexité, $0 \leq \lambda^{-1}(1 - e^{-\lambda X}) \leq X$. Le théorème de convergence dominée implique que

$$\lambda^{-1}(1 - L_X(\lambda)) = \mathbf{E}[\lambda^{-1}(1 - e^{-\lambda X})] \xrightarrow{\lambda \rightarrow 0+} \mathbf{E}[X] .$$

Cela montre donc que L_X admet une dérivée à droite en 0 et que $L_X'(0+) = -\mathbf{E}[X]$.

Réciproquement, supposons que L_X admette une dérivée à droite finie en $0+$. D'après ce qui précède, il suffit de montrer que $\mathbf{E}[X] < \infty$. Or le lemme de Fatou implique

$$\mathbf{E}[X] = \mathbf{E}[\liminf_{\lambda \rightarrow 0+} \lambda^{-1}(1 - e^{-\lambda X})] \leq \liminf_{\lambda \rightarrow 0+} \lambda^{-1}(1 - L_X(\lambda)) = -L_X'(0+) < \infty ,$$

ce qui termine la preuve de (ii).

Montrons (iii) par récurrence. La propriété est vraie pour $p = 1$. On la suppose vraie pour p . Si L_X admet une dérivée $(p+1)^{\text{ième}}$ à droite, alors elle admet une dérivée $p^{\text{ième}}$ à droite et l'hypothèse de récurrence implique que X a un moment d'ordre p . On peut donc supposer dans restriction que X a un moment d'ordre p . On pose $m_p = \mathbf{E}[X^p]$ et on note ν la mesure sur \mathbb{R}_+ ayant pour densité $x \mapsto x^p / m_p$ par rapport à μ_X . On a donc

$$\int_{\mathbb{R}} e^{-\lambda x} d\nu(x) = \frac{1}{m_p} \int_{\mathbb{R}} x^p e^{-\lambda x} d\mu(x) .$$

Le (i) implique que pour tout $\lambda > 0$,

$$\int_{\mathbb{R}} e^{-\lambda x} d\nu(x) = \frac{(-1)^p}{m_p} L_X^{(p)}(\lambda) .$$

Par cette égalité et le (ii), X admet un moment d'ordre $p + 1$ ssi $\int_{\mathbb{R}} x d\nu(x) < \infty$, ce qui équivaut à ce que L_X admette une dérivée $(p + 1)^{\text{ième}}$ à droite. Cela termine la preuve du point (iii) par récurrence. ■

Exemple II.11.1 Soit $X : \Omega \rightarrow \mathbb{R}_+$, une v.a. \mathcal{F} -mesurable de loi exponentielle de paramètre $c \in]0, \infty[$. On rappelle que cela signifie que la loi de X admet la densité $f(t) = \mathbf{1}_{\mathbb{R}_+}(t)ce^{-ct}$, $t \in \mathbb{R}$. La densité étant nulle sur l'ensemble des réels négatifs on peut calculer sa transformée de Laplace et on voit facilement

$$L_X(\lambda) = \mathbf{E}[e^{-\lambda X}] = \int_0^\infty ce^{-ct}e^{-\lambda t} dt = \frac{c}{c + \lambda} = \frac{1}{1 + c^{-1}\lambda}.$$

Si $\lambda \in]0, c[$, on a le développement en série

$$L_X(\lambda) = \sum_{p \in \mathbb{N}} (-1)^p c^{-p} \lambda^p.$$

Par le théorème précédent, on en déduit que

$$\mathbf{E}[X^p] = (-1)^p L_X^{(p)}(0+) = p!c^{-p}.$$

pour tout $p \in \mathbb{N}$. □

On rappelle le principe des zéros isolés : soit $U \subset \mathbb{C}$, un ouvert connexe. Soit $f : U \rightarrow \mathbb{C}$, une fonction analytique, c'est-à-dire développable en série entière au voisinage de chaque point ; si f n'est pas la fonction nulle, alors l'ensemble $\{z \in U : f(z) = 0\}$ est sans point d'accumulation. On utilise le principe des zéros isolés pour démontrer le théorème suivant.

Théorème II.11.2 (Injectivité de la trans. de Laplace) Soient X, Y , des v.a. positives \mathcal{F} -mesurables. Si l'ensemble

$$\{\lambda \in \mathbb{R}_+ : L_X(\lambda) = L_Y(\lambda)\}$$

possède un point d'accumulation $\lambda_0 > 0$, alors X et Y ont même loi. Cela implique notamment

$$L_X = L_Y \iff \mu_X = \mu_Y.$$

Preuve : on pose $U = \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$ et $\bar{U} = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq 0\}$. Pour tout $z \in \bar{U}$, on peut poser

$$L_X(z) = \mathbf{E}[e^{-zX}].$$

Le théorème de continuité des intégrales à paramètres implique facilement que L_X est continue sur \bar{U} . En raisonnant comme dans la proposition II.11.1 (i), on montre que L_X est analytique sur U . De même pour L_Y . On note S l'ensemble des zéros de $L_X - L_Y$. Si cet ensemble admet un point d'accumulation dans U alors $L_X = L_Y$ sur U . Par continuité, $L_X = L_Y$ sur \bar{U} . Notamment on a pour tout $u \in \mathbb{R}$, $\hat{\mu}_X(u) = \hat{\mu}_Y(u)$, ce qui implique que X et Y ont même loi par injectivité de la transformée de Fourier. ■

II.12 Exercices

Exercice II.12.1 Calculer l'espérance, la variance et la fonction génératrice d'une variable aléatoire X dont la loi est

- de Bernoulli de paramètre $p \in [0, 1]$,
- binomiale de paramètres $n \geq 0$ et $p \in [0, 1]$,
- géométrique de paramètre $p \in]0, 1[$,
- de Poisson de paramètre $\lambda > 0$.

Exercice II.12.2 Soit $m \geq 1$ un entier. Donner un exemple d'une variable aléatoire réelle qui admet un moment d'ordre m mais pas de moment d'ordre $m + 1$.

Exercice II.12.3 Montrer qu'une variable aléatoire qui suit la loi normale centrée réduite admet des moments de tous les ordres et les calculer.

Exercice II.12.4 Soit X une variable aléatoire réelle qui admet des moments de tous les ordres. Pour tout $n \geq 0$ on pose $m_n = \mathbf{E}[X^n]$. Pour chacune des assertions suivantes, dire si elle est vraie ou fausse, en justifiant par une démonstration ou un contre-exemple.

- Pour tout $n \geq 0$, $m_n \geq 0$.
- La suite $(m_n)_{n \geq 0}$ est monotone.
- Pour tous $k, n \geq 1$ tels que $k \leq n$, on a $m_{2n} \geq (m_{2k})^{\frac{n}{k}}$
- Pour tous $n \geq 1$ et $x_1, \dots, x_n \in \mathbb{R}$, on a $\sum_{i,j=1}^n x_i x_j m_{i+j} \geq 0$.

Donner une suite de réels qui n'est pas la suite des moments d'une variable aléatoire réelle.

Exercice II.12.5 Soit $X \sim \mathcal{N}(0, 1)$. Déterminer la loi de $\exp X$. Cette loi s'appelle la loi log-normale, car c'est la loi d'une variable aléatoire dont le logarithme suit une loi normale.

Exercice II.12.6 Soit (X, Y) un vecteur aléatoire sur \mathbb{R}^2 dont la loi admet la densité

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

Déterminer les lois de $X, Y, X + Y, X^2 + Y^2$.

Exercice II.12.7 a. Soient $\theta > 0$ un nombre réel et $k, l \geq 1$ deux nombres entiers. Déterminer l'unique réel c tel que la fonction

$$f_{(X,Y)}(x, y) = cx^{k-1}y^{l-1}e^{-\theta(x+y)}\mathbf{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(x, y)$$

soit la densité d'une mesure de probabilités sur \mathbb{R}^2 . Déterminer la loi de X . Cette loi s'appelle la loi Gamma de paramètres θ et k : on écrit $X \sim \Gamma(\theta, k)$. Quelle est cette loi lorsque $k = 1$?

b. Soit (X, Y) un vecteur aléatoire dont la loi admet la densité décrite à la question précédente. Déterminer la loi de $X + Y$.

Exercice II.12.8 Soient μ et $\sigma \geq 0$ deux nombres réels. Soit X une variable aléatoire de loi normale centrée réduite. Calculer la loi de $\sigma X + \mu$. En déduire l'espérance et la variance de la loi $\mathcal{N}(\mu, \sigma^2)$.

Exercice II.12.9 a. Soient X et Y deux variables aléatoires telles que $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{E}(1)$.

- Déterminer la loi de $\log |X|$.
- Déterminer la loi de X^4 .
- Déterminer la loi de $\lfloor Y \rfloor$, où pour tout réel x , $\lfloor x \rfloor$ est la partie entière de x , définie par $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$.
- Déterminer la loi de $\cos(Y)$.

Exercice II.12.10 Soit $\theta > 0$ un réel. Soit (X, Y) un vecteur aléatoire dont la loi admet la densité

$$f_{(X,Y)}(x, y) = \theta^2 e^{-\theta(x+y)} \mathbf{1}_{[0,+\infty[}(x) \mathbf{1}_{[0,+\infty[}(y).$$

Déterminer la loi du vecteur aléatoire $(X + Y, X - Y)$.

Exercice II.12.11 Soient p et q deux réels compris entre 0 et 1. Soit (X, Y) un vecteur aléatoire tel que X suive la loi de Bernoulli de paramètre p et Y la loi de Bernoulli de paramètre q . Montrer qu'on a

$$\max(p + q - 1, 0) - pq \leq \mathbf{cov}(X, Y) \leq \min(p, q) - pq$$

et que les deux bornes de cette égalité peuvent être atteintes.

Exercice II.12.12 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire sur \mathbb{R}^n dont la loi admet la densité

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}.$$

a. Soit $R : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une transformation orthogonale, c'est-à-dire une isométrie linéaire. Déterminer la loi de $R(X)$.

b. Montrer que la loi du vecteur aléatoire $Z = \frac{(X_1, \dots, X_n)}{\sqrt{X_1^2 + \dots + X_n^2}}$ est une mesure de probabilité sur \mathbb{R}^n qui donne une probabilité 1 à la sphère unité, et qui donne la même probabilité à deux parties qui diffèrent l'une de l'autre par une rotation.

On admettra qu'il existe une unique telle mesure de probabilités. On l'appelle la mesure uniforme sur la sphère unité de \mathbb{R}^n .

c. Déterminer la loi de la distance au plan de l'équateur d'un point choisi uniformément au hasard sur la Terre.

Chapitre III

Indépendance.

III.1 Indépendance d'événements et de classes d'événements.

III.1.a Indépendance d'événements (rappel).

On rappelle la définition suivante de l'indépendance mutuelle d'une famille d'événement.

Définition III.1.1 ◀▶ (Indépendance d'événements) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. On introduit les notions suivantes.

- (a) Soient $A, B \in \mathcal{F}$. Les événements A et B sont dits *indépendants sous \mathbf{P}* si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.
- (b) Soient $A_1, \dots, A_n \in \mathcal{F}$. Les événements A_1, \dots, A_n sont dits *mutuellement indépendants sous \mathbf{P}* si pour tout $2 \leq k \leq n$ et tous $1 \leq j_1 < \dots < j_k \leq n$, on

$$\mathbf{P}(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = \mathbf{P}(A_{j_1})\mathbf{P}(A_{j_2}) \dots \mathbf{P}(A_{j_k}) .$$

Il y a donc $2^n - 1 - n$ égalités à vérifier.

- (c) Soit $A_i \in \mathcal{F}, i \in I$, une famille d'événements. On dit qu'ils sont *mutuellement indépendants sous \mathbf{P}* si pour tout sous-ensemble d'indices $J \subset I$ qui est fini et non-vide, on a

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j) .$$

On remarque que les $(A_i)_{i \in I}$ sont mutuellement indépendants sous \mathbf{P} ssi si pour tout sous-ensemble d'indices $J \subset I$ qui est fini et qui compte au moins deux éléments, les événements $(A_j)_{j \in J}$ sont mutuellement indépendants sous \mathbf{P} au sens du (b). \square

III.1.b Indépendance de classes d'événements.

Définition III.1.2 ◀▶ (Indépendance de classe d'ensembles) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. On introduit les notions suivantes.

- (a) Soient $\mathcal{R}_1, \dots, \mathcal{R}_n \subset \mathcal{F}$, des classes non-vides d'événements. Les classes d'événements $\mathcal{R}_1, \dots, \mathcal{R}_n$ sont *mutuellement indépendantes sous \mathbf{P}* si

$$\forall A_1 \in \mathcal{R}_1, \dots, \forall A_n \in \mathcal{R}_n, \quad \text{les événements } A_1, \dots, A_n \text{ sont indépendants.}$$

- (b) Soit $\mathcal{R}_i \subset \mathcal{F}$, $i \in I$, une famille de classes non-vides d'événements. Les classes \mathcal{R}_i , $i \in I$, sont mutuellement indépendantes sous \mathbf{P} si pour tout sous-ensemble d'indices $J \subset I$, qui est fini et qui contient au moins deux éléments, les classes $(\mathcal{R}_j)_{j \in J}$ sont mutuellement indépendantes au sens du (b). \square

Au vu du point (c) de la définition I.2.2, les classes d'événements $(\mathcal{R}_i)_{i \in I}$ sont mutuellement indépendantes sous \mathbf{P} ssi

$$\forall J \subset I, J \text{ fini et non-vide}, \forall A_j \in \mathcal{R}_j, j \in J, \quad \mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j). \quad (\text{III.1})$$

La proposition suivante est une conséquence immédiate des définitions précédentes.

Proposition III.1.1 $\blacktriangleleft \blacktriangleright$ (« Qui peut le plus peut le moins ») Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $\mathcal{R}_i \subset \mathcal{F}$, $i \in I$, une famille de classes non-vides d'événements. Soit $I' \subset I$ et soient $\mathcal{R}'_i \subset \mathcal{R}_i$, $i \in I'$, des classes non-vides d'événements. On suppose que les classes d'événements \mathcal{R}_i , $i \in I$ sont mutuellement indépendantes sous \mathbf{P} . Alors il en est de même pour les classes d'événements \mathcal{R}'_i , $i \in I'$.

Lemme III.1.2 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $\mathcal{R}_1, \dots, \mathcal{R}_n \subset \mathcal{F}$, des classes non-vides d'événements. On suppose que $\Omega \in \mathcal{R}_k$ pour tout $1 \leq k \leq n$. Alors,

$$\mathcal{R}_1, \dots, \mathcal{R}_n \text{ indépendants} \iff \forall A_1 \in \mathcal{R}_1, \dots, \forall A_n \in \mathcal{R}_n, \mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \dots \mathbf{P}(A_n).$$

Preuve : l'implication \implies est claire. Montrons l'implication contraire. On suppose donc que

$$\forall A_1 \in \mathcal{R}_1, \dots, \forall A_n \in \mathcal{R}_n, \mathbf{P}(A_1 \cap \dots \cap A_n) = \mathbf{P}(A_1) \dots \mathbf{P}(A_n). \quad (\text{III.2})$$

Soient $B_1 \in \mathcal{R}_1, \dots, B_n \in \mathcal{R}_n$. Il faut montrer que les événements B_1, \dots, B_n sont mutuellement indépendants sous \mathbf{P} : soient $2 \leq k \leq n$ et $1 \leq j_1 < \dots < j_k \leq n$. Pour tout $1 \leq \ell \leq k$, on pose $A_{j_\ell} := B_{j_\ell}$ et si $j \notin \{j_1, \dots, j_k\}$, on pose $A_j := \Omega$, qui appartient à \mathcal{R}_j par hypothèse. Par (III.2), on a donc

$$\begin{aligned} \mathbf{P}(B_{j_1} \cap \dots \cap B_{j_k}) &= \mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= \mathbf{P}(A_1) \mathbf{P}(A_2) \dots \mathbf{P}(A_n) = \mathbf{P}(B_{j_1}) \dots \mathbf{P}(B_{j_k}). \end{aligned}$$

Cela implique que pour tous $B_1 \in \mathcal{R}_1, \dots, B_n \in \mathcal{R}_n$, les événements B_1, \dots, B_n sont mutuellement indépendants sous \mathbf{P} . Les classes d'événements $\mathcal{R}_1, \dots, \mathcal{R}_n$ sont donc mutuellement indépendantes sous \mathbf{P} . \blacksquare

Théorème III.1.3 $\blacktriangleleft \blacktriangleright$ (« L'indépendance de Pi-systèmes implique celle des tribus engendrées ») Soit $\mathcal{P}_i \subset \mathcal{F}$, $i \in I$, une famille de pi-systèmes. On suppose que les pi-systèmes \mathcal{P}_i , $i \in I$, sont mutuellement indépendants sous \mathbf{P} . Alors, les tribus engendrées $\sigma(\mathcal{P}_i)$, $i \in I$, sont également mutuellement indépendantes sous \mathbf{P} .

Preuve : au vu de la définition III.1.2 (b), il suffit de prouver le théorème dans le cas où I est fini. On choisit donc $I = \{1, \dots, n\}$. On montre d'abord l'assertion suivante.

- Pour tous pi-systèmes $\mathcal{P}_i^* \subset \mathcal{F}$, $1 \leq i \leq n$, on a l'implication suivante :

$$(\mathcal{P}_1^*, \dots, \mathcal{P}_n^* \text{ mut. ind. sous } \mathbf{P}) \implies (\sigma(\mathcal{P}_1^*), \mathcal{P}_2^*, \dots, \mathcal{P}_n^* \text{ mut. ind. sous } \mathbf{P}). \quad (\text{III.3})$$

En effet, on suppose $\mathcal{P}_1^*, \dots, \mathcal{P}_n^*$ indépendants, on fixe $A_2 \in \mathcal{P}_2^*, \dots, A_n \in \mathcal{P}_n^*$ et on pose

$$\mathcal{L} = \{B \in \sigma(\mathcal{P}_1^*) : \mathbf{P}(B \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(B)\mathbf{P}(A_2 \cap \dots \cap A_n)\}.$$

On a donc $\mathcal{P}_1^* \subset \mathcal{L}$. On montre facilement que \mathcal{L} est une classe monotone (exercice). Le théorème de la classe monotone implique alors que $\sigma(\mathcal{P}_1^*) \subset \mathcal{L}$ et donc que $\sigma(\mathcal{P}_1^*) = \mathcal{L}$. Donc, pour tout $B \in \sigma(\mathcal{P}_1^*)$, $\mathbf{P}(B \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(B)\mathbf{P}(A_2 \cap \dots \cap A_n)$. Mais comme $\mathcal{P}_2^*, \dots, \mathcal{P}_n^*$ sont indépendants, on a $\mathbf{P}(A_2 \cap \dots \cap A_n) = \mathbf{P}(A_2) \dots \mathbf{P}(A_n)$. On a donc montré que

$$\forall B \in \sigma(\mathcal{P}_1^*), \forall A_2 \in \mathcal{P}_2^*, \dots, \forall A_n \in \mathcal{P}_n^*, \quad \mathbf{P}(B \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(B)\mathbf{P}(A_2) \dots \mathbf{P}(A_n).$$

Le lemme III.2 implique ensuite que les classes d'événements $\sigma(\mathcal{P}_1^*), \mathcal{P}_2^*, \dots, \mathcal{P}_n^*$ sont mutuellement indépendantes sous \mathbf{P} , ce qui termine la preuve de (III.3).

En appliquant (III.3) à $(\mathcal{P}_1^*, \dots, \mathcal{P}_n^*) = (\mathcal{P}_1, \dots, \mathcal{P}_n)$, on montre que $\sigma(\mathcal{P}_1), \mathcal{P}_2, \dots, \mathcal{P}_n$ sont mutuellement indépendants sous \mathbf{P} . Or $\sigma(\mathcal{P}_1)$ est une tribu donc un pi-système. On applique alors (III.3) à la suite de pi-systèmes

$$(\mathcal{P}_1^*, \dots, \mathcal{P}_n^*) = (\mathcal{P}_2, \dots, \mathcal{P}_n, \sigma(\mathcal{P}_1)),$$

ce qui montre la mutuelle indépendance des $\sigma(\mathcal{P}_2), \mathcal{P}_3, \dots, \mathcal{P}_n, \sigma(\mathcal{P}_1)$. Ainsi de suite jusqu'à montrer que $\sigma(\mathcal{P}_1), \dots, \sigma(\mathcal{P}_n)$ sont mutuellement indépendantes sous \mathbf{P} , ce qui termine la preuve du théorème. ■

Proposition III.1.4 ◀▶ (Indépendance par paquets) Soit I un ensemble d'indices. Soit $I_k \subset I$, $k \in K$, une partition de I . Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $\mathcal{F}_i \subset \mathcal{F}$, $i \in I$, une famille de tribus. Pour tout $k \in K$, on pose $\mathcal{G}_k = \sigma(\mathcal{F}_i; i \in I_k)$. Alors,

$$\mathcal{F}_i, i \in I \text{ mut. ind. sous } \mathbf{P} \implies \mathcal{G}_k, k \in K \text{ mut. ind. sous } \mathbf{P}.$$

Preuve : pour tout $k \in K$, on note \mathcal{P}_k la classe des intersections finies d'événements dans $\bigcup_{i \in I_k} \mathcal{F}_i$:

$$\mathcal{P}_k = \{A_1 \cap \dots \cap A_n; n \in \mathbb{N}, A_1, \dots, A_n \in \bigcup_{i \in I_k} \mathcal{F}_i\}.$$

Il est facile de voir que \mathcal{P}_k est un pi-système tel que $\sigma(\mathcal{P}_k) = \mathcal{G}_k$. L'indépendance des tribus \mathcal{F}_i , $i \in I$, entraîne facilement celle des \mathcal{P}_k , $k \in K$. Le théorème III.1.3 permet de conclure. ■

III.2 Indépendance de variables aléatoires.

III.2.a Résultats généraux.

Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit (E, \mathcal{E}) , un espace mesurable et $X : \Omega \rightarrow E$, une v.a. $(\mathcal{F}, \mathcal{E})$ -mesurable. On rappelle que la tribu engendrée par X est donnée par la tribu pré-image de \mathcal{E} par X , c'est-à-dire

$$\sigma(X) = \{\{X \in B\}; B \in \mathcal{E}\} \quad (\text{III.4})$$

qui est la plus petite tribu sur Ω rendant X mesurable. On introduit la définition suivante.

Définition III.2.1 ◀▶ (Indépendance de v.a.) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit I , un ensemble d'indices. Pour tout $i \in I$, soit (E_i, \mathcal{E}_i) , un espace mesurable et $X_i : \Omega \rightarrow E_i$, une v.a. $(\mathcal{F}, \mathcal{E}_i)$ -mesurable.

On dit que les variables $X_i, i \in I$, sont mutuellement indépendantes sous \mathbf{P} si les tribus $\sigma(X_i) \subset \mathcal{F}, i \in I$, sont mutuellement indépendantes sous \mathbf{P} . \square

Comme conséquence facile de la définition, nous montrons si on prend des fonctions déterministes de v.a. indépendantes, on obtient des v.a. qui sont encore indépendantes.

Proposition III.2.1 ◀▶ («Des fonctions déterministes de variables indépendantes restent indépendantes») Soit I , un ensemble d'indices. Pour tout $i \in I$, soit (E_i, \mathcal{E}_i) , un espace mesurable et $X_i : \Omega \rightarrow E_i$, une v.a. $(\mathcal{F}, \mathcal{E}_i)$ -mesurable. Pour tout $i \in I$, soit (E'_i, \mathcal{E}'_i) et $f_i := E_i \rightarrow E'_i$, une fonction $(\mathcal{E}_i, \mathcal{E}'_i)$ -mesurable. On pose alors

$$\forall i \in I, \forall \omega \in \Omega, \quad X'_i(\omega) = f_i(X_i(\omega)) .$$

Si les variables aléatoires $(X_i)_{i \in I}$ sont mutuellement indépendantes sous \mathbf{P} , alors il en est de même pour les variables aléatoires $(X'_i)_{i \in I}$.

Preuve : par (III.4), on a $\sigma(X'_i) = \{\{X'_i \in B'\}; B' \in \mathcal{E}'_i\}$. Or

$$\begin{aligned} \{X'_i \in B'\} &= \{\omega \in \Omega : f_i(X_i(\omega)) \in B'\} \\ &= \{\omega \in \Omega : X_i(\omega) \in f_i^{-1}(B')\} \\ &= \{X_i \in f_i^{-1}(B')\} \in \sigma(X_i) \end{aligned}$$

car $f_i^{-1}(B') \in \mathcal{E}_i$ puisque $f_i := E_i \rightarrow E'_i$ est une fonction $(\mathcal{E}_i, \mathcal{E}'_i)$ -mesurable. On a donc montré que $\sigma(X'_i) \subset \sigma(X_i)$. Si les $(X_i)_{i \in I}$ sont mutuellement indépendantes sous \mathbf{P} , cela signifie que les tribus $\sigma(X_i), i \in I$ sont mutuellement indépendantes sous \mathbf{P} et la proposition III.1.1 « qui peut le plus peut le moins » (page 98) implique que les tribus $\sigma(X'_i), i \in I$ sont mutuellement indépendantes sous \mathbf{P} , ce qui signifie par définition que les variables $(X'_i)_{i \in I}$ sont mutuellement indépendantes sous \mathbf{P} . ■

La proposition suivante donne plusieurs reformulations de la définition de l'indépendance des variables aléatoires.

Proposition III.2.2 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit I , un ensemble d'indices. Pour tout $i \in I$, on fixe les objets suivants.

- Un espace mesurable (E_i, \mathcal{E}_i) et un pi-système $\mathcal{C}_i \subset \mathcal{E}_i$ tel que $\sigma(\mathcal{C}_i) = \mathcal{E}_i$.
- Une v.a. $X_i : \Omega \rightarrow E_i$ qui est $(\mathcal{F}, \mathcal{E}_i)$ -mesurable et dont la loi sous \mathbf{P} est notée $\mu_i : \mu_i(B) = \mathbf{P}(X_i \in B), B \in \mathcal{E}_i$.

Alors, les assertions suivantes sont équivalentes.

- (i) Les variables $X_i, i \in I$, sont mutuellement indépendantes sous \mathbf{P} .
- (ii) Pour tout $J \subset I$, fini non-vide et pour tous $B_j \in \mathcal{E}_j, j \in J$,

$$\mathbf{P}\left(\bigcap_{j \in J} \{X_j \in B_j\}\right) = \prod_{j \in J} \mathbf{P}(X_j \in B_j) .$$

- (iii) Pour tout $J \subset I$, fini non-vide et pour tous $C_j \in \mathcal{C}_j, j \in J$,

$$\mathbf{P}\left(\bigcap_{j \in J} \{X_j \in C_j\}\right) = \prod_{j \in J} \mathbf{P}(X_j \in C_j) .$$

(iv) Pour tout $J = \{j_1, \dots, j_n\} \subset I$, où les j_k sont distincts,

la loi de $(X_{j_1}, \dots, X_{j_n})$ sous \mathbf{P} est $\mu_{j_1} \otimes \dots \otimes \mu_{j_n}$.

(v) Pour tout $J \subset I$, fini non-vide et pour toutes fonctions \mathcal{E}_j -mesurables bornées $h_j : E_j \rightarrow \mathbb{R}$, $j \in J$,

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

Preuve : au vu des définitions III.2.1 et III.1.2, il suffit de traiter le cas où I est fini. On prend donc $I = \{1, \dots, n\}$. Par définition, (i) est équivalent à l'indépendance des tribus $\sigma(X_i)$, $i \in I$. Or (III.4) et (III.1) montrent que (i) \Leftrightarrow (ii). L'implication (ii) \Rightarrow (iii) est triviale. Supposons (iii) : on pose

$$\mathcal{P}_i = \{ \{X_i \in C\} ; C \in \mathcal{C}_i \} .$$

C'est clairement un pi-système tel que $\sigma(X_i) = \sigma(\mathcal{P}_i)$. Or (iii) signifie que les pi-systèmes \mathcal{P}_i , $i \in I$ sont indépendants : le théorème III.1.3 implique alors que les tribus $\sigma(\mathcal{P}_i)$, $i \in I$, sont indépendantes, ce qui implique (i). On a donc montré que (iii) \Rightarrow (i) et donc que (i), (ii) et (iii) sont équivalents.

On suppose (ii). On rappelle ensuite que $\mu_1 \otimes \dots \otimes \mu_n$ est l'unique mesure de probabilité sur $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ telle que $(\mu_1 \otimes \dots \otimes \mu_n)(B_1 \times \dots \times B_n) = \mu_1(B_1) \dots \mu_n(B_n)$, pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$. On note μ la loi de $X = (X_1, \dots, X_n)$. Il est clair que pour tous $B_1 \in \mathcal{E}_1, \dots, B_n \in \mathcal{E}_n$

$$\mu(B_1 \times \dots \times B_n) = \mathbf{P}(X \in B_1 \times \dots \times B_n) = \mathbf{P} \left(\bigcap_{1 \leq j \leq n} \{X_j \in B_j\} \right).$$

Or, par (ii), $\mathbf{P}(\bigcap_{1 \leq j \leq n} \{X_j \in B_j\}) = \prod_{1 \leq j \leq n} \mathbf{P}(X_j \in B_j) = \mu_1(B_1) \dots \mu_n(B_n)$. Cela entraîne donc $\mu = \mu_1 \otimes \dots \otimes \mu_n$, et donc (iv). On a montré que (ii) \Rightarrow (iv). L'implication (iv) \Rightarrow (v) est une conséquence directe du théorème de Fubini. L'implication (v) \Rightarrow (ii) est triviale. ■

Pour des variables aléatoires réelles, on obtient le corollaire suivant

Corollaire III.2.3 $\blacktriangleleft \blacktriangleright$ Soit $(X_i)_{i \in I}$, une famille de v.a. réelles \mathcal{F} -mesurables. Elles sont indépendantes ssi pour tout $\{j_1, \dots, j_n\} \subset I$ où les j_k sont distincts, et pour tous $x_1, \dots, x_n \in \mathbb{R}$, on ait

$$\mathbf{P}(X_{j_1} \leq x_1; \dots; X_{j_n} \leq x_n) = \mathbf{P}(X_{j_1} \leq x_1) \dots \mathbf{P}(X_{j_n} \leq x_n) .$$

Preuve : on applique l'équivalence de (i) et de (iii) de la proposition III.2.2 à $(E_i, \mathcal{E}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\mathcal{C}_i = \{\mathbb{R}\} \cup \{]-\infty, a] ; a \in \mathbb{R} \}$, qui est un pi-système générant $\mathcal{B}(\mathbb{R})$. ■

On déduit immédiatement du corollaire précédent le résultat suivant sur le maximum et le minimum de variables indépendantes.

Proposition III.2.4 $\blacktriangleleft \blacktriangleright$ Soient X_1, \dots, X_n , des v.a. réelles indépendantes. On pose

$$X^* = \max_{1 \leq k \leq n} X_k \quad \text{et} \quad X_* = \min_{1 \leq k \leq n} X_k .$$

Les fonctions de répartition de X^* et de X_* sont données par

$$\forall x \in \mathbb{R}, \quad F_{X^*}(x) = \prod_{1 \leq k \leq n} F_{X_k}(x) \quad \text{et} \quad F_{X_*}(x) = 1 - \prod_{1 \leq k \leq n} (1 - F_{X_k}(x)) .$$

III.2.b Variables indépendantes de loi diffuse, statistique d'ordre.

Nous montrons d'abord un résultat général montrant que des variables indépendantes de loi diffuse prennent p.s. des valeurs distinctes.

Plus précisément, on se place sur \mathbb{R}^2 , muni de la topologie de la norme et de la tribu $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R})^{\otimes 2}$ correspondante. On note $\Delta = \{(x, x); x \in \mathbb{R}\}$ qui est la diagonale du plan \mathbb{R}^2 . C'est un ensemble fermé donc un borélien de \mathbb{R}^2 . Soient μ_1 et μ_2 deux mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Par Fubini on observe the following

$$\mu_1 \otimes \mu_2(\Delta) = \int_{\mathbb{R}} \mu_2(dx_2) \left(\int_{\mathbb{R}} \mu_1(dx_1) \mathbf{1}_{\Delta}(x_1, x_2) \right).$$

Si on fixe $x_2 \in \mathbb{R}$, alors $\mathbf{1}_{\Delta}(x_1, x_2) = 0$ si $x_1 \neq x_2$ et $\mathbf{1}_{\Delta}(x_1, x_2) = 1$ si $x_1 = x_2$. Par conséquent, $\mathbf{1}_{\Delta}(x_1, x_2) = \mathbf{1}_{\{x_2\}}(x_1)$ et donc

$$\mu_1 \otimes \mu_2(\Delta) = \int_{\mathbb{R}} \mu_2(dx_2) \left(\int_{\mathbb{R}} \mu_1(dx_1) \mathbf{1}_{\{x_2\}}(x_1) \right) = \int_{\mathbb{R}} \mu_2(dx_2) \mu_1(\{x_2\}).$$

De même en l'autre coordonnée :

$$\mu_1 \otimes \mu_2(\Delta) = \int_{\mathbb{R}} \mu_2(dx_2) \mu_1(\{x_2\}) = \int_{\mathbb{R}} \mu_1(dx_1) \mu_2(\{x_1\}).$$

Cela montre le fait suivant.

$$\text{Si } \mu_1 \text{ ou } \mu_2 \text{ est diffuse alors } \mu_1 \otimes \mu_2(\Delta) = 0. \quad (\text{III.5})$$

Cela implique le résultat suivant.

Lemme III.2.5 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $X, Y : \Omega \rightarrow \mathbb{R}$, deux v.a. \mathcal{F} -mesurables supposées **indépendantes et dont les lois sont diffuses**. Alors, $\mathbf{P}(X = Y) = 0$.

Plus généralement, soit $X_n : \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, une suite de v.a. \mathcal{F} -mesurables supposées **indépendantes et dont les lois sont diffuses**. Alors,

$$\mathbf{P}(\exists m, n \in \mathbb{N} : m \neq n \text{ et } X_m = X_n) = 0.$$

Preuve : on remarque que $\{X = Y\} = \{(X, Y) \in \Delta\}$, où Δ désigne la diagonale du plan \mathbb{R}^2 . On note μ_1 la loi de X et μ_2 celle de Y . Par la proposition III.2.2 (i) \Leftrightarrow (iv) (voir page 100), puisque les variables sont indépendantes, la loi de $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ est $\mu_1 \otimes \mu_2$. Comme μ_1 et μ_2 sont supposées diffuses, on a

$$\mathbf{P}(X = Y) = \mathbf{P}((X, Y) \in \Delta) = \mu_1 \otimes \mu_2(\Delta) = 0,$$

par (III.5). Cela prouve la première partie du lemme.

On remarque ensuite que

$$A := \{\exists m, n \in \mathbb{N} : m \neq n \text{ et } X_m = X_n\} = \bigcup_{m, n \in \mathbb{N} : m < n} \{X_m = X_n\}.$$

Par sigma sous-additivité, $\mathbf{P}(A) \leq \sum_{m, n \in \mathbb{N} : m < n} \mathbf{P}(X_m = X_n) = 0$, ce qui termine la preuve. ■

Statistique d'ordre de n v.a. réelles indépendantes de lois diffuses. Soit $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. On note $\Lambda_n(\mathbf{x}) = (x_{(k)})_{1 \leq k \leq n}$ le réarrangement croissant des réels (x_1, \dots, x_n) , c'est-à-dire que

$$x_{(1)} \leq \dots \leq x_{(k)} \leq x_{(k+1)} \leq \dots \leq x_{(n)} \quad \text{et} \quad \{x_k; 1 \leq k \leq n\} = \{x_{(k)}; 1 \leq k \leq n\}.$$

Lemme III.2.6 *L'application $\Lambda_n : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est mesurable.*

Preuve : pour tout $y \in \mathbb{R}$ on pose $\phi_y(x_1, \dots, x_n) = \sum_{1 \leq j \leq n} \mathbf{1}_{]-\infty, y]}(x_j)$. La fonction $\phi_y : \mathbb{R}^n \rightarrow \mathbb{N}$ est clairement mesurable. Pour tout $y_1, \dots, y_n \in \mathbb{R}$, on pose $C_{y_1, \dots, y_n} = \prod_{1 \leq k \leq n}]-\infty, y_k]$. On remarque alors que

$$\Lambda_n^{-1}(C_{y_1, \dots, y_n}) = \bigcap_{1 \leq k \leq n} \{\phi_{y_k} \geq k\} \in \mathcal{B}(\mathbb{R}^n).$$

Or il est facile de montrer que $\mathcal{B}(\mathbb{R}^n)$ est engendrée par les ensembles du type C_{y_1, \dots, y_n} , $y_1, \dots, y_n \in \mathbb{R}$, ce qui implique la mesurabilité de Λ_n . ■

Cela permet de poser la définition suivante.

Définition III.2.2 (*Statistique d'ordre*) Soient X_1, \dots, X_n , des v.a. réelles \mathcal{F} -mesurables. On note $(X_{(k)})_{1 \leq k \leq n}$ le réarrangement croissant de ces variables :

$$X_{(1)} \leq \dots \leq X_{(k)} \leq X_{(k+1)} \leq \dots \leq X_{(n)} \quad \text{et} \quad \{X_k; 1 \leq k \leq n\} = \{X_{(k)}; 1 \leq k \leq n\}.$$

La v.a. $(X_{(k)})_{1 \leq k \leq n} : \Omega \rightarrow \mathbb{R}^n$ est \mathcal{F} -mesurable : c'est *statistique d'ordre de la suite* $(X_k)_{1 \leq k \leq n}$. □

La proposition suivante décrit la statistique d'ordre de variables indépendantes de loi diffuse.

Proposition III.2.7 *Soit X_1, \dots, X_n , des variables indépendantes de même loi μ qui est supposée diffuse. Alors les assertions suivantes sont vérifiées.*

(i) *On note \mathbf{S}_n le groupe des permutations de $\{1, \dots, n\}$. Il existe $\sigma : \Omega \rightarrow \mathbf{S}_n$, une permutation aléatoire \mathcal{F} -mesurable de loi uniforme indépendante de $(X_{(k)})_{1 \leq k \leq n}$ telle que*

$$\mathbf{P}\text{-p.s.} \quad X_{(k)} = X_{\sigma(k)}, \quad 1 \leq k \leq n. \quad (\text{III.6})$$

(ii) *Pour toute fonction mesurable $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$, on a*

$$\mathbf{E}[f(X_{(1)}, \dots, X_{(n)})] = n! \int_{\{x_1 < \dots < x_n\}} f(x_1, \dots, x_n) \mu(dx_1) \dots \mu(dx_n).$$

(iii) *On note $F(x) = \mu(]-\infty, x])$, $x \in \mathbb{R}$, la fonction de répartition de μ . Alors, pour tout $1 \leq k \leq n$ et pour toute fonction mesurable $g : \mathbb{R} \rightarrow \mathbb{R}_+$,*

$$\mathbf{E}[g(X_{(k)})] = \int_{\mathbb{R}} g(x) k \binom{n}{k} F(x)^{k-1} (1-F(x))^{n-k} \mu(dx).$$

Preuve : on pose $A = \{X_j \neq X_k, 1 \leq j < k \leq n\}$, qui l'événement que les variables X_1, \dots, X_n soient toutes distinctes. Comme μ est diffuse le lemme III.2.5 implique que $\mathbf{P}(A) = 1$. Pour toute permutation $\gamma \in \mathbf{S}_n$, on pose

$$A_\gamma = \{X_{\gamma(1)} < X_{\gamma(2)} < \dots < X_{\gamma(n)}\}.$$

Il est clair que les événements $A_\gamma, \gamma \in \mathbf{S}_n$ sont deux-à-deux disjoints et que $A = \bigcup_{\gamma \in \mathbf{S}_n} A_\gamma$. On définit alors la permutation aléatoire σ en posant $\sigma = \gamma$ sur A_γ et $\sigma = \text{Id}$ sur $\Omega \setminus A$. Comme $\mathbf{P}(A) = 1$, σ satisfait bien (III.6).

On observe ensuite que pour toute permutation $\gamma \in \mathbf{S}_n$, les v.a. $(X_{\gamma(k)})_{1 \leq k \leq n}$ sont indépendantes et de même loi μ . Donc

$$(X_{\gamma(1)}, X_{\gamma(2)}, \dots, X_{\gamma(n)}) \stackrel{\text{loi}}{=} (X_1, X_2, \dots, X_n).$$

On en déduit alors que pour toute fonction mesurable $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ et toute permutation $\gamma \in \mathbf{S}_n$,

$$\begin{aligned} \mathbf{E}[f(X_{(1)}, \dots, X_{(n)}) \mathbf{1}_{\{\sigma=\gamma\}}] &= \mathbf{E}[f(X_{\gamma(1)}, \dots, X_{\gamma(n)}) \mathbf{1}_{A_\gamma}] \\ &= \mathbf{E}[f(X_1, \dots, X_n) \mathbf{1}_{\{X_1 < X_2 < \dots < X_n\}}] \\ &= \int_{\{x_1 < \dots < x_n\}} f(x_1, \dots, x_n) \mu(dx_1) \dots \mu(dx_n). \end{aligned} \quad (\text{III.7})$$

En sommant sur $\gamma \in \mathbf{S}_n$, cela montre (ii). On en déduit que $\int_{\{x_1 < \dots < x_n\}} \mu(dx_1) \dots \mu(dx_n) = 1/n!$ et (III.7) implique d'une part que σ est de loi uniforme sur \mathbf{S}_n et d'autre part que

$$\mathbf{E}[f(X_{(1)}, \dots, X_{(n)}) \mathbf{1}_{\{\sigma=\gamma\}}] = \mathbf{E}[f(X_{(1)}, \dots, X_{(n)})] \mathbf{P}(\sigma = \gamma),$$

ce qui implique que σ est indépendante de $(X_{(k)})_{1 \leq k \leq n}$, ce qui termine la preuve de (i).

Montrons (iii) : en raisonnant comme précédemment on montre que pour tout $1 \leq k \leq n$ et tout $x \in \mathbb{R}$,

$$\int_{\{x_1 < \dots < x_{k-1} < x\}} \mu(dx_1) \dots \mu(dx_{k-1}) = \frac{1}{(k-1)!} \int_{]-\infty, x[}^{[k-1]} \mu(dx_1) \dots \mu(dx_{k-1}) = \frac{F(x)^{k-1}}{(k-1)!}.$$

De même, on a

$$\int_{\{x < x_{k+1} < \dots < x_n\}} \mu(dx_{k+1}) \dots \mu(dx_n) = \frac{1}{(n-k)!} \int_{]x, \infty[}^{[n-k]} \mu(dx_{k+1}) \dots \mu(dx_n) = \frac{(1-F(x))^{n-k}}{(n-k)!}.$$

Par (ii) et Fubini, on a donc

$$\begin{aligned} \mathbf{E}[g(X_{(k)})] &= n! \int_{\mathbb{R}} \mu(dx) g(x) \left(\int_{\{x_1 < \dots < x_{k-1} < x\}} \mu(dx_1) \dots \mu(dx_{k-1}) \right) \left(\int_{\{x < x_{k+1} < \dots < x_n\}} \mu(dx_{k+1}) \dots \mu(dx_n) \right) \\ &= \frac{n!}{(k-1)!(n-k)!} \int_{\mathbb{R}} \mu(dx) g(x) F(x)^{k-1} (1-F(x))^{n-k}, \end{aligned}$$

ce qui entraîne (iii). ■

Application au lemme dit « des réveils ». On montre la proposition suivante qui est une propriété très utile des variables exponentielles connue sous le nom de « lemme des réveils ».

Lemme III.2.8 (Lemme des réveils) Soient X_1, \dots, X_n , des variables exponentielles de paramètres respectifs c_1, \dots, c_n . On pose $Y = \min_{1 \leq k \leq n} X_k$. Alors les assertions suivantes sont vérifiées.

- (i) Y est une v.a. exponentielle de paramètre $c_1 + \dots + c_n$.
- (ii) Il existe $N : \Omega \rightarrow \{1, \dots, n\}$, une v.a. \mathcal{F} -mesurable telle que

$$\mathbf{P}\text{-p.s. } X_N = Y \quad \text{et} \quad X_N < \min \{X_k ; k \in \{1, \dots, n\} \setminus \{N\}\} .$$

Par ailleurs

$$\forall k \in \{1, \dots, n\}, \quad \mathbf{P}(N = k) = \frac{c_k}{c_1 + \dots + c_n} .$$

- (iii) N et Y sont indépendantes.

Preuve : on remarque d'abord que pour tout $t \in \mathbb{R}_+$,

$$\mathbf{P}(Y > t) = \mathbf{P}(X_1 > t; \dots; X_n > t) = \mathbf{P}(X_1 > t) \dots \mathbf{P}(X_n > t) = e^{-(c_1 + \dots + c_n)t} ,$$

ce qui implique (i). Comme les v.a. X_k sont indépendantes et à densité, le lemme III.2.5 implique qu'elles sont p.s. distinctes : cela implique l'existence de N comme au (ii). Soit $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, une fonction mesurable et $k \in \{1, \dots, n\}$. On pose $c = c_1 + \dots + c_n$ et $c^* = c - c_k$. Le théorème de transfert implique le résultat suivant :

$$\begin{aligned} \mathbf{E}[f(Y)\mathbf{1}_{\{N=k\}}] &= \mathbf{E}[f(X_k)\mathbf{1}_{\{\forall j \in \{1, \dots, n\} \setminus \{k\}, X_k < X_j\}}] \\ &= \int_0^\infty c_k e^{-c_k t} f(t) e^{-c^* t} dt = \frac{c_k}{c} \int_0^\infty c e^{-c t} f(t) dt \end{aligned}$$

En prenant f constante à 1 on voit que $\mathbf{P}(N = k) = c_k/c$, ce qui complète la preuve de (ii). De plus on a $\mathbf{E}[f(Y)\mathbf{1}_{\{N=k\}}] = \mathbf{E}[f(Y)]\mathbf{P}(N = k)$. Pour toute fonction $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, on a

$$\begin{aligned} \mathbf{E}[f(Y)g(N)] &= \sum_{1 \leq k \leq n} g(k)[f(Y)\mathbf{1}_{\{N=k\}}] = \sum_{1 \leq k \leq n} g(k)\mathbf{E}[f(Y)]\mathbf{P}(N = k) \\ &= \mathbf{E}[f(Y)] \sum_{1 \leq k \leq n} g(k)\mathbf{P}(N = k) = \mathbf{E}[f(Y)]\mathbf{E}[g(N)] , \end{aligned}$$

ce qui montre que Y et N sont indépendantes. ■

III.2.c Indépendance, espérance et covariance.

Si deux variables réelles X_1 et X_2 sont \mathbf{P} -intégrables, c'est-à-dire si $\mathbf{E}[|X_1|]$ et $\mathbf{E}[|X_2|]$ sont des quantités finies, cela n'est pas forcément le cas du produit $X_1 X_2$. En revanche, lorsque les variables sont indépendantes, cela est vrai comme le montre la proposition suivante.

Donnons une conséquence très utile de la proposition III.2.2.

Proposition III.2.9 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient X_1, \dots, X_n , des v.a. complexes \mathcal{F} -mesurables intégrables indépendantes. Alors le produit $X_1 \dots X_n$ est également intégrable et

$$\mathbf{E}[X_1 \dots X_n] = \mathbf{E}[X_1] \dots \mathbf{E}[X_n] . \tag{III.8}$$

De même si $X_1, \dots, X_n : \Omega \rightarrow [0, \infty]$ sont \mathcal{F} -mesurables indépendantes, on a (III.8).

Preuve : par la proposition III.2.2 la loi de $X = (X_1, \dots, X_n)$ est le produit des lois des X_i : $\mu_X = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$. Par le théorème de transfert et Fubini positif

$$\begin{aligned} \mathbf{E}[|X_1| \dots |X_n|] &= \int_{\mathbb{R}^n} |x_1| \dots |x_n| \mu_{X_1}(dx_1) \dots \mu_{X_n}(dx_n) \\ &= \left(\int_{\mathbb{R}} |x_1| \mu_{X_1}(dx_1) \right) \dots \left(\int_{\mathbb{R}} |x_n| \mu_{X_n}(dx_n) \right) \\ &= \mathbf{E}[|X_1|] \dots \mathbf{E}[|X_n|] . \end{aligned}$$

Cela démontre le résultat voulu lorsque les variables sont à valeurs dans $[0, \infty]$. Si les variables X_1, \dots, X_n sont intégrables, il en est de même pour $X_1 \dots X_n$ et par transfert et Fubini, on a

$$\begin{aligned} \mathbf{E}[X_1 \dots X_n] &= \int_{\mathbb{R}^n} x_1 \dots x_n \mu_{X_1}(dx_1) \dots \mu_{X_n}(dx_n) \\ &= \left(\int_{\mathbb{R}} x_1 \mu_{X_1}(dx_1) \right) \dots \left(\int_{\mathbb{R}} x_n \mu_{X_n}(dx_n) \right) \\ &= \mathbf{E}[X_1] \dots \mathbf{E}[X_n] , \end{aligned}$$

qui est le résultat voulu. ■

On en déduit le résultat suivant.

Lemme III.2.10 ◀▶ Soient X, Y , deux v.a. réelles \mathcal{F} -mesurables admettant un moment d'ordre deux. On les suppose indépendantes. Alors, $\mathbf{cov}(X, Y) = 0$.

◀▶ **Preuve :** $\mathbf{cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0$ si X et Y sont indépendantes par la proposition III.2.9. ■

Exemple III.2.1 Soit X , une v.a. \mathcal{F} -mesurable de loi uniforme sur $[0, 1]$. Soit $\xi : \Omega \rightarrow \{-1, 1\}$, une v.a. \mathcal{F} -mesurable telle que $\mathbf{P}(\xi = 1) = \mathbf{P}(\xi = -1) = 1/2$. On suppose que X et ξ sont indépendantes. On pose alors $Y = \xi X$. On remarque que $\mathbf{P}(Y > 1/2) = \mathbf{P}(\xi = 1; X > 1/2) = \mathbf{P}(\xi = 1)\mathbf{P}(X > 1/2) = 1/4$. On a donc

$$\mathbf{P}(X > 1/2; Y > 1/2) = \mathbf{P}(\xi = 1; X > 1/2) = 1/4 \neq 1/8 = \mathbf{P}(X > 1/2)\mathbf{P}(Y > 1/2) .$$

On en déduit que X et Y ne sont pas indépendantes. Or $\mathbf{E}[Y] = \mathbf{E}[\xi]\mathbf{E}[X] = 0$ et $\mathbf{E}[XY] = \mathbf{E}[\xi X^2] = \mathbf{E}[\xi]\mathbf{E}[X^2] = 0$. Donc $\mathbf{cov}(X, Y) = 0$. La réciproque du lemme III.2.10 est donc fautive en général. □

Proposition III.2.11 ◀▶ Soient X_1, \dots, X_n , des v.a. réelles \mathcal{F} -mesurables admettant chacune un moment d'ordre deux. On suppose $\mathbf{cov}(X_i, X_j) = 0$, pour tous $1 \leq i < j \leq n$. Alors,

$$\mathbf{var}(X_1 + \dots + X_n) = \mathbf{var}(X_1) + \dots + \mathbf{var}(X_n) .$$

◀▶ **Preuve :** $\mathbf{cov}(X_i, X_j) = 0$ implique que $\mathbf{E}[X_i X_j] = \mathbf{E}[X_i]\mathbf{E}[X_j]$. On rappelle également que $\mathbf{var}(Y) = \mathbf{E}[Y^2] - (\mathbf{E}[Y])^2$, pour toute v.a. Y admettant un moment d'ordre deux. En développant le carré suivant, on obtient :

$$\mathbf{E}[(X_1 + \dots + X_n)^2] = \sum_{1 \leq i \leq n} \mathbf{E}[X_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbf{E}[X_i X_j] = \sum_{1 \leq i \leq n} \mathbf{E}[X_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbf{E}[X_i]\mathbf{E}[X_j]$$

$$\begin{aligned}
&= \sum_{1 \leq i \leq n} \text{var}(X_i) + \sum_{1 \leq i \leq n} \mathbf{E}[X_i]^2 + 2 \sum_{1 \leq i < j \leq n} \mathbf{E}[X_i] \mathbf{E}[X_j] \\
&= \sum_{1 \leq i \leq n} \text{var}(X_i) + (\mathbf{E}[X_1 + \dots + X_n])^2,
\end{aligned}$$

ce qui implique le résultat voulu. ■

Le lemme III.2.10 et la proposition III.2.11 qui précède impliquent immédiatement la proposition suivante.

Proposition III.2.12 ◀▶ Soient X_1, \dots, X_n , des v.a. réelles \mathcal{F} -mesurables admettant chacune un moment d'ordre deux. On les suppose deux-à-deux indépendantes. Alors,

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

III.2.d Convolution ; loi de sommes de variables indépendantes

Rappel sur le produit de convolution des mesures.

Définition III.2.3 ◀▶ Soient $\mu, \nu : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$, deux mesures de probabilité sur \mathbb{R}^n . La *convolution* de μ et de ν , notée $\mu * \nu$ est la mesure π sur \mathbb{R}^n qui est la mesure image par l'application $\text{Add} : (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbf{x} + \mathbf{y}$, c'est à-dire que pour toute fonction $f : \mathbb{R}^n \rightarrow [0, \infty]$ mesurable, on a

$$\int_{\mathbb{R}^n} f d\pi = \int_{\mathbb{R}^n} d\mu(\mathbf{x}) \int_{\mathbb{R}^n} d\nu(\mathbf{y}) f(\mathbf{x} + \mathbf{y})$$

. On utilise la notation $\pi = \mu * \nu$ pour désigner la convolution de μ avec ν . □

Soient $\mu, \nu, \pi : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$, trois mesures de probabilité sur \mathbb{R}^n . On a les propriétés suivantes.

• Le produit de convolution est commutatif : $\mu * \nu = \nu * \mu$. En effet,

$$\int_{\mathbb{R}^n} f d\mu * \nu = \int_{\mathbb{R}^n} d\mu(\mathbf{x}) \int_{\mathbb{R}^n} d\nu(\mathbf{y}) f(\mathbf{x} + \mathbf{y}) = \int_{\mathbb{R}^n} d\nu(\mathbf{y}) \int_{\mathbb{R}^n} d\mu(\mathbf{x}) f(\mathbf{x} + \mathbf{y})$$

par Fubini pour la seconde égalité. □

• La masse de Dirac en 0 est élément neutre pour $*$: $\mu * \delta_0 = \mu$. En effet, pour tout $\mathbf{x} \in \mathbb{R}^n$ fixé, on a $\int_{\mathbb{R}^n} \delta_0(d\mathbf{y}) f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x})$. Donc

$$\int_{\mathbb{R}^n} f d\mu * \delta_0 = \int_{\mathbb{R}^n} \mu(d\mathbf{x}) \int_{\mathbb{R}^n} \delta_0(d\mathbf{y}) f(\mathbf{x} + \mathbf{y}) = \int_{\mathbb{R}^n} f d\mu$$

pour toute fonction mesurable $f : \mathbb{R}^n \rightarrow [0, \infty]$, ce qui implique bien que $\mu * \delta_0 = \mu$. □

• Le produit de convolution de deux probabilité est une probabilité : $(\mu * \nu)(\mathbb{R}^n) = 1$. En effet,

$$\int_{\mathbb{R}^n} f d\mu * \nu = \int_{\mathbb{R}^n} d\mu(\mathbf{x}) \int_{\mathbb{R}^n} d\nu(\mathbf{y}) \mathbf{1} = \int_{\mathbb{R}^n} d\mu(\mathbf{x}) \nu(\mathbb{R}^n) = \mu(\mathbb{R}^n) \nu(\mathbb{R}^n)$$

ce qui implique bien que $(\mu * \nu)(\mathbb{R}^n) = 1$. □

• Le produit de convolution est associatif : $(\mu * \nu) * \pi = \mu * (\nu * \pi)$. La vérification est facile mais longue : nous laissons ce détail aux lecteurs. □

Exemple III.2.2 (*Convolution de loi sur \mathbb{Z}*) Lorsque l'on considère des lois discrètes la convolution est plus concrète. Soit μ une loi de probabilité sur \mathbb{Z} , l'ensemble des entiers relatifs. La mesure est caractérisée par les quantités $\mu(\{p\})$, $p \in \mathbb{Z}$: il s'agit en effet d'une mesure purement atomique qui s'écrit

$$\mu = \sum_{p \in \mathbb{Z}} \mu(\{p\}) \delta_p .$$

Soit ν , une autre mesure de probabilité sur \mathbb{Z} . Alors il est facile de vérifier que

$$\forall p \in \mathbb{Z}, \quad (\mu * \nu)(\{p\}) = \sum_{k \in \mathbb{Z}} \mu(\{p - k\}) \nu(\{k\}) . \quad (\text{III.9})$$

On remarque qu'il s'agit du produit de Cauchy des suites $(\mu(\{p\}))_{p \in \mathbb{Z}}$ et $(\nu(\{p\}))_{p \in \mathbb{Z}}$ (ce qui s'explique très bien plus loin). On voit que (III.9) définit une nouvelle probabilité sur \mathbb{Z} . \square

La proposition suivante montre la transformée de Fourier est particulièrement bien adaptée au produit de convolution.

Proposition III.2.13 $\blacktriangleleft \blacktriangleright$ Soient $\mu, \nu : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$, deux mesures de probabilités sur \mathbb{R}^n . Alors on a

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{\mu * \nu}(\mathbf{u}) = \widehat{\mu}(\mathbf{u}) \widehat{\nu}(\mathbf{u}) .$$

Preuve : on utilise le théorème de Fubini et au fait que $e^{i\langle \mathbf{u}, \mathbf{x} + \mathbf{y} \rangle} = e^{i\langle \mathbf{u}, \mathbf{x} \rangle} e^{i\langle \mathbf{u}, \mathbf{y} \rangle}$:

$$\begin{aligned} \widehat{\mu * \nu}(\mathbf{u}) &= \int_{\mathbb{R}^n} \mu(d\mathbf{x}) \int_{\mathbb{R}^n} \nu(d\mathbf{y}) e^{i\langle \mathbf{u}, \mathbf{x} + \mathbf{y} \rangle} \\ &= \int_{\mathbb{R}^n} \mu(d\mathbf{x}) \int_{\mathbb{R}^n} \nu(d\mathbf{y}) e^{i\langle \mathbf{u}, \mathbf{x} \rangle} e^{i\langle \mathbf{u}, \mathbf{y} \rangle} \\ &= \int_{\mathbb{R}^n} \mu(d\mathbf{x}) e^{i\langle \mathbf{u}, \mathbf{x} \rangle} \int_{\mathbb{R}^n} \nu(d\mathbf{y}) e^{i\langle \mathbf{u}, \mathbf{y} \rangle} \\ &= \widehat{\mu}(\mathbf{u}) \widehat{\nu}(\mathbf{u}) , \end{aligned}$$

qui est bien le résultat voulu. \blacksquare

Un récurrence facile (laissée au lecteur) utilisant la proposition précédente entraîne la proposition suivante.

Proposition III.2.14 $\blacktriangleleft \blacktriangleright$ Soient $\mu_k : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$, $1 \leq k \leq p$ des mesures de probabilités sur \mathbb{R}^n . On pose $\mu := \mu_1 * \dots * \mu_p$. Alors,

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{\mu}(\mathbf{u}) = \widehat{\mu}_1(\mathbf{u}) \dots \widehat{\mu}_p(\mathbf{u}) .$$

Interprétation probabiliste du produit de convolution.

Proposition III.2.15 $\blacktriangleleft \blacktriangleright$ Soient $\mathbf{X}, \mathbf{Y} : \Omega \rightarrow \mathbb{R}^n$, deux vecteurs aléatoires \mathcal{F} -mesurables. On les suppose *indépendants*. Alors la loi de $\mathbf{X} + \mathbf{Y}$ est la convolution des lois de \mathbf{X} et de \mathbf{Y} :

$$\mu_{\mathbf{X} + \mathbf{Y}} = \mu_{\mathbf{X}} * \mu_{\mathbf{Y}} .$$

Preuve : soit $f : \mathbb{R}^n \rightarrow [0, \infty]$, une fonction mesurable. Par la proposition III.2.2 (i) \Leftrightarrow (iv) (page 100), la loi du vecteur aléatoire de \mathbb{R}^{2n} donné par (\mathbf{X}, \mathbf{Y}) est la loi produit $\mu_{\mathbf{X}} \otimes \mu_{\mathbf{Y}}$. Par conséquent, pour toute fonction mesurable $f : \mathbb{R}^n \rightarrow [0, \infty]$, le théorème de transfert et la définition de la convolution de deux mesures de probabilité impliquent ce qui suit :

$$\begin{aligned} \mathbf{E}[f(\mathbf{X} + \mathbf{Y})] &= \int_{\mathbb{R}^n \times \mathbb{R}^n} f(\mathbf{x} + \mathbf{y}) \mu_{\mathbf{X}}(d\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) \\ &= \int_{\mathbb{R}^n} f d\mu_{\mathbf{X}} * \mu_{\mathbf{Y}} , \end{aligned}$$

ce qui implique bien que la loi de $\mathbf{X} + \mathbf{Y}$ est $\mu_{\mathbf{X} + \mathbf{Y}}$. ■

Un récurrence très simple implique le résultat suivant.

Proposition III.2.16 $\blacktriangleleft \blacktriangleright$ Soient $\mathbf{X}_k : \Omega \rightarrow \mathbb{R}^n$, $1 \leq k \leq p$, des vecteurs aléatoires \mathcal{F} -mesurables que l'on suppose **indépendants**. On note μ_1, \dots, μ_p leurs lois respectives. Alors, la loi μ de $\mathbf{X} := \mathbf{X}_1 + \dots + \mathbf{X}_p$ est $\mu_1 * \dots * \mu_p$

Les fonctions caractéristiques sont plus adaptées aux sommes de variables indépendantes.

Proposition III.2.17 $\blacktriangleleft \blacktriangleright$ Soient $\mathbf{X}_k : \Omega \rightarrow \mathbb{R}^n$, $1 \leq k \leq p$, des vecteurs aléatoires \mathcal{F} -mesurables que l'on suppose **indépendants**. On note μ_1, \dots, μ_p leurs lois respectives. On pose

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_p .$$

Alors,

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{\mu}_{\mathbf{X}}(\mathbf{u}) = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle}] \dots \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X}_p \rangle}] = \widehat{\mu}_1(\mathbf{u}) \dots \widehat{\mu}_p(\mathbf{u}) .$$

Preuve : on fixe $\mathbf{u} \in \mathbb{R}^n$. On utilise le principe que des fonctions déterministes de v.a. indépendantes restent indépendantes (c'est-à-dire la proposition III.2.1, page 100) pour montrer que

$$e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle}, \dots, e^{i\langle \mathbf{u}, \mathbf{X}_p \rangle} \text{ sont indépendants.}$$

Or $e^{i\langle \mathbf{u}, \mathbf{X} \rangle} = e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle} \dots e^{i\langle \mathbf{u}, \mathbf{X}_p \rangle}$, donc en utilisant la proposition III.2.9 (voir page 105), on montre que

$$\begin{aligned} \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] &= \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle} \dots e^{i\langle \mathbf{u}, \mathbf{X}_p \rangle}] \\ &= \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle}] \dots \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X}_p \rangle}] \\ &= \widehat{\mu}_1(\mathbf{u}) \dots \widehat{\mu}_p(\mathbf{u}) , \end{aligned}$$

ce qui est bien le résultat voulu. ■

On montre ensuite le critère d'indépendance suivant utilisant la fonction caractéristique.

Proposition III.2.18 $\blacktriangleleft \blacktriangleright$ Soient X_1, \dots, X_n , des v.a. **réelles** \mathcal{F} -mesurables. Elles sont indépendantes si et seulement si on a

$$\forall u_1, \dots, u_n \in \mathbb{R}, \quad \mathbf{E}[e^{iu_1 X_1 + \dots + iu_n X_n}] = \mathbf{E}[e^{iu_1 X_1}] \dots \mathbf{E}[e^{iu_n X_n}] . \quad (\text{III.10})$$

◀► **Preuve :** si les variables sont indépendantes, alors la proposition III.2.1, page 100, implique les variables $e^{iu_1X_1}, \dots, e^{iu_nX_n}$ sont indépendante et la proposition III.2.9 (voir page 105) implique (III.10).

Montrons la réciproque en supposant (III.10). On note μ_k la loi de X_k et on pose $\nu = \mu_1 \otimes \dots \otimes \mu_n$. On pose également $\mathbf{X} = (X_1, \dots, X_n)$, qui est vecteur aléatoire \mathcal{F} -mesurable de \mathbb{R}^n (par la proposition II.9.9, page 81). Soit $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$. On observe, par Fubini que

$$\begin{aligned} \widehat{\nu}(\mathbf{u}) &= \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\nu \\ &= \int_{\mathbb{R}} \dots \int_{\mathbb{R}} e^{iu_1x_1 + \dots + iu_nx_n} \mu_1(dx_1) \dots \mu_n(dx_n) \\ &= \left(\int_{\mathbb{R}} e^{iu_1x_1} \mu_1(dx_1) \right) \dots \left(\int_{\mathbb{R}} e^{iu_nx_n} \mu_n(dx_n) \right) \\ &= \mathbf{E}[e^{iu_1X_1}] \dots \mathbf{E}[e^{iu_nX_n}] \\ &= \mathbf{E}[e^{iu_1X_1 + \dots + iu_nX_n}] \\ &= \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] = \widehat{\mu}_{\mathbf{X}}(\mathbf{u}). \end{aligned}$$

On a donc montré que $\widehat{\nu} = \widehat{\mu}_{\mathbf{X}}$. L'injectivité de la transformée de Fourier (théorème II.10.4, page 87) implique laors que $\nu = \mu_{\mathbf{X}}$, c'est-à-dire que le vecteur (X_1, \dots, X_n) a pour loi $\mu_1 \otimes \dots \otimes \mu_n$. La proposition III.2.2 (iv) \Rightarrow (i) (voir page 100) montre alors que les v.a. X_1, \dots, X_n sont mutuellement indépendantes, ce qui complète la preuve. ■

Applications aux lois géométriques, binomiales et de Poisson. Donnons quelques applications des résultat précédents aux lois discrètes usuelles. La proposition suivante montre comment les variables géométriques apparaissent fréquemment et pourquoi le paramètre des géométrique s'appelle « paramètre de succès ».

Proposition III.2.19 Soit $(Y_n)_{n \in \mathbb{N}}$, une suite de v.a. de Bernoulli indépendantes de paramètre p . On pose

$$X = \inf\{n \in \mathbb{N} : Y_n = 1\}$$

avec la convention que $\inf \emptyset = \infty$. Alors, pour tout $k \in \mathbb{N}$, on a $\mathbf{P}(X = k) = p(1-p)^k$. Cela implique que $X < \infty$ p.s. et X est une variable géométrique de paramètre de succès p .

Preuve : on fixe $k \in \mathbb{N}$, on a $\{X = k\} = \{Y_k = 1\} \cap \bigcap_{0 \leq j \leq k-1} \{Y_j = 0\}$. Comme les v.a. Y_j sont indépendantes les événements $\{Y_k = 1\}, \{Y_0 = 0\}, \dots, \{Y_{k-1} = 0\}$ sont indépendants et on a

$$\mathbf{P}(X = k) = \mathbf{P}(Y_k = 1) \prod_{0 \leq j \leq k-1} \mathbf{P}(Y_j = 0) = p(1-p)^k.$$

On remarque ensuite que les événements $\{X = k\}$, $k \in \mathbb{N}$ sont deux-à-deux disjoints et que $\{X < \infty\} = \bigcup_{k \in \mathbb{N}} \{X = k\}$ donc

$$\mathbf{P}(X < \infty) = \sum_{k \in \mathbb{N}} \mathbf{P}(X = k) = \sum_{k \in \mathbb{N}} p(1-p)^k = 1,$$

ce qui permet de conclure. ■

La proposition suivante montre pourquoi les lois binomiales apparaissent fréquemment.

Proposition III.2.20 Soient Y_1, \dots, Y_n , des v.a. de Bernoulli indépendantes de paramètre p . Alors, la variable aléatoire $X := Y_1 + \dots + Y_n$ est de loi binomiale de paramètres (n, p) . On a donc d'une part

$$\mathbf{E}[X] = \mathbf{E}[Y_1] + \dots + \mathbf{E}[Y_n] = n\mathbf{E}[Y_1] = np,$$

d'autre part, par la proposition III.2.12, page 107, on a

$$\mathbf{var}(X) = \mathbf{var}[Y_1] + \dots + \mathbf{var}[Y_n] = n\mathbf{var}[Y_1] = np(1-p).$$

Preuve : on pose $X = Y_1 + \dots + Y_n$. Pour tout $r \in [0, 1]$, $r^X = r^{Y_1} \dots r^{Y_n}$. Or les variables r^{Y_1}, \dots, r^{Y_n} sont indépendantes (ce sont des fonctions déterministes de v.a. indépendantes). La proposition III.2.9, page 105 et (II.20) impliquent que

$$\mathbf{E}[r^X] = \mathbf{E}[r^{Y_1}] \dots \mathbf{E}[r^{Y_n}] = (1 - p + pr)^n.$$

Comme les fonctions génératrices caractérisent les lois, par (II.21) suit bien une loi binomiale de paramètre (n, p) , ce qui prouve la proposition. ■

La proposition suivante montre une propriété remarquable des lois de Poisson.

Proposition III.2.21 Soient X_1, \dots, X_n , des v.a. de Poisson de paramètres respectifs $\theta_1, \dots, \theta_n$. On les suppose indépendantes. Alors $X_1 + \dots + X_n$ est une v.a. de Poisson de paramètre $\theta_1 + \dots + \theta_n$.

Preuve : on observe que

$$\mathbf{E}[r^{X_1 + \dots + X_n}] = \mathbf{E}[r^{X_1}] \dots \mathbf{E}[r^{X_n}] = \exp(-(\theta_1 + \dots + \theta_n)(1 - r)),$$

ce qui permet de conclure puisque les fonctions génératrices caractérisent les lois sur \mathbb{N} . ■

III.3 Lemme de Borel-Cantelli, réciproques partielles et lois des grands nombres de Borel.

III.3.a Lemme de Borel-Cantelli et loi des grands nombres de Borel.

Dans cette section, $(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace de probabilité de référence. On commence par quelques notations importantes. On rappelle tout d'abord que pour un événement $A \in \mathcal{F}$

" $\omega \in A$ " s'interprète et se lit comme "le résultat ω réalise l'événement A ".

De manière un peu absurde et redondante on peut écrire $A = \{\omega \in \Omega : \omega \text{ réalise } A\}$.

Soit $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, une suite d'événements. On pose

$$A^* = \{\omega \in \Omega : \omega \in A_n \text{ pour une infinité de } n\}$$

et

$$A_* = \{\omega \in \Omega : \omega \in A_n \text{ à partir d'un certain rang}\} .$$

Autrement dit A^* est l'événement "une infinité de A_n sont réalisés" et A_* est l'événement "les A_n sont tous réalisés à partir d'un certain rang". On vérifie l'écriture suivante :

$$A^* = \bigcap_{N \geq 0} \bigcup_{n \geq N} A_n \quad \text{et} \quad A_* = \bigcup_{N \geq 0} \bigcap_{n \geq N} A_n . \quad (\text{III.11})$$

Les égalités (III.11) montrent que $A^* \in \mathcal{F}$ et que $A_* \in \mathcal{F}$. On remarque ensuite que

$$\mathbf{1}_{A^*} = \limsup_{n \rightarrow \infty} \mathbf{1}_{A_n} \quad \text{et} \quad \mathbf{1}_{A_*} = \liminf_{n \rightarrow \infty} \mathbf{1}_{A_n} . \quad (\text{III.12})$$

Cela justifie les deux notations suivantes pour les ensembles A^* et A_* :

$$\limsup_{n \rightarrow \infty} A_n := A^* = \bigcap_{N \geq 0} \bigcup_{n \geq N} A_n \quad \text{et} \quad \liminf_{n \rightarrow \infty} A_n := A_* = \bigcup_{N \geq 0} \bigcap_{n \geq N} A_n .$$

On remarque alors que

$$\Omega \setminus \left(\limsup_{n \rightarrow \infty} A_n \right) = \liminf_{n \rightarrow \infty} (\Omega \setminus A_n) \quad \text{et} \quad \Omega \setminus \left(\liminf_{n \rightarrow \infty} A_n \right) = \limsup_{n \rightarrow \infty} (\Omega \setminus A_n) .$$

Donnons une autre représentation des ensembles A^* et A_* . Pour cela on introduit les fonctions de comptage suivantes

$$N = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n} \quad \text{et} \quad N' = \sum_{n \in \mathbb{N}} \mathbf{1}_{\Omega \setminus A_n}$$

On voit que $N, N' : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ sont \mathcal{F} -mesurables et on vérifie que

$$A^* = \{N = \infty\} \quad \text{et} \quad A_* = \{N' < \infty\} . \quad (\text{III.13})$$

Lemme III.3.1 ◀▶ (Borel-Cantelli) Soient $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. On suppose que $\sum_{n \geq 0} \mathbf{P}(A_n) < \infty$. Alors, $\mathbf{P}(\limsup_n A_n) = 0$ et donc

$$\mathbf{P}(\text{aucun des } A_n \text{ n'est réalisé à partir d'un certain rang}) = 1 .$$

◀▶ **Preuve** : on pose $N = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}$. Par interversion intégrale/espérance positive, on a $\mathbf{E}[N] = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n)$. Donc $N < \infty$ p.s. et on conclut par (III.13). ■

En application du lemme de Borel-Cantelli on énonce et prouve la loi des grands nombres de Borel : ce résultat est généralisé dans le chapitre (l'hypothèse de moments d'ordre 4 n'est pas optimale).

Proposition III.3.2 ◀▶ (Loi des grands nombre de Borel) Soit $(X_n)_{n \geq 1}$, une suite de v.a. réelles \mathcal{F} -mesurables. On les suppose indépendantes et de même loi. On suppose de plus qu'elles admettent un moment d'ordre 4 (ce qui implique par Jensen qu'elles sont intégrables). Alors,

$$\mathbf{P}\text{-p.s.} \quad \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[X_1] .$$

Preuve : on pose $Y_n := X_n - \mathbf{E}[X_1]$ et $S_n := Y_1 + \dots + Y_n$, si bien que $\frac{1}{n}S_n = \frac{1}{n}(X_1 + \dots + X_n) - \mathbf{E}[X_1]$. Il est clair que les $(Y_n)_{n \geq 1}$ sont indépendantes et de même loi. Par ailleurs, elles admettent également un moment d'ordre 4 et elles sont d'espérance nulle : $\mathbf{E}[Y_n] = 0$. En développant simplement la somme on constate que

$$S_n^4 = \sum_{1 \leq i_1, \dots, i_4 \leq n} \mathbf{E}[Y_{i_1} \dots Y_{i_4}] .$$

Mais on observe que $\mathbf{E}[Y_{i_1} \dots Y_{i_4}] = 0$ si $\{i_1, \dots, i_4\}$ compte 4 ou 3 éléments car dans ces cas, une variable distincte des autres est à la puissance 1 et on rappelle que les Y_i sont d'espérance nulle et indépendantes. Pour les mêmes raisons, $\mathbf{E}[Y_{i_1} \dots Y_{i_4}] = 0$ si $\{i_1, \dots, i_4\}$ compte 2 éléments et si l'une des variables dans l'espérance est à la puissance 3 et l'autre (nécessairement) à la puissance 1. Un calcul élémentaire implique donc que

$$\mathbf{E}[S_n^4] = \sum_{1 \leq i \leq n} \mathbf{E}[Y_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbf{E}[Y_i^2 Y_j^2] = n\mathbf{E}[Y_1^4] + 3n(n-1)\mathbf{E}[Y_1^2]^2 \leq Cn^2 ,$$

où $C := 3\mathbf{E}[Y_1^4]$, car $\mathbf{E}[Y_1^2]^2 \leq \mathbf{E}[Y_1^4]$ par Jensen. En utilisant l'inégalité de Markov pour la fonction $\phi(x) = x^4$, on obtient pour tout $\varepsilon \in]0, \infty[$,

$$\mathbf{P}\left(\left|\frac{1}{n}S_n\right| > \varepsilon\right) \leq \frac{\mathbf{E}[S_n^4]}{n^4 \varepsilon^4} \leq \frac{C}{n^2 \varepsilon^4} .$$

On en déduit donc que

$$\sum_{n \geq 1} \mathbf{P}\left(\left|\frac{1}{n}S_n\right| > n^{-\frac{1}{8}}\right) \leq \sum_{n \geq 1} \frac{C}{n^{3/2}} < \infty ,$$

et le lemme de Borel-Cantelli implique que **P**-p.s. $\left|\frac{1}{n}S_n\right| \leq n^{-\frac{1}{8}}$ pour tout n assez grand, ce qui implique le résultat voulu. ■

Comme déjà mentionné, l'hypothèse de moment d'ordre 4 n'est pas optimale et nous revenons plus loin sur la loi des grands nombres. Néanmoins, la proposition précédente contient les cas des variables bornées et sa preuve est élémentaire. Nous renvoyons aux commentaires qui suivent l'énoncé général de la loi des grands nombres (théorème IV.4.3 page 138).

III.3.b Réciproques partielles du lemme de Borel-Cantelli.

Le lemme Borel-Cantelli admet la réciproque partielle suivante.

Lemme III.3.3 ◀▶ (Lemme de Borel) Soient $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. On fait les hypothèses suivantes.

- (a) Les événements A_n , $n \in \mathbb{N}$, sont mutuellement indépendants.
- (b) $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) = \infty$.

Alors,

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbf{P}\left(\text{une infinité de } A_n \text{ sont réalisés}\right) = 1 .$$

Preuve : pour tout $n \in \mathbb{N}$, on pose $N_n = \sum_{0 \leq k \leq n} \mathbf{1}_{A_k}$. L'indépendance des A_n entraîne

$$\mathbf{E}\left[e^{-N_n}\right] = \prod_{1 \leq k \leq n} \mathbf{E}\left[e^{-\mathbf{1}_{A_k}}\right] .$$

Mais $\mathbf{E}[e^{-1A_k}] = e^{-1}\mathbf{P}(A_k) + 1 - \mathbf{P}(A_k) = 1 - (1 - e^{-1})\mathbf{P}(A_k)$. On utilise ensuite l'inégalité de convexité $1 - r \leq e^{-r}$, qui implique que

$$\mathbf{E}[e^{-N_n}] \leq \exp\left(- (1 - e^{-1}) \sum_{0 \leq k \leq n} \mathbf{P}(A_k)\right).$$

On en déduit que $\lim_n \mathbf{E}[e^{-N_n}] = 0$. Or, si on pose $N = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}$, on a $\lim_n \uparrow N_n = N$ et par convergence dominée $\lim_n \mathbf{E}[e^{-N_n}] = \mathbf{E}[e^{-N}]$. On a donc montré que $\mathbf{E}[e^{-N}] = 0$, c'est-à-dire que p.s. $N = \infty$, ce qui implique le résultat voulu. ■

Le lemme suivant est une généralisation utile du lemme de Borel sous des hypothèses plus faibles.

Lemme III.3.4 (Kochen-Stone) Soit $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. On pose $N_n = \mathbf{1}_{A_0} + \dots + \mathbf{1}_{A_n}$ et on fait les hypothèses suivantes.

$$\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) = \infty \quad \text{et} \quad c := \limsup_{n \rightarrow \infty} \frac{\mathbf{E}[N_n^2]}{(\mathbf{E}[N_n])^2} < \infty.$$

Alors,

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbf{P}\left(\text{une infinité de } A_n \text{ sont réalisés}\right) > 1/c.$$

Preuve : pour tous $m, n \in \mathbb{N}$ tels que $m \leq n$, on pose $N_{m,n} = N_n - N_m$. On observe que $\bigcup_{m < k \leq n} A_k = \{N_{m,n} > 0\}$. Par l'inégalité du lemme II.4.5, page 56, on a donc

$$\mathbf{P}\left(\bigcup_{m < k \leq n} A_k\right) \geq \frac{(\mathbf{E}[N_{m,n}])^2}{\mathbf{E}[N_{m,n}^2]}. \quad (\text{III.14})$$

Pour tout $n \in \mathbb{N}$, on pose ensuite $a_n = \mathbf{E}[N_n^2]$ et $b_n = \mathbf{E}[N_n]$. En développant le carré sous l'espérance et en utilisant Cauchy-Schwarz, on obtient les inégalités suivantes

$$\begin{aligned} \mathbf{E}[N_{m,n}^2] &= a_n + a_m - 2\mathbf{E}[N_m N_n] \\ &\leq a_n + a_m + 2\mathbf{E}[N_m N_n] \\ &\leq a_n + a_m + 2(a_m a_n)^{1/2} \\ &= (a_m^{1/2} + a_n^{1/2})^2. \end{aligned}$$

Donc

$$\frac{(\mathbf{E}[N_{m,n}])^2}{\mathbf{E}[N_{m,n}^2]} \geq \frac{(b_n - b_m)^2}{(a_m^{1/2} + a_n^{1/2})^2} = \left(\frac{1 - b_n^{-1} b_m}{(b_n^{-2} a_n)^{1/2} + (b_n^{-2} a_m)^{1/2}}\right)^2. \quad (\text{III.15})$$

Or par hypothèse $\lim_n b_n = \infty$ et $\limsup_n b_n^{-2} a_n = c < \infty$. On déduit de (III.14) et de (III.15) que pour tout $m \in \mathbb{N}$,

$$\mathbf{P}\left(\bigcup_{k > m} A_k\right) = \lim_{n \rightarrow \infty} \mathbf{P}\left(\bigcup_{m < k \leq n} A_k\right) \geq \liminf_{n \rightarrow \infty} \frac{(\mathbf{E}[N_{m,n}])^2}{\mathbf{E}[N_{m,n}^2]} \geq \frac{1}{c}.$$

On a donc

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \lim_{m \rightarrow \infty} \mathbf{P}\left(\bigcup_{k > m} A_k\right) \geq \frac{1}{c},$$

qui est bien le résultat désiré. ■

Corollaire III.3.5 Soient $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. On fait les hypothèses suivantes.

(a) Les événements A_n , $n \in \mathbb{N}$, sont deux-à-deux indépendants, c'est-à-dire que pour tous $m < n$,
 $\mathbf{P}(A_m \cap A_n) = \mathbf{P}(A_m)\mathbf{P}(A_n)$.

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

Alors,

$$\mathbf{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbf{P}\left(\text{une infinité de } A_n \text{ sont réalisés}\right) = 1.$$

Preuve : avec les notation du lemme de Kochen-Stone, on voit, en développant le carré que

$$\begin{aligned} \mathbf{E}[N_n^2] &= \sum_{0 \leq k \leq n} \mathbf{P}(A_k) + 2 \sum_{0 \leq k < \ell \leq n} \mathbf{P}(A_k \cap A_\ell) = \sum_{0 \leq k \leq n} \mathbf{P}(A_k) + 2 \sum_{0 \leq k < \ell \leq n} \mathbf{P}(A_k)\mathbf{P}(A_\ell) \\ &= \sum_{0 \leq k \leq n} \mathbf{P}(A_k) - \mathbf{P}(A_k)^2 + (\mathbf{E}[N_n])^2 \leq \mathbf{E}[N_n] + (\mathbf{E}[N_n])^2. \end{aligned}$$

Cela implique que $\limsup_n \mathbf{E}[N_n^2]/(\mathbf{E}[N_n]) \leq 1$ et le lemme de Kochen-Stone implique le résultat voulu. ■

◀► **Application.** On note Σ , un ensemble fini contenant au moins deux éléments. On voit Σ comme un alphabet. Soit $X_n : \Omega \rightarrow \Sigma$, $n \in \mathbb{N}$, un suite de v.a. \mathcal{F} -mesurables indépendantes et de même loi μ telle que

$$\forall a \in \Sigma, \quad \mathbf{P}(X_n = a) = \mu(a) > 0.$$

Alors, on montre le résultat suivant

$$\mathbf{P}\text{-p.s.} \quad \forall k \geq 1, \forall (a_1, \dots, a_k) \in \Sigma^k, \quad \#\{n \in \mathbb{N} : (X_n, \dots, X_{n+k-1}) = (a_1, \dots, a_k)\} = \infty. \quad (\text{III.16})$$

Preuve de (III.16) : pour tout $w = (a_1, \dots, a_k) \in \Sigma^k$ et tout $n \in \mathbb{N}$, on pose

$$A_{w,n} = \{X_n = a_1; \dots; X_{n+k-1} = a_k\} \quad \text{et} \quad A_w = \limsup_{m \rightarrow \infty} A_{w,km}.$$

Il est clair que les événements $A_{w,km}$, $m \in \mathbb{N}$, sont indépendants. De plus on a

$$\mathbf{P}(A_{w,n}) = \mu(a_1) \dots \mu(a_k) > 0.$$

Donc $\sum_{m \in \mathbb{N}} \mathbf{P}(A_{w,km}) = \infty$ et le lemme III.3.3 implique $\mathbf{P}(A_w) = 1$. On pose ensuite

$$A = \bigcap_{k \geq 1} \bigcap_{w \in \Sigma^k} A_w.$$

Comme $\bigcup_{k \geq 1} \Sigma^k$ est dénombrable, on a $A \in \mathcal{F}$ et $\mathbf{P}(A) = 1$. Or

$$\forall \omega \in A, \forall k \geq 1, \forall (a_1, \dots, a_k) \in \Sigma^k, \quad \#\{n \in \mathbb{N} : (X_n(\omega), \dots, X_{n+k-1}(\omega)) = (a_1, \dots, a_k)\} = \infty,$$

ce qui permet de conclure. ■

Imaginons un singe tapant à la machine : il appuie sur les touches totalement au hasard ; ici Σ représente les lettres et les caractères habituels. Si ce singe vit un temps infini, et tape une infinité de fois, il finira par écrire *L'Homme Sans Qualité* de R. Musil et toutes les fins possibles de ce roman ; il écrira non seulement tous les modes d'emploi de toutes les friteuses en vente dans le commerce, mais aussi celui des friteuses du futur et plus il le fera une infinité de fois.

Exemple III.3.1 ◀► Une application de (III.16) concerne la *marche aléatoire symétrique* sur \mathbb{Z} qui se définit comme suit. Soient $\xi_n : \Omega \rightarrow \{-1, 1\}$, $n \geq 1$, une suite de v.a. \mathcal{F} -mesurables indépendantes et de même loi donnée par

$$\mathbf{P}(\xi_n = -1) = \mathbf{P}(\xi_n = 1) = 1/2 .$$

On pose

$$S_0 = 0 \quad \text{et} \quad \forall n \geq 1, \quad S_n = \xi_1 + \dots + \xi_n .$$

On note W_n l'ensemble des fonctions $w : \{0, \dots, n\} \rightarrow \mathbb{Z}$ telles que $w(0) = 0$ et $|w(k+1) - w(k)| = 1$, pour tout $0 \leq k < n$. On voit que W_n est l'ensemble des trajectoires possibles de (S_0, \dots, S_n) . Par ailleurs on vérifie facilement que $\#W_n = 2^n$ et que pour tout $w \in W_n$, on a

$$\mathbf{P}((S_0, \dots, S_n) = w) = 2^{-n} = \frac{1}{\#W_n} .$$

La loi de (S_0, \dots, S_n) est donc uniforme sur W_n . Pour tout $u \in \mathbb{R}$, on remarque que

$$\mathbf{E}[e^{iuS_n}] = \prod_{1 \leq k \leq n} \mathbf{E}[e^{iu\xi_k}] = \left(\frac{1}{2}(e^{iu} + e^{-iu})\right)^n = \sum_{0 \leq \ell \leq n} 2^{-n} \binom{n}{\ell} e^{iu(2\ell-n)} .$$

On en déduit que pour tout $k \in \{-n, \dots, n\}$,

$$\mathbf{P}(S_n = k) = 2^{-n} \binom{n}{\frac{1}{2}(n+k)} \quad \text{si } n \text{ et } k \text{ ont même parité et } \mathbf{P}(S_n = k) = 0 \text{ sinon.}$$

On en déduit que

$$\#\{w \in W_n : w(n) = k\} = \binom{n}{\frac{1}{2}(n+k)} \quad \text{si } n-k \text{ pair et } \#\{w \in W_n : w(n) = k\} = 0 \text{ sinon.}$$

On peut retrouver ce résultat par un raisonnement combinatoire élémentaire.

On veut montrer que p.s. la suite $(S_n)_{n \in \mathbb{N}}$ est non-bornée, c'est-à-dire que

$$\mathbf{P}\text{-p.s.} \quad \limsup_{n \rightarrow \infty} |S_n| = \infty . \tag{III.17}$$

Preuve de (III.17) : on remarque que pour tout $k \geq 1$,

$$\{\#\{n \in \mathbb{N} : \xi_n = \xi_{n+1} = \dots = \xi_{n+2k} = 1\} = \infty\} \subset \{\limsup_{n \rightarrow \infty} |S_n| \geq k\} ,$$

ce qui entraîne (III.17) par (III.16). ■

III.4 Construction des suites de v.a. réelles indépendantes.

Le but de cette section est de montrer le résultat d'existence suivant.

Théorème III.4.1 ◀► Soient $\mu_n : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$, $n \in \mathbb{N}$, une suite de mesures de probabilité. Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et une suite $X_n : \Omega \rightarrow \mathbb{R}$ de v.a. \mathcal{F} -mesurables satisfaisant les propriétés suivantes.

(i) Les v.a. X_n , $n \in \mathbb{N}$, sont indépendantes.

(ii) Pour tout $n \in \mathbb{N}$, la loi de X_n est μ_n , c'est-à-dire que $\mu_n(B) = \mathbf{P}(X_n \in B)$, pour tout $B \in \mathcal{B}(\mathbb{R})$.

La preuve du théorème de fait en plusieurs étapes. On commence par rappeler quelques résultats sur la décomposition des nombres réels en base b , où ici b désigne un nombre entier supérieur ou égal à 2. On rappelle en particulier la proposition suivante.

Proposition III.4.2 À tout réel $x \in [0, 1[$, on associe une suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ telle que

(a) $D_n^{(b)}(x) \in \{0, 1, \dots, b-1\}$, $n \in \mathbb{N}$.

(b) La suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ n'est pas constante à $b-1$ à partir d'un certain rang.

(c) $x = \sum_{n \in \mathbb{N}} b^{-n-1} D_n^{(b)}(x)$.

Par ailleurs, si on note S_b l'ensemble des suites à valeurs dans $\{0, \dots, b-1\}$ qui ne sont pas constantes à $b-1$ à partir d'un certain rang, alors

$$x \in [0, 1[\longmapsto (D_n^{(b)}(x))_{n \in \mathbb{N}} \in S_b \text{ est une bijection.}$$

La suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ est appelé le **développement en base b du réel x** .

Preuve : supposons qu'il existe une suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ satisfaisant (a), (b) et (c). Alors pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} b^{n+1}x &= b^n D_0^{(b)}(x) + b^{n-1} D_1^{(b)}(x) + \dots + b^2 D_{n-2}^{(b)}(x) + b D_{n-1}^{(b)}(x) + D_n^{(b)}(x) \\ &\quad + \sum_{p>n} b^{n-p} D_p^{(b)}(x). \end{aligned}$$

On remarque que :

$$N := b^n D_0^{(b)}(x) + b^{n-1} D_1^{(b)}(x) + \dots + b^2 D_{n-2}^{(b)}(x) + b D_{n-1}^{(b)}(x) + D_n^{(b)}(x)$$

est un entier ; en posant $q = p - n$, on remarque aussi que :

$$\begin{aligned} r := \sum_{p>n} b^{n-p} D_p^{(b)}(x) &< \sum_{p>n} b^{n-p} (b-1) \\ &< \sum_{q=1n} b^{-q} (b-1) = 1. \end{aligned}$$

L'inégalité stricte se justifie par le fait la suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ n'est pas constante à $b-1$ à partir d'un certain rang. On a donc écrit $b^{n+1}x = N + r$, avec $N \in \mathbb{N}$ et $r \in [0, 1[$. On obtient donc le résultat suivant sur la partie entière de $b^{n+1}x$:

$$\lfloor b^{n+1}x \rfloor = N = b^n D_0^{(b)}(x) + \dots + b^2 D_{n-2}^{(b)}(x) + b D_{n-1}^{(b)}(x) + D_n^{(b)}(x)$$

Au rang $n - 1$ on a donc

$$\lfloor b^n x \rfloor = b^{n-1} D_0^{(b)}(x) + \dots + b D_{n-2}^{(b)}(x) + D_{n-1}^{(b)}(x).$$

Par conséquent on vérifie que

$$\forall n \in \mathbb{N}, \quad D_n^{(b)}(x) = \lfloor b^{n+1}x \rfloor - b \lfloor b^n x \rfloor \quad (\text{III.18})$$

On a donc montré que s'il existe une suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ satisfaisant (a), (b) et (c), alors elle nécessairement donnée par (III.18).

Raisonnons maintenant dans l'autre direction : on définit la suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ par (III.18) et nous voulons vérifier qu'elle satisfait (a), (b) et (c). Pour cela on remarque tout d'abord que $\lfloor b^{n+1}x \rfloor \leq b^{n+1}x < \lfloor b^{n+1}x \rfloor + 1$, c'est-à-dire $b^{n+1}x - 1 < \lfloor b^{n+1}x \rfloor \leq b^{n+1}x$. De même $b^n x - 1 < \lfloor b^n x \rfloor \leq b^n x$ et donc $b^{n+1}x - b < b \lfloor b^n x \rfloor \leq b^{n+1}x$ et par conséquent,

$$-1 < D_n^{(b)}(x) = \lfloor b^{n+1}x \rfloor - b \lfloor b^n x \rfloor \leq b - 1.$$

Cela montre que $D_n^{(b)}(x) \in \{0, \dots, b-1\}$, pour tout $n \in \mathbb{N}$, ce qui est la propriété (a).

Montrons ensuite que la suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ n'est pas stationnaire à la valeur $b-1$. Supposons le contraire, c'est-à-dire qu'il existe un entier $n_0 \in \mathbb{N}$ tel que $D_n^{(b)}(x) = b-1$ pour tout $n \geq n_0$. Pour tout $y \in \mathbb{R}$, on note $\{y\} := y - \lfloor y \rfloor$, sa partie fractionnaire. On a nécessairement $0 \leq \{y\} < 1$ (la partie fractionnaire ne peut être égale à 1). Par ailleurs, on remarque facilement que $D_n^{(b)}(x) = b\{b^n x\} - \{b^{n+1}x\}$, pour tout $n \in \mathbb{N}$. Donc,

$$\forall n \geq n_0, \quad b\{b^n x\} - \{b^{n+1}x\} = b-1, \quad \text{c'est-à-dire} \quad 1 - \{b^{n+1}x\} = b(1 - \{b^n x\}).$$

Cela implique que

$$\forall n \in \mathbb{N}, \quad 1 - \{b^{n+n_0}x\} = b^n(1 - \{b^{n_0}x\}) \quad \text{et donc} \quad 0 \leq 1 - \{b^{n_0}x\} \leq b^{-n},$$

car $1 - \{b^{n+n_0}x\} \leq 1$. En faisant tendre n vers l'infini, cela implique que $1 - \{b^{n_0}x\} = 0$, ce qui est impossible pour une partie fractionnaire. Par conséquent, la suite $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ n'est pas stationnaire à la valeur $b-1$, ce qui est la propriété (b).

La propriété est (c) est facile à montrer : on vérifie que

$$\begin{aligned} \sum_{n \in \mathbb{N}} b^{-n-1} D_n^{(b)}(x) &= \sum_{n \in \mathbb{N}} (b^{-(n+1)} \lfloor b^{n+1}x \rfloor - b^{-n} \lfloor b^n x \rfloor) \\ &= \lim_{n \rightarrow \infty} b^{-(n+1)} \lfloor b^{n+1}x \rfloor - b^{-0} \lfloor b^0 x \rfloor \end{aligned}$$

par télescopage. Or $b^{-0} \lfloor b^0 x \rfloor = \lfloor x \rfloor = 0$ et $\lim_{n \rightarrow \infty} b^{-(n+1)} \lfloor b^{n+1}x \rfloor = x$, ce qui termine la preuve de (c).

Le caractère bijectif de l'application $\phi : x \in [0, 1[\mapsto (D_n^{(b)}(x))_{n \in \mathbb{N}} \in S_b$ est ensuite très facile à démontrer car ϕ admet la réciproque φ qui à toute suite $(d_n)_{n \in \mathbb{N}} \in S_b$ associe $\varphi((d_n)_{n \in \mathbb{N}}) = \sum_{n \in \mathbb{N}} b^{-n-1} d_n$. Cela complète la preuve de la proposition. ■

On montre ensuite le résultat suivant.

Proposition III.4.3 *On définit l'espace de probabilité suivant :*

$$\Omega := [0, 1[, \quad \mathcal{F} := \mathcal{B}([0, 1]), \quad \mathbf{P} = \ell, \quad \text{la mesure de Lebesgue.}$$

Soit b , un entier supérieur ou égal à 2. Pour tout $x \in [0, 1[$, on note $(D_n^{(b)}(x))_{n \in \mathbb{N}}$ son développement en base b .

Alors les variables $D_n^{(b)} : \Omega \rightarrow \{0, 1, \dots, b-1\}$ sont \mathcal{F} -mesurables, mutuellement indépendantes sous \mathbf{P} et de loi uniforme sur $\{0, \dots, b-1\}$, c'est-à-dire que

$$\forall a \in \{0, \dots, b-1\}, \quad \mathbf{P}(D_n^{(b)} = a) = 1/b.$$

Preuve : on rappelle (III.18) qui montre que les fonctions $D_n^{(b)} : [0, 1[\rightarrow \{0, 1, \dots, b-1\}$ sont $\mathcal{B}([0, 1[)$ -mesurables car la partie entière est mesurable. Cela signifie strictement la même chose que de dire que $D_n^{(b)} : \Omega \rightarrow \{0, 1, \dots, b-1\}$ est \mathcal{F} -mesurable.

Soient $\varepsilon_0, \dots, \varepsilon_n \in \{0, 1, \dots, b-1\}$. On pose $z = \sum_{0 \leq k \leq n} b^{-k-1} \varepsilon_k$. L'unicité du développement en base b implique immédiatement que

$$\begin{aligned} \{D_0^{(b)} = \varepsilon_0\} \cap \{D_1^{(b)} = \varepsilon_1\} \cap \dots \cap \{D_n^{(b)} = \varepsilon_n\} &= \{x \in [0, 1[: D_0^{(b)}(x) = \varepsilon_0 \text{ et } \dots \text{ et } D_n^{(b)}(x) = \varepsilon_n\} \\ &= [z, z + b^{-n-1}[. \end{aligned}$$

On en déduit alors que

$$\mathbf{P}(D_0^{(b)} = \varepsilon_0; D_1^{(b)} = \varepsilon_1; \dots; D_n^{(b)} = \varepsilon_n) = \ell([z, z + b^{-n-1}[) = \frac{1}{b^{n+1}}. \quad (\text{III.19})$$

On fixe $j \in \{0, \dots, n\}$. On remarque ensuite que pour tout $\varepsilon \in \{0, 1, \dots, b-1\}$,

$$\{D_n^{(b)} = \varepsilon\} = \bigcup_{\substack{\varepsilon_k \in \{0, \dots, b-1\} \\ 0 \leq k \leq n-1}} \{D_0^{(b)} = \varepsilon_0\} \cap \{D_1^{(b)} = \varepsilon_1\} \cap \dots \cap \{D_{n-1}^{(b)} = \varepsilon_{n-1}\} \cap \{D_n^{(b)} = \varepsilon\}.$$

Comme les b^n événements du membre de droite sont deux-à-deux disjoints et qu'ils ont tous la même probabilité b^{-n-1} , on en déduit que

$$\forall \varepsilon \in \{0, 1\}, \quad \mathbf{P}(D_n^{(b)} = \varepsilon) = b^n b^{-n-1} = \frac{1}{b}.$$

Donc par (III.19)

$$\forall \varepsilon_0, \dots, \varepsilon_n \in \{0, \dots, b-1\}, \quad \mathbf{P}(D_0^{(b)} = \varepsilon_0; D_1^{(b)} = \varepsilon_1; \dots; D_n^{(b)} = \varepsilon_n) = \prod_{0 \leq k \leq n} \mathbf{P}(D_k^{(b)} = \varepsilon_k).$$

Cela implique que pour tout $n \in \mathbb{N}$, $D_0^{(b)}, \dots, D_n^{(b)}$ sont mutuellement indépendantes sous \mathbf{P} , ce qui implique que les v.a. $D_n^{(b)}$, $n \in \mathbb{N}$ sont mutuellement indépendantes sous \mathbf{P} , ce qui termine la preuve de la proposition. ■

Une première application aux nombres normaux.

Définition III.4.1 (*Nombres normaux*) Soit $x \in [0, 1[$. Soit b , un entier supérieur ou égal à 2. On dit que x est normal en base b si pour tout $a \in \{0, \dots, b-1\}$, la fréquence d'apparition de la « b -cimale » a est asymptotiquement de $1/b$, c'est-à-dire

$$\forall a \in \{0, \dots, b-1\}, \quad \frac{1}{n} (\mathbf{1}_{\{D_0^{(b)}(x)=a\}} + \dots + \mathbf{1}_{\{D_n^{(b)}(x)=a\}}) \xrightarrow{n \rightarrow \infty} 1/b.$$

On dit que le nombre x est normal s'il est normal pour toutes les bases $b \geq 2$. □

En application de la proposition III.4.3 et de la loi des grands nombres de Borel, on montre le théorème suivant qui est dû à Borel.

Théorème III.4.4 *Lebesgue presque tout réel $x \in [0, 1[$ est normal.*

Preuve : on se place, comme dans la proposition III.4.3 sur l'espace de probabilité.

$$\Omega := [0, 1[, \quad \mathcal{F} := \mathcal{B}([0, 1]), \quad \mathbf{P} = \ell, \text{ la mesure de Lebesgue.}$$

On fixe $b \geq 2$ et $a \in \{0, \dots, b-1\}$ et on pose

$$X_n = \mathbf{1}_{\{D_n^{(b)}=a\}}.$$

Comme les X_n sont des fonctions déterministes des $D_n^{(b)}$, et comme les $D_n^{(b)}$ sont indépendantes (proposition III.4.3), les variables X_n sont également indépendante (proposition III.2.1, page 100). Par ailleurs, par la proposition III.4.3

$$\mathbf{P}(X_n = 1) = 1 - \mathbf{P}(X_n = 0) = \mathbf{P}(D_n^{(b)} = a) = 1/b.$$

Il est clair que $\mathbf{E}[X_n^4] = \mathbf{E}[X_n] = 1/b$. Donc la suite $(X_n)_{n \in \mathbb{N}}$ est une suite de v.a. indépendantes, de même loi et qui admettent un moment d'ordre 4. La loi des grands nombres de Borel (théorème III.3.2, page 112) s'applique et on a

$$\mathbf{P}\text{-p.s.} \quad \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[X_1] = 1/b.$$

Si l'on retraduit ce que cela signifie, on voit que $X_1 + \dots + X_n = \mathbf{1}_{\{D_0^{(b)}=a\}} + \dots + \mathbf{1}_{\{D_n^{(b)}=a\}}$ et comme $\Omega = [0, 1[$ et $\mathbf{P} = \ell$, cela montre qu'il existe $N_{a,b} \in \mathcal{B}([0, 1[)$ tel que

$$\ell(N_{a,b}) = 0 \quad \text{et} \quad \forall x \in [0, 1[\setminus N_{a,b}, \quad \frac{1}{n} (\mathbf{1}_{\{D_0^{(b)}(x)=a\}} + \dots + \mathbf{1}_{\{D_n^{(b)}(x)=a\}}) \xrightarrow[n \rightarrow \infty]{} 1/b.$$

On pose alors $N = \bigcup_{b \geq 2} \bigcup_{0 \leq a < b} N_{a,b}$. Il est clair que $N \in \mathcal{B}([0, 1[)$ car c'est une union dénombrable de boréliens. D'autre part, la sigma-sous-additivité de ℓ implique que :

$$\ell(N) \leq \sum_{b \geq 2} \sum_{0 \leq a < b} \ell(N_{a,b}) = 0.$$

On a donc

$$\forall x \in [0, 1[\setminus N, \quad \forall b \geq 2, \quad \forall a \in \{0, \dots, b-1\}, \quad \frac{1}{n} (\mathbf{1}_{\{D_0^{(b)}(x)=a\}} + \dots + \mathbf{1}_{\{D_n^{(b)}(x)=a\}}) \xrightarrow[n \rightarrow \infty]{} 1/b,$$

ce qui signifie que ℓ -presque tout réel x est normal. ■

Remarque III.4.1 Bien que presque tout nombre réel de $[0, 1[$ soit normal, il semble assez difficile d'en exhiber un, concrètement. On peut montrer que le nombre dont le développement en base 10 s'écrivant :

$$x = 0, 1234567891011121314151617181920212223 \dots$$

est normal. Mais la preuve de ce fait est vraiment difficile. Par ailleurs on ne sait pas si la partie fractionnaire de π est un nombre normal, bien que cela ait été supposé. □

Suite de la preuve du théorème III.4.1. Pour simplifier les notations, on pose

$$\forall n \in \mathbb{N}, \quad D_n := D_n^{(2)}.$$

On montre alors la proposition suivante.

Proposition III.4.5 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité et soient $Y_n : \Omega \rightarrow \{0, 1\}$, $n \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables mutuellement indépendantes sous \mathbf{P} et telles que $\mathbf{P}(Y_n = 0) = \mathbf{P}(Y_n = 1) = \frac{1}{2}$, pour tout $n \in \mathbb{N}$. On pose

$$U = \sum_{n \geq 0} 2^{-n-1} Y_n.$$

Alors, $U : \Omega \rightarrow [0, 1]$ est \mathcal{F} -mesurable de loi uniforme sur $[0, 1]$.

Preuve : pour tout $n \in \mathbb{N}$ et tout $0 \leq m < 2^{n+1}$, on pose $I_{n,m} = [m2^{-n-1}, (m+1)2^{-n-1}[$. On décompose le nombre m en base deux : il existe $\varepsilon_0, \dots, \varepsilon_n \in \{0, 1\}$ tels que $m = \sum_{0 \leq k \leq n} \varepsilon_k 2^{n-k}$, on a donc $m2^{-n-1} = \sum_{0 \leq k \leq n} 2^{-k-1} \varepsilon_k$. Il est alors facile de vérifier que

$$\{U \in I_{n,m}\} = \{Y_0 = \varepsilon_0; Y_1 = \varepsilon_1; \dots; Y_n = \varepsilon_n\}.$$

Cela implique alors que

$$\begin{aligned} \mathbf{P}(U \in I_{n,m}) &= \mathbf{P}\left(\bigcap_{0 \leq k \leq n} \{Y_k = \varepsilon_k\}\right) = \prod_{0 \leq k \leq n} \mathbf{P}(Y_k = \varepsilon_k) \\ &= 2^{-n-1} = \ell(I_{n,m}). \end{aligned}$$

On pose $\mathcal{P} = \{[0, 1]\} \cup \{I_{n,m}; n \in \mathbb{N}, 0 \leq m < 2^{n+1}\}$. Il est clair que \mathcal{P} est un pi-système générant les Boréliens de $[0, 1]$. Or on vient de montrer que la loi de U sous \mathbf{P} coïncide sur \mathcal{P} avec $\ell(\cdot \cap [0, 1])$. Le théorème I.1.12 d'unicité du prolongement des mesures implique que la loi de U sous \mathbf{P} est $\ell(\cdot \cap [0, 1])$, ce qui montre la proposition. ■

Fin de la preuve du théorème III.4.1. Soit $(\Omega, \mathcal{F}, \mathbf{P})$ et $D_n : \Omega \rightarrow \{0, 1\}$ comme dans la proposition III.4.3. Pour tout $n \in \mathbb{N}$, on pose $\mathcal{F}_n = \sigma(D_n)$. Les tribus \mathcal{F}_n , $n \in \mathbb{N}$ sont donc mutuellement indépendantes sous \mathbf{P} . Soit $\phi : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, une bijection. Pour tout $p \in \mathbb{N}$, on pose

$$U_n = \sum_{p \geq 0} 2^{-p-1} D_{\phi(n,p)} \quad \text{et} \quad \mathcal{G}_n = \sigma(\mathcal{F}_{\phi(n,p)}; p \in \mathbb{N}).$$

Le théorème III.1.4 d'indépendance par paquets implique que les tribus \mathcal{G}_n , $n \in \mathbb{N}$, sont mutuellement indépendantes sous \mathbf{P} . Or U_n est \mathcal{G}_n -mesurable donc $\sigma(U_n) \subset \mathcal{G}_n$. La proposition III.1.1 implique alors l'indépendance mutuelle des U_n , $n \in \mathbb{N}$, sous \mathbf{P} . La proposition III.4.5 implique ensuite que $U_n : \Omega \rightarrow [0, 1]$ est \mathcal{F} -mesurable de loi uniforme sur $[0, 1]$.

Pour tout $n \in \mathbb{N}$, on pose ensuite

$$\forall y \in]0, 1[, \quad G_n(y) = \inf \{x \in \mathbb{R} : \mu_n(]-\infty, x]) > y\},$$

qui est le pseudo-inverse continu à droite de la fonction de répartition de μ_n . La proposition II.7.5, page 72, implique que $X_n := G_n(U_n)$ a pour loi μ_n . Comme $\sigma(X_n) \subset \sigma(U_n)$, La proposition III.1.1 implique alors l'indépendance mutuelle des X_n , $n \in \mathbb{N}$, ce qui termine la preuve du théorème. ■

III.5 Exercices.

Exercice III.5.1 Déterminer à quelle condition une variable aléatoire réelle est indépendante d'elle-même.

Exercice III.5.2 Soient $E = \{x_1, x_2, x_3\}$ et $F = \{y_1, y_2, y_3\}$ deux ensembles finis. Pour chacune des matrices $P = (P_{ij})_{i,j=1,2,3}$ ci-dessous, on considère un couple (X, Y) de variables aléatoires à valeurs dans $E \times F$ tel que pour tous $i, j \in \{1, 2, 3\}$, on ait $\mathbf{P}(X = x_i, Y = y_j) = P_{ij}$. Déterminer si les variables aléatoires X et Y sont indépendantes.

$$P = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{12} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{6} \end{pmatrix}, P = \begin{pmatrix} 0 & 0 & 0 \\ \frac{3}{17} & \frac{12}{17} & \frac{2}{17} \\ 0 & 0 & 0 \end{pmatrix}, P = \begin{pmatrix} \frac{1}{4} & \frac{3}{32} & \frac{3}{96} \\ \frac{2}{15} & \frac{1}{20} & \frac{1}{60} \\ \frac{17}{60} & \frac{17}{160} & \frac{17}{480} \end{pmatrix} P = \begin{pmatrix} \frac{1}{4} & \frac{3}{32} & \frac{3}{96} \\ \frac{2}{15} & \frac{1}{20} & \frac{1}{20} \\ \frac{1}{4} & \frac{1}{10} & \frac{3}{80} \end{pmatrix}.$$

Exercice III.5.3 a. Soient X et Y deux variables aléatoires indépendantes de loi gaussienne centrée réduite. Déterminer la loi de $\frac{X}{Y}$.

b. Soit Z une variable aléatoire qui suit la loi de Cauchy standard. Déterminer la loi de $\frac{1}{Z}$.

Exercice III.5.4 Calculer la loi de la somme de deux variables aléatoires indépendantes, l'une de loi de binomiale de paramètres n et p , l'autre de paramètres m et p , où $p \in [0, 1]$ et m, n sont deux entiers.

Exercice III.5.5 Montrer que si la somme de deux variables aléatoires discrètes indépendantes a la loi de Bernoulli de paramètre $p \in [0, 1]$, alors l'une des deux variables aléatoires est constante.

Exercice III.5.6 Soient X et Y deux variables aléatoires normales centrées réduites indépendantes. On pose $Z = X - Y$. Calculer la matrice de covariance du vecteur (X, Y, Z) et vérifier que $\text{Var}(X + Y + Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)$.

Exercice III.5.7 Soit z un nombre complexe.

a. Montrer que l'intégrale

$$I(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{zx} dx$$

est convergente.

b. Montrer que pour tout $n \geq 1$, on a l'inégalité

$$\left| \sum_{k=0}^n \frac{z^k x^k}{k!} \right| \leq e^{|z||x|}$$

et en déduire que

$$I(z) = \lim_{n \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \sum_{k=0}^n \frac{z^k x^k}{k!} dx.$$

c. Calculer $I(z)$.

d. Déterminer la fonction caractéristique d'une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$.

e. Soit X une variable aléatoire de loi normale. Montrer qu'on a l'égalité

$$\mathbf{E}[e^{X}] = e^{\mathbf{E}[X] + \frac{1}{2}\text{Var}(X)}.$$

Exercice III.5.8 Soient X_1, \dots, X_n des variables indépendantes et identiquement distribuées de loi normale centrée réduite. Soient a_1, \dots, a_n des réels. Déterminer la loi de $a_1X_1 + \dots + a_nX_n$.

Exercice III.5.9 Pour tout réel $t > 0$, on pose

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx.$$

Soient θ et t deux réels strictement positifs. On appelle loi gamma de paramètres θ et t la loi de densité

$$\frac{\theta^t}{\Gamma(t)} x^{t-1} e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x).$$

a. Vérifier que la fonction $\Gamma(t)$ est bien définie sur \mathbb{R}_+^* . Vérifier également que la densité donnée ci-dessus est bien d'intégrale égale à 1.

b. Calculer la fonction caractéristique de la loi $\Gamma(\theta, t)$.

c. Soient t_1, \dots, t_n des réels strictement positifs. Soient X_1, \dots, X_n des variables aléatoires indépendantes de lois respectives $\Gamma(\theta, t_1), \dots, \Gamma(\theta, t_n)$. Calculer la loi de la somme $X_1 + \dots + X_n$.

Exercice III.5.10 Pour toute variable aléatoire réelle X qui admet un moment d'ordre 3, on définit $k_3(X) = \mathbf{E}[X^3] - 3\mathbf{E}[X^2]\mathbf{E}[X] + 2\mathbf{E}[X]^3$.

Montrer que si X et Y sont deux variables aléatoires réelles indépendantes qui admettent un moment d'ordre 3, alors $k_3(X + Y) = k_3(X) + k_3(Y)$.

Exercice III.5.11 Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités. Soient $U, V : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{R}$ deux variables aléatoires indépendantes de loi uniforme sur l'intervalle $[-1, 1]$.

a. Calculer la fonction caractéristique de la variable aléatoire $U + V$.

b. Soit $X : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow \mathbb{R}$ une variable aléatoire. On suppose que X, U, V sont indépendantes. Montrer que la loi de $X + U + V$ admet une densité par rapport à la mesure de Lebesgue.

Exercice III.5.12 Soit Ω un ensemble. Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω . Déterminer les relations d'inclusion qui existent toujours entre les parties suivantes de Ω :

$$\bigcup_{M \geq 1} \bigcap_{N \geq M} \bigcup_{n \geq N} A_n, \quad \bigcap_{M \geq 1} \bigcup_{N \geq M} \bigcap_{n \geq N} A_n, \quad \bigcup_{N \geq 1} \bigcap_{n \geq N} A_n, \quad \bigcap_{N \geq 1} \bigcup_{n \geq N} A_n, \quad \bigcup_{n \geq 1} A_n, \quad \bigcap_{n \geq 1} A_n, \quad \Omega, \quad \emptyset.$$

Montrer par un exemple que six de ces parties peuvent être deux à deux distinctes.

Exercice III.5.13 Soit Ω un ensemble. Soit $(X_n)_{n \geq 1}$ une suite de fonctions à valeurs réelles définies sur Ω .

a. Décrire en français et sans utiliser les expressions “quel que soit” ni “il existe” les parties suivantes de Ω :

$$A = \bigcup_{a \in \mathbb{R}} \bigcup_{b \in \mathbb{R}} \bigcap_{n \geq 1} \{\omega \in \Omega : a \leq X_n(\omega) \leq b\},$$

$$B = \bigcup_{N \geq 1} \bigcap_{n \geq N} \bigcap_{m \geq n} \{\omega \in \Omega : X_n(\omega) - X_m(\omega) \geq 0\},$$

$$C = \bigcup_{\ell \in \mathbb{R}_+} \bigcap_{k \geq 1} \bigcup_{N \geq 1} \bigcap_{n \geq N} \left\{ \omega \in \Omega : |X_n(\omega) - \ell| \leq \frac{1}{k} \right\},$$

$$D = \bigcup_{k \geq 1} \bigcap_{N \geq 1} \bigcup_{n \geq N} \bigcup_{m \geq N} \left\{ \omega \in \Omega : |X_n(\omega) - X_m(\omega)| > \frac{1}{k} \right\},$$

$$L = \left(\bigcap_{\varepsilon > 0} \bigcap_{p \geq 1} \bigcup_{n \geq p} \{ \omega \in \Omega : X_n(\omega) \geq 2 - \varepsilon \} \right) \cap \left(\bigcap_{\varepsilon > 0} \bigcup_{p \geq 1} \bigcap_{n \geq p} \{ \omega \in \Omega : X_n(\omega) \leq 4 + \varepsilon \} \right).$$

b. Faire l'opération de traduction inverse pour les parties suivantes de Ω :

	L'ensemble des $\omega \in \Omega$ tels que la suite $(X_n(\omega))_{n \geq 1} \dots$
<i>E</i>	... soit à termes positifs,
<i>F</i>	... ne soit pas bornée supérieurement,
<i>G</i>	... tende vers $+\infty$,
<i>H</i>	... ne converge pas vers un réel positif,
<i>I</i>	... soit $o(\log n)$,
<i>J</i>	... ne soit ni croissante ni décroissante,
<i>K</i>	... ait une limite inférieure rationnelle.

Exercice III.5.14 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes. On suppose que la série $\sum_{n \geq 1} X_n$ converge presque sûrement. Montrer que pour tout réel $c > 0$, on a $\sum_{n \geq 1} \mathbf{P}(|X_n| > c) < +\infty$.

Exercice III.5.15 a. Soit $(a_n)_{n \geq 1}$ une suite de nombre réels. Montrer que $\limsup a_n = \infty$ si et seulement si pour tout $k \geq 1$ et tout $n_0 \geq 1$ il existe $n \geq n_0$ tel que $a_n \geq k$.

b. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires positives indépendantes et toutes de même loi. On suppose $\mathbf{E}[X_1] = \infty$. Montrer que

$$\mathbf{P} \left(\limsup \frac{X_n}{n} = \infty \right) = 1.$$

Exercice III.5.16 Soit $(A_n)_{n \geq 1}$ une suite d'événements sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbf{P})$. On pose $A_* = \bigcup_{p \geq 1} \bigcap_{n \geq p} A_n$.

a. Montrer que si $\lim_{n \rightarrow +\infty} \mathbf{P}(A_n) = 0$, alors $\mathbf{P}(A_*) = 0$.

b. On pose aussi $A^* = \bigcap_{p \geq 1} \bigcup_{n \geq p} A_n$. On suppose

$$\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = 0 \quad \text{et} \quad \sum_{n=1}^{\infty} \mathbf{P}(A_n \cap A_{n+1}^c) < \infty.$$

Montrer que $\mathbf{P}(A^*) = 0$.

Chapitre IV

Convergence de variables aléatoires. Lois des grands nombres.

IV.1 Resultats généraux de convergence.

IV.1.a Convergence presque sûre.

On note $(\Omega, \mathcal{F}, \mathbf{P})$, l'espace de probabilité de référence pour cette section, espace sur lequel toutes les v.a. mentionnées sont a priori définies, sauf mention explicite du contraire. On munit \mathbb{R}^d d'une norme $\|\cdot\|$ et de la tribu $\mathcal{B}(\mathbb{R}^d)$ des boréliens correspondante. On pose

$$\mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F}) = \{X : \Omega \rightarrow \mathbb{R}^d, \mathcal{F}\text{-mesurable}\}.$$

Définition IV.1.1 ◀▶ (Convergence p.s.) Soient $X_n \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$ et $X \in \mathcal{L}_E(\Omega, \mathcal{F})$. La suite $(X_n)_{n \geq 0}$ converge \mathbf{P} -presque sûrement vers X si et seulement s'il existe $N \in \mathcal{F}$ tel que $\mathbf{P}(N) = 0$ et

$$\forall \omega \in \Omega \setminus N, \quad \lim_{n \rightarrow \infty} \|X_n(\omega) - X(\omega)\| = 0.$$

S'il n'y a pas ambiguïté, on parle simplement de convergence p.s. au lieu de convergence \mathbf{P} -presque sûre. \square

On montre tout d'abord le critère de convergence suivant.

Proposition IV.1.1 ◀▶ Soient X et $X_n \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$. On suppose

$$\forall \varepsilon \in]0, \infty[, \quad \sum_{n \geq 0} \mathbf{P}(\|X - X_n\| > \varepsilon) < \infty.$$

Alors, $\lim_n X_n = X$ presque sûrement.

◀▶ **Preuve** : pour tout $\varepsilon \in]0, \infty[$, le lemme de Borel-Cantelli implique l'existence de $N_\varepsilon \in \mathcal{F}$ tel que $\mathbf{P}(N_\varepsilon) = 0$ et tel pour tout $\omega \in \Omega \setminus N_\varepsilon$, on ait $\|X(\omega) - X_n(\omega)\| \leq \varepsilon$ pour tout n assez grand; donc pour tout $\omega \in \Omega \setminus N_\varepsilon$, $\limsup_n \|X(\omega) - X_n(\omega)\| \leq \varepsilon$. On choisit $(\varepsilon_p)_{p \in \mathbb{N}}$, une suite décroissant strictement vers 0 et on pose $N = \bigcup_{p \in \mathbb{N}} N_{\varepsilon_p}$. On a $N \in \mathcal{F}$ (car \mathcal{F} est en tant que tribu, stable par union dénombrable); on a également $0 \leq \mathbf{P}(N) \leq \sum_{p \in \mathbb{N}} \mathbf{P}(N_{\varepsilon_p}) = 0$, par sigma sous-additivité de la mesure de probabilité \mathbf{P} ; donc $\mathbf{P}(N) = 0$. Comme $\Omega \setminus N = \bigcap_{p \in \mathbb{N}} (\Omega \setminus N_{\varepsilon_p})$, on a

$$\forall \omega \in \Omega \setminus N, \quad \forall p \in \mathbb{N}, \quad \limsup_{n \rightarrow \infty} \|X(\omega) - X_n(\omega)\| \leq \varepsilon_p$$

et donc pour tout $\omega \in \Omega \setminus N$, $\limsup_n \|X(\omega) - X_n(\omega)\| = 0$, ce qui entraîne que $X_n \rightarrow X$ presque sûrement. ■

Dans la preuve de la proposition précédente on a utilisé le principe qu'une intersection dénombrable d'événement presque sûrs et un événement presque sûr, ce que l'on rappelle sous la forme du lemme suivant.

Lemme IV.1.2 ◀▶ (Une intersection dénombrable d'év. presque sûrs est un év. presque sûr) Soient $B_n \in \mathcal{F}$, $n \in \mathbb{N}$ une suite d'événements presque sûrs, c'est-à-dire tels que $\mathbf{P}(B_n) = 1$, $n \in \mathbb{N}$. Alors $B = \bigcap_{n \in \mathbb{N}} B_n$ est un événement \mathbf{P} -presque sûr, c'est-à-dire que $B \in \mathcal{F}$ et $\mathbf{P}(B) = 1$.

◀▶ **Preuve** : on pose $N_n := \Omega \setminus B_n$. On a donc $N_n \in \mathcal{F}$ et $\mathbf{P}(N_n) = 1 - \mathbf{P}(B_n) = 0$. On pose ensuite $N = \bigcup_{n \in \mathbb{N}} N_n$. Alors $N \in \mathcal{F}$. Par sigma sous-additivité de \mathbf{P} , on a $0 \leq \mathbf{P}(N) \leq \sum_{n \in \mathbb{N}} \mathbf{P}(N_n) = 0$. Donc $\mathbf{P}(N) = 0$. Or on a

$$\Omega \setminus N = \Omega \setminus \left(\bigcup_{n \in \mathbb{N}} N_n \right) = \bigcap_{n \in \mathbb{N}} (\Omega \setminus N_n) = \bigcap_{n \in \mathbb{N}} B_n = B .$$

Donc $\mathbf{P}(B) = 1 - \mathbf{P}(N) = 1$. ■

Donnons un exemple d'application de la proposition IV.1.1.

Exemple IV.1.1 (Maximum de v.a. bornées indépendantes) Soit $c \in \mathbb{R}$ et soit $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$, une mesure de probabilité sur \mathbb{R} . On suppose que

$$\mu(]c, \infty[) = 0 \quad \text{et} \quad \forall \varepsilon > 0, \quad \mu([c - \varepsilon, c]) > 0 .$$

Soient $X_n \in \mathcal{L}_{\mathbb{R}}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$, une suite de v.a. que l'on suppose indépendantes et de même loi μ . On pose

$$\forall n \in \mathbb{N}, \quad M_n = \max_{1 \leq k \leq n} X_k .$$

On s'intéresse à la convergence presque sûre des M_n . On remarque tout d'abord que puisque X_n a pour loi μ ,

$$\mathbf{P}(X_n > c) = \mu(]c, \infty[) = 0 ,$$

C'est-à-dire que pour tout $n \in \mathbb{N}$, \mathbf{P} -p.s. $X_n \leq c$. Comme l'intersection des événements presque sûrs $\{X_n \leq c\}$ est encore un événement presque sûr (voir le lemme IV.1.2 qui précède), on a $\mathbf{P}(\bigcap_{n \in \mathbb{N}} \{X_n \leq c\}) = \mathbf{P}(\forall n \in \mathbb{N}, X_n \leq c)$, c'est-à-dire que \mathbf{P} -p.s. pour tout $n \in \mathbb{N}$, $X_n \leq c$. Donc \mathbf{P} -p.s. pour tout $n \in \mathbb{N}$, $M_n \leq c$. Soit $\varepsilon > 0$. On a donc $\mathbf{P}(|M_n - c| > \varepsilon) = \mathbf{P}(M_n < c - \varepsilon)$. On observe ensuite que

$$\begin{aligned} \mathbf{P}(M_n < c - \varepsilon) &= \mathbf{P}(X_1 < c - \varepsilon; X_2 < c - \varepsilon; \dots; X_n < c - \varepsilon) \\ &= \mathbf{P}(X_1 < c - \varepsilon) \mathbf{P}(X_2 < c - \varepsilon) \dots \mathbf{P}(X_n < c - \varepsilon) \\ &= (\mathbf{P}(X_1 < c - \varepsilon))^n . \end{aligned}$$

Or, puisque $\mathbf{P}(X_1 > c) = 0$, on a

$$\mathbf{P}(X_1 < c - \varepsilon) = 1 - \mathbf{P}(X_1 \geq c - \varepsilon) = 1 - \mathbf{P}(c \geq X_1 \geq c - \varepsilon) = 1 - \mu([c - \varepsilon, c]) =: \rho_\varepsilon .$$

Par hypothèse, pour tout $\varepsilon > 0$, on a $0 \leq \rho_\varepsilon < 1$. On donc montré que

$$\forall \varepsilon \in]0, \infty[, \quad \sum_{n \in \mathbb{N}} \mathbf{P}(|M_n - c| > \varepsilon) = \sum_{n \in \mathbb{N}} \rho_\varepsilon^n = \frac{1}{1 - \rho_\varepsilon} < \infty ,$$

et la proposition IV.1.1, page 125, implique que \mathbf{P} -p.s. $\lim_{n \rightarrow \infty} M_n = c$. □

On donne ensuite le critère suivant permettant de démontrer l'existence d'une limite presque sûre sans connaître cette limite (il s'agit d'une sorte de critère de Cauchy pour la convergence presque sûre).

Proposition IV.1.3 *Soit $\varepsilon_n \in]0, \infty[$, $n \in \mathbb{N}$, une suite telle que*

$$\sum_{n \geq 0} \varepsilon_n < \infty .$$

Soient $X_n \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$. On suppose

$$\sum_{n \geq 0} \mathbf{P}(\|X_{n+1} - X_n\| > \varepsilon_n) < \infty .$$

Alors, il existe $X \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$ telle que $\lim_n X_n = X$ presque sûrement.

Preuve : le lemme de Borel-Cantelli implique l'existence de $B \in \mathcal{F}$ tel que $\mathbf{P}(B) = 1$ et tel pour tout $\omega \in B$, on ait $\|X_{n+1} - X_n\| \leq \varepsilon_n$ pour tout $n \geq n_0(\omega)$. L'inégalité triangulaire implique que pour tous $m, n \geq p \geq n_0(\omega)$, on a

$$\|X_n(\omega) - X_m(\omega)\| \leq \sum_{k \geq p} \varepsilon_k .$$

Comme $\sum_{n \in \mathbb{N}} \varepsilon_n < \infty$, pour tout $\omega \in B$, la suite $(X_n(\omega))_{n \in \mathbb{N}}$ est de Cauchy. Puisque \mathbb{R}^d est complet, il existe $X(\omega) \in E$ tel que $\lim_n \|X_n(\omega) - X(\omega)\| = 0$. On fixe ensuite $x_0 \in E$ et pour tout $\omega \in \Omega \setminus B$, on pose $X(\omega) = x_0$. On a donc montré que $\lim_n X_n = X$ presque sûrement car $\mathbf{P}(B) = 1$.

Il reste à montrer que X est $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$ -mesurable. Pour cela, on rappelle que les fonctions continues bornées sur \mathbb{R}^d engendrent la tribu Borélienne $\mathcal{B}(\mathbb{R}^d)$. Cela implique que si pour toute fonction continue bornée $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $f(X)$ est \mathcal{F} -mesurable alors X est $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$ -mesurable. Or on a \mathbf{P} -p.s. $\lim_n f(X_n) = f(X)$, ce qui implique que $f(X)$ est \mathcal{F} -mesurable une limite presque sûre de v.a. réelles \mathcal{F} -mesurable (les $f(X_n)$ le sont) est \mathcal{F} -mesurable. Cela termine la preuve de la proposition. ■

IV.1.b Convergence en probabilité.

On note $(\Omega, \mathcal{F}, \mathbf{P})$, l'espace de probabilité de référence pour cette section, espace sur lequel toutes les v.a. mentionnées sont a priori définies, sauf mention explicite du contraire. On munit \mathbb{R}^d d'une norme $\|\cdot\|$ et de la tribu $\mathcal{B}(\mathbb{R}^d)$ des boréliens correspondante. On pose

$$\mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F}) = \{X : \Omega \rightarrow \mathbb{R}^d, \mathcal{F}\text{-mesurable}\} .$$

Définition IV.1.2 ◀▶ (Convergence en probabilité) Soient $X_n \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$ et $X \in \mathcal{L}_E(\Omega, \mathcal{F})$. La suite $(X_n)_{n \geq 0}$ converge en \mathbf{P} -probabilité vers X ssi

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}(\|X_n - X\| > \varepsilon) = 0.$$

S'il n'y a pas ambiguïté, on parle simplement de convergence en probabilité, plutôt que de convergence en \mathbf{P} -probabilité. \square

Le théorème suivant fait le lien entre la convergence en probabilité et la convergence presque sûre.

Théorème IV.1.4 ◀▶ Soient $X_n \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$ et $X \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$. Les assertions suivantes sont vérifiées

- (i) Si $\lim_n X_n = X$ p.s., alors $\lim_n X_n = X$ en probabilité.
- (ii) Si $\lim_n X_n = X$ en probabilité, il existe une suite strictement croissante d'indices $(n_k)_{k \geq 0}$ tels que $\lim_k X_{n_k} = X$ presque sûrement.
- (iii) On suppose que $\lim_n X_n = X$ en probabilité. Soit $Y \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$ tel que l'on ait aussi $\lim_n X_n = Y$ en probabilité. Alors $X = Y$ presque sûrement.

◀▶ **Preuve :** on montre (i). Supposons que $\lim_n X_n = X$ presque sûrement. On pose

$$Y_n = \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}}.$$

Alors $0 \leq Y_n \leq 1$ et p.s. $\lim_n Y_n = 0$. Par convergence dominée, $\lim_n \mathbf{E}[Y_n] = 0$, c'est-à-dire $\lim_n \mathbf{P}(\|X_n - X\| > \varepsilon) = 0$. Comme ε peut être arbitrairement petit, cela montre que $\lim_n X_n = X$ en probabilité.

Montrons (ii). Supposons que $\lim_n X_n = X$ en probabilité. Par définition, pour tout $k \in \mathbb{N}$, il existe $p_k \in \mathbb{N}$, tel que pour tout $n \geq p_k$, on ait $\mathbf{P}(\|X_n - X\| > 2^{-k}) \leq 2^{-k}$. On pose $n_k = p_1 + \dots + p_k$ et on a bien $\mathbf{P}(\|X_{n_k} - X\| > 2^{-k}) \leq 2^{-k}$, et donc

$$\sum_{k \in \mathbb{N}} \mathbf{P}(\|X_{n_k} - X\| > 2^{-k}) < \infty.$$

Le lemme de Borel-Cantelli implique p.s. que pour tout k suffisamment grand, $\|X_{n_k} - X\| \leq 2^{-k}$ et donc $\lim_k X_{n_k} = X$ presque sûrement, ce qui prouve (ii).

Le point (iii) est une conséquence immédiate du (ii). Une preuve directe sans utiliser (ii) est bien entendu possible (nous laissons cela au lecteur). \blacksquare

Exemple IV.1.2 La convergence en probabilité n'entraîne pas nécessairement la convergence presque sûre comme le montre le contreexemple suivant. On choisit

$$(\Omega, \mathcal{F}, \mathbf{P}) := ([0, 1], \mathcal{B}([0, 1]), \ell).$$

On pose $I_0 = [0, 1]$ et pour tout entier n tel que

$$1 + 2 + \dots + 2^{p-1} = 2^p - 1 < n \leq 2^{p+1} - 1 = 1 + 2 + \dots + 2^p \quad \text{on pose} \quad I_n = [2^{-p}n - 1, 2^{-p}(n+1) - 1].$$

On vérifie que les I_n sont des sous-intervalles dyadiques de $[0, 1]$ et que la suite des I_n les parcourt tous par ordre croissant pour la longueur et, à longueur égale, de gauche à droite. On pose $X_n = \mathbf{1}_{I_n}$. Il est clair que pour tout $\varepsilon > 0$,

$$\mathbf{P}(X_n > \varepsilon) = \ell(I_n) = 2^{-\lfloor \log_2 n \rfloor} \xrightarrow[n \rightarrow \infty]{} 0.$$

Donc $\lim_n X_n = 0$ en probabilité. Par ailleurs, tout $\omega \in \Omega = [0, 1]$ appartient à une infinité de I_n et n'appartient pas à une infinité de I_n . Donc

$$\forall \omega \in \Omega, \quad \limsup_{n \rightarrow \infty} X_n(\omega) = 1 \quad \text{et} \quad \limsup_{n \rightarrow \infty} X_n(\omega) = 0.$$

La suite $(X_n)_{n \in \mathbb{N}}$ ne converge donc pas presque sûrement. □

Proposition IV.1.5 ◀ On munit \mathbb{R}^{d_1} d'une norme notée $\|\cdot\|_1$ et \mathbb{R}^{d_2} d'une norme $\|\cdot\|_2$. On munit $\mathbb{R}^{d_1+d_2} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ de la norme

$$\forall (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, \quad \|(\mathbf{x}_1, \mathbf{x}_2)\| := \max(\|\mathbf{x}_1\|_1, \|\mathbf{x}_2\|_2).$$

Soient $X_n \in \mathcal{L}_{\mathbb{R}^{d_1}}(\Omega, \mathcal{F})$, $X'_n \in \mathcal{L}_{\mathbb{R}^{d_2}}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$. Soient $X \in \mathcal{L}_{\mathbb{R}^{d_1}}(\Omega, \mathcal{F})$ et $X' \in \mathcal{L}_{\mathbb{R}^{d_2}}(\Omega, \mathcal{F})$. Alors, il y a équivalence entre les deux assertions suivantes.

- (a) $\lim_n X_n = X$ et $\lim_n X'_n = X'$ en probabilité.
- (b) $\lim_n (X_n, X'_n) = (X, X')$ en probabilité dans $\mathbb{R}^{d_1+d_2}$.

Preuve : pour tout $\varepsilon > 0$, on remarque d'abord que

$$\mathbf{P}(\|(X_n, X'_n) - (X, X')\| > \varepsilon) \leq \mathbf{P}(\|X_n - X\|_1 > \varepsilon) + \mathbf{P}(\|X'_n - X'\|_2 > \varepsilon),$$

ce qui montre que (a) \implies (b). On vérifie également que

$$\max(\mathbf{P}(\|X_n - X\|_1 > \varepsilon), \mathbf{P}(\|X'_n - X'\|_2 > \varepsilon)) \leq \mathbf{P}(\|(X_n, X'_n) - (X, X')\| > \varepsilon),$$

qui entraîne que (b) \implies (a). ■

On s'intéresse ensuite aux propriétés de continuité de la convergence en probabilité. Plus précisément, On munit \mathbb{R}^{d_1} d'une norme notée $\|\cdot\|_1$ et \mathbb{R}^{d_2} d'une norme $\|\cdot\|_2$; soit $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$, une fonction $(\mathcal{B}(\mathbb{R}^{d_1}), \mathcal{B}(\mathbb{R}^{d_2}))$ -mesurable. On définit le *module de continuité* de f en $x \in \mathbb{R}^{d_1}$ de la manière suivante : pour tout réel $\delta > 0$, on pose

$$w_f(x, \delta) := \sup \{ \|f(y) - f(z)\|_2; y, z \in B(x, \delta) \},$$

où $B(x, \delta) := \{y \in \mathbb{R}^{d_1} : \|x - y\|_1 < \delta\}$ est la boule ouverte de centre x et de rayon δ . On rappelle que

$$f \text{ est continue en } x \text{ ssi } \lim_{\delta \rightarrow 0} w_f(x, \delta) = 0.$$

On montre le lemme suivant.

Lemme IV.1.6 Soit $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$, une fonction $(\mathcal{B}(\mathbb{R}^{d_1}), \mathcal{B}(\mathbb{R}^{d_2}))$ -mesurable. Alors les assertions suivantes sont vérifiées.

- (i) Pour tout $\delta > 0$, la fonction $x \in \mathbb{R}^{d_1} \mapsto w_f(x, \delta) \in [0, \infty]$ est $\mathcal{B}(\mathbb{R}^{d_1})$ -mesurable.
(ii) $C := \{x \in \mathbb{R}^{d_1} : f \text{ est continue en } x\} \in \mathcal{B}(\mathbb{R}^{d_1})$.

Preuve : pour tous réels $a, \delta > 0$, on pose

$$U_{a,\delta} = \{x \in \mathbb{R}^{d_1} : \exists y, z \in B(x, \delta) \text{ tels que } \|f(y) - f(z)\|_2 > a\}.$$

On voit que $U_{a,\delta} = \{x \in \mathbb{R}^{d_1} : w_f(x, \delta) > a\}$. On vérifie que $U_{a,\delta}$ est un ouvert. En effet, soit $x \in U_{a,\delta}$. Il existe $y, z \in B(x, \delta)$ tels que $\|f(y) - f(z)\|_2 > a$. On pose $\eta := \min(\delta - \|x - y\|_1, \delta - \|x - z\|_1)$. Donc $\eta > 0$, et pour tout $x_* \in B(x, \eta)$, on a bien $\|x_* - y\|_1 \leq \|x_* - x\|_1 + \|x - y\|_1 < \delta$. De même, $\|x_* - z\|_1 < \delta$. Donc, $y, z \in B(x_*, \delta)$ et on a bien $B(x_*, \eta) \subset U_{a,\delta}$. Cela montre (i).

Pour montrer (ii), on pose pour tout $x \in \mathbb{R}^{d_1}$, $g(x) := \limsup_p w_f(x, 2^{-p})$. Par (i), c'est une application $\mathcal{B}(\mathbb{R}^{d_1})$ -mesurable. On remarque ensuite que pour tout $p \in \mathbb{N}$, et tout $x \in \mathbb{R}^{d_1}$, $w_f(x, 2^{-p-1}) \leq w_f(x, 2^{-p})$. Donc

$$C = \{x \in \mathbb{R}^{d_1} : f \text{ est continue en } x\} = \{x \in \mathbb{R}^{d_1} : \lim_{\delta \rightarrow 0} w_f(x, \delta) = 0\} = \{g = 0\},$$

ce qui termine la preuve. ■

Proposition IV.1.7 ◀▶ Soit $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$, une fonction mesurable. Soient $X_n \in \mathcal{L}_{\mathbb{R}^{d_1}}(\Omega, \mathcal{F})$, $n \in \mathbb{N}$ et $X \in \mathcal{L}_{\mathbb{R}^{d_1}}(\Omega, \mathcal{F})$. On fait les hypothèses suivantes.

- (a) $\lim_n X_n = X$ en probabilité.
(b) Presque sûrement f est continue en X .

Alors $\lim_n f(X_n) = f(X_\infty)$ en probabilité.

Preuve : on munit \mathbb{R}^{d_1} de la norme $\|\cdot\|_1$ et \mathbb{R}^{d_2} de la norme $\|\cdot\|_2$. Pour tous réels $\varepsilon, \delta > 0$, on vérifie que

$$\begin{aligned} \mathbf{P}(\|f(X_n) - f(X)\|_2 > \varepsilon) &= \mathbf{P}(\|f(X_n) - f(X)\|_2 > \varepsilon; \|X_n - X\|_1 < \delta) \\ &\quad + \mathbf{P}(\|f(X_n) - f(X)\|_2 > \varepsilon; \|X_n - X\|_1 \geq \delta) \\ &\leq \mathbf{P}(w_f(X, \delta) > \varepsilon) + \mathbf{P}(\|X_n - X\|_1 > \delta/2). \end{aligned}$$

Donc pour tous $\varepsilon, \delta > 0$, l'hypothèse (a) implique que

$$\limsup_{n \rightarrow \infty} \mathbf{P}(\|f(X_n) - f(X)\|_2 > \varepsilon) \leq \mathbf{P}(w_f(X, \delta) > \varepsilon).$$

Pour tout $p \in \mathbb{N}$, on pose $A_p = \{w_f(X, 2^{-p}) > \varepsilon\}$. On a donc $A_{p+1} \subset A_p$ et

$$A := \bigcap_{p \in \mathbb{N}} A_p \subset \{f \text{ n'est pas continue en } X\}.$$

Donc $\lim_p \mathbf{P}(A_p) = \mathbf{P}(A) = 0$ par l'hypothèse (b), ce qui permet de conclure. ■

On déduit des deux propositions précédentes les propriétés de continuité suivantes de la convergence en probabilité (bien d'autres énoncés du même genre en découlent).

Proposition IV.1.8 ◀► Soient $X_n, Y_n, X, Y \in \mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$. On suppose que $\lim_{n \rightarrow \infty} X_n = X$ en probabilité et $\lim_{n \rightarrow \infty} Y_n = Y$ en probabilité. Alors les assertions suivantes sont vérifiées.

- (i) $\lim_n X_n + Y_n = X + Y$ en probabilité.
- (ii) Si $d=1$ ou bien $d=2$ avec $\mathbb{R}^2 \equiv \mathbb{C}$, alors $\lim_n X_n Y_n = XY$ en probabilité.
- (iii) Si $d=1$ ou bien $d=2$ avec $\mathbb{R}^2 \equiv \mathbb{C}$ et si p.s. $Y \neq 0$, alors $\lim_n X_n/Y_n = X/Y$ en probabilité.
- (iv) Si $d=1$, alors $\lim_n X_n \wedge Y_n = X \wedge Y$ et $\lim_n X_n \vee Y_n = X \vee Y$ en probabilité.

Preuve : la proposition IV.1.5 implique que $\lim_n (X_n, Y_n) = (X, Y)$ en probabilité dans \mathbb{R}^{2d} . On applique ensuite la proposition IV.1.7 à $f(x, y) = x + y$ pour obtenir (i), à $f(x, y) = xy$ pour obtenir (ii), à $f(x, y) = x/y$ qui est continue sur $\mathbb{C} \times \mathbb{C}^*$ pour obtenir (iii), et enfin à $f(x, y) = x \wedge y$ et $f(x, y) = x \vee y$ pour obtenir (iv). ■

IV.2 Convergence L^p .

Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Dans tout la section \mathbb{K} désigne \mathbb{R} ou \mathbb{C} et on note

$$\mathcal{L}_{\mathbb{K}}(\Omega, \mathcal{F}) = \{X : \Omega \rightarrow \mathbb{K}, \mathcal{F}\text{-mesurable}\}.$$

Pour tout $p \in [1, \infty[$, on pose

$$\mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P}) = \{X \in \mathcal{L}_{\mathbb{K}}(\Omega, \mathcal{F}) : \mathbf{E}[|X|^p] < \infty\}.$$

$\mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$ est l'espace des v.a. admettant un moment d'ordre p . On définit également,

$$\forall X \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P}), \quad \|X\|_p := (\mathbf{E}[|X|^p])^{\frac{1}{p}}.$$

Proposition IV.2.1 ◀► On conserve les notations ci-dessus et on rappelle que $p \in [1, \infty[$. Alors les propriétés suivantes sont vérifiées pour tout $X, Y \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$.

- (i) Pour tout $c \in \mathbb{K}$, $\|c \cdot X\|_p = |c| \cdot \|X\|_p$.
- (ii) (Inégalité triangulaire) $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$.
- (iii) Si $\|X\|_p = 0$, alors $X = 0$ \mathbf{P} -p.s.

Cela montre que $\mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$ est un espace vectoriel et que $\|\cdot\|_p$ est une pseudonorme sur cet espace.

Preuve : les preuves de (i) et de (iii) sont immédiates (la preuve de (iii) provient de la proposition II.2.7 (i), page 41, qui dans sa version probabiliste dit que si une variable positive est d'espérance nulle, alors elle est nulle presque sûrement). La preuve de (ii) est une conséquence de l'inégalité de Minkowski (proposition II.4.4, page 55). Les autres affirmations de la proposition découlent immédiatement des propriétés (i), (ii) et (iii). ■

Définition IV.2.1 ◀▶ (Convergence L^p) Soit $p \in [1, \infty[$. Soient $X_n, X \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$, $n \in \mathbb{N}$. On dit la suite $(X_n)_{n \in \mathbb{N}}$ converge au sens L^p vers X si

$$\lim_{n \rightarrow \infty} \mathbf{E}[|X_n - X|^p] = 0,$$

ce qui est équivalent à $\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$. □

Soient $X, Y \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$. L'inégalité triangulaire pour $\|\cdot\|_p$ (c'est-à-dire le point (ii) de la proposition IV.2.1, page 131) implique que

$$\|X\|_p = \|X - Y + Y\|_p \leq \|X - Y\|_p + \|Y\|_p$$

De même on a $\|Y\|_p \leq \|X - Y\|_p + \|X\|_p$. Donc

$$\forall X, Y \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P}), \quad \left| \|X\|_p - \|Y\|_p \right| \leq \|X - Y\|_p. \quad (\text{IV.1})$$

Il s'agit d'un fait général pour les normes : une simple conséquence de l'inégalité triangulaire. On en déduit le lemme suivant.

Lemme IV.2.2 ◀▶ Soit $p \in [1, \infty[$. Soient $X_n, X \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$, $n \in \mathbb{N}$ tel que $X_n \rightarrow X$ au sens L^p . Alors, $\lim_n \mathbf{E}[|X_n|^p] = \mathbf{E}[|X|^p]$, ce qui est équivalent à $\|X_n\|_p \rightarrow \|X\|_p$ lorsque $n \rightarrow \infty$.

◀▶ **Preuve :** (IV.1) implique que $\left| \|X_n\|_p - \|X\|_p \right| \leq \|X_n - X\|_p$ qui implique immédiatement le résultat désiré. ■

On pose ensuite

$$\mathcal{N} = \{X : \Omega \rightarrow \mathbb{K} \text{ } \mathcal{F}\text{-mesurable tel que } \mathbf{P}\text{-p.s. } X = 0\}.$$

Il est clair que \mathcal{N} est un sous-espace vectoriel des espaces $\mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$, $p \in [1, \infty[$. Soient $X, Y \in \mathcal{L}_{\mathbb{K}}^p(\Omega, \mathcal{F}, \mathbf{P})$. On note $X \sim Y$ la relation \mathbf{P} -p.s. $X = Y$, c'est-à-dire $X - Y \in \mathcal{N}$. Clairement il s'agit d'une relation d'équivalence. On définit alors le \mathbb{K} -espace vectoriel quotient

$$L^p(\Omega, \mathcal{F}, \mathbf{P}) = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P}) / \sim.$$

L'élément nul, simplement noté 0, est la classe d'équivalence de 0, c'est-à-dire l'ensemble des fonctions nulles \mathbf{P} -presque sûrement. Par ailleurs, si $X \sim Y$, alors $\|X\|_p = \|Y\|_p$: cela induit une fonction sur $L^p(\Omega, \mathcal{F}, \mathbf{P})$, que l'on note également $\|\cdot\|_p$. On remarque que si $\|X\|_p = 0$, alors la classe de \sim -équivalence de X est l'élément nul de $L^p(\Omega, \mathcal{F}, \mathbf{P})$.

Convention : un élément \tilde{X} de $L^p(\Omega, \mathcal{F}, \mathbf{P})$ n'est pas une fonction mais une classe de fonctions. Cependant dans la pratique, on le confond systématiquement avec n'importe lequel de ses représentants $X \in \tilde{X}$. □

Proposition IV.2.3 ◀▶ Soit $p \in]1, \infty[$. On note q son exposant conjugué.

- (i) Pour tous $X \in L^p(\Omega, \mathcal{F}, \mathbf{P})$ et $Y \in L^q(\Omega, \mathcal{F}, \mathbf{P})$, on a $XY \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ et $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.
- (ii) $\|\cdot\|_p$ est une norme sur $L^p(\Omega, \mathcal{F}, \mathbf{P})$.

Preuve : le (i) est une conséquence de l'inégalité de Hölder (proposition II.4.3, page 55) et (ii) est une conséquence de la proposition IV.2.1, page 131, et du passage au quotient. ■

On rappelle la conséquence suivante des inégalités de Jensen (voir l'inégalité (II.24) page 54.

Proposition IV.2.4 ◀▶ Soient $p_1, p_2 \in [1, \infty[$ tels que $p_1 \leq p_2$.

$$\forall X \in L^{p_2}(\Omega, \mathcal{F}, \mathbf{P}), \quad \|X\|_{p_1} \leq \|X\|_{p_2}$$

et donc $L^{p_2}(\Omega, \mathcal{F}, \mathbf{P}) \subset L^{p_1}(\Omega, \mathcal{F}, \mathbf{P})$.

La proposition suivante implique que la convergence au sens L^p implique la convergence en probabilité.

Proposition IV.2.5 ◀▶ ($L^p \Rightarrow \text{proba.}$) Soit $p \in [1, \infty[$. Soient $X_n, X \in L^p(\Omega, \mathcal{F}, \mathbf{P})$, $n \in \mathbb{N}$. Alors

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0 \quad \implies \quad X_n \xrightarrow[n \rightarrow \infty]{\text{proba.}} X.$$

◀▶ **Preuve :** l'inégalité suivante, de type Markov, implique que pour tout $\varepsilon > 0$ et tout $n \in \mathbb{N}$,

$$\mathbf{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-p} \mathbf{E}[|X_n - X|^p] = \varepsilon^{-p} \|X_n - X\|_p^p,$$

ce qui implique le résultat. ■

En combinant le résultat précédent et le théorème IV.1.4 (ii), on obtient immédiatement le résultat suivant, qui a été déjà prouvé par une méthode différente.

Proposition IV.2.6 ◀▶ Soit $p \in [1, \infty[$. Soient $X_n, X \in L^p(\Omega, \mathcal{F}, \mathbf{P})$. On suppose que $\lim_n \|X_n - X\|_p = 0$. Alors, il existe une suite strictement croissante d'indices $(n_k)_{k \geq 0}$ tels que \mathbf{P} -p.s. $\lim_k X_{n_k} = X$.

On rappelle le théorème suivant.

Théorème IV.2.7 ◀▶ Pour tout $p \in [1, \infty[$, $L^p(E, \mathcal{E}, \mu)$ muni de la norme $\|\cdot\|_p$ est complet.

Preuve : soient $X_n \in L^p(\Omega, \mathcal{F}, \mathbf{P})$, $n \in \mathbb{N}$, que l'on suppose de Cauchy, c'est-à-dire que

$$\sup_{m_1, m_2 \geq m} \|X_{m_1} - X_{m_2}\|_p \xrightarrow{m \rightarrow \infty} 0.$$

Soit $m(k)$ tel que pour tous $m_1, m_2 \geq m(k)$, on ait $\|X_{m_1} - X_{m_2}\|_p \leq 2^{-2k}$. On pose $n_k = m(0) + \dots + m(k) + k$, alors $(n_k)_{k \in \mathbb{N}}$ est une suite d'entiers strictement croissante telle que

$$\forall m_1, m_2 \geq n_k, \quad \|X_{m_1} - X_{m_2}\|_p \leq 2^{-2k}.$$

Par une inégalité de type Markov, on en déduit notamment que

$$\mathbf{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) \leq 2^{pk} \mathbf{E}[|X_{n_{k+1}} - X_{n_k}|^p] = 2^{pk} \|X_{n_{k+1}} - X_{n_k}\|_p^p \leq 2^{pk} 2^{-2pk} = 2^{-pk}.$$

On a donc $\mathbf{P}(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) < \infty$. Comme $\sum_{k \in \mathbb{N}} 2^{-k}$, la proposition IV.1.3 (voir page 127) s'applique et montre qu'il existe une v.a. X telle que \mathbf{P} -p.s. $X_{n_k} \rightarrow X$ lorsque $k \rightarrow \infty$.

On remarque ensuite que pour tous $k, \ell \in \mathbb{N}$ et pour tout $n \leq n_k$,

$$\mathbf{E}[|X_{n_{k+\ell}} - X_n|^p] \leq \|X_{n_{k+\ell}} - X_n\|_p^p \leq 2^{-pk}.$$

On fixe k et n tels que $n \geq n_k$. Lorsque $\ell \rightarrow \infty$, $X_{n_{k+\ell}} \rightarrow X$ presque sûrement et par Fatou, on a

$$\mathbf{E}[|X - X_n|^p] = \mathbf{E}\left[\liminf_{\ell \rightarrow \infty} |X_{n_{k+\ell}} - X_n|^p\right] \leq \liminf_{\ell \rightarrow \infty} \mathbf{E}[|X_{n_{k+\ell}} - X_n|^p] \leq 2^{-pk}.$$

On a donc pour tout $k \in \mathbb{N}$,

$$\forall n \geq n_k, \quad \mathbf{E}[|X - X_n|^p] \leq 2^{-pk},$$

ce qui implique que $\lim_{n \rightarrow \infty} \mathbf{E}[|X - X_n|^p] = 0$, et donc $X_n \rightarrow X$ au sens L^p . ■

IV.3 Lois faibles des grands nombres.

On note $(\Omega, \mathcal{F}, \mathbf{P})$, l'espace de probabilité de référence pour cette section, espace sur lequel toutes les v.a. mentionnées sont a priori définies, sauf mention explicite du contraire.

Cette section est consacrée à convergence en probabilité de la somme de variables indépendantes ou faiblement dépendantes, essentiellement à l'aide de calculs de moment d'ordre 2. L'énoncé le plus simple est la loi faible des grands nombres.

Théorème IV.3.1 ◀▶ (Loi faible des grands nombres) Soit $(X_n)_{n \geq 1}$, une suite de v.a. réelles indépendantes de même loi. On pose $m = \mathbf{E}[X_n]$ et on suppose que ces v.a. admettent un moment d'ordre 2. Pour tout $n \geq 1$, on pose $S_n = X_1 + \dots + X_n$. Alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n = m \text{ en probabilité.}$$

▶▶ **Preuve :** la proposition III.2.12 implique que $\mathbf{var}(S_n) = n \mathbf{var}(X_1)$. L'inégalité de Tchebychev implique alors que

$$\mathbf{P}\left(\left|\frac{1}{n} S_n - m\right| > \varepsilon\right) = \mathbf{P}(|S_n - nm| > n\varepsilon) \leq \frac{\mathbf{var}(S_n)}{n^2 \varepsilon^2} = \frac{\mathbf{var}(X_1)}{n \varepsilon^2}, \quad (\text{IV.2})$$

ce qui implique immédiatement le résultat désiré. ■

Exemple IV.3.1 Donnons un exemple d'application à l'analyse. Soit $f : [0, 1] \rightarrow \mathbb{R}$, une fonction continue. On lui associe son *polynôme de Bernstein d'ordre n* donné par

$$Q_n(p) = \sum_{0 \leq k \leq n} f(k/n) \binom{n}{k} p^k (1-p)^{n-k}.$$

On fixe $p \in [0, 1]$ et on se donne $(X_n)_{n \geq 1}$, une suite de v.a. de Bernoulli indépendantes de paramètre p . On pose $S_n = X_1 + \dots + X_n$. On remarque alors que

$$Q_n(p) = \mathbf{E}[f(S_n/n)].$$

On note w le module d'uniforme continuité de f :

$$\forall \varepsilon \in \mathbb{R}_+, \quad w(\varepsilon) = \sup \{ |f(x) - f(y)| ; x, y \in [0, 1] : |x - y| \leq \varepsilon \} .$$

Puisque f est continue sur le compact $[0, 1]$, elle est uniformément continue, c'est-à-dire $\lim_0 w(\varepsilon) = 0$. Pour tout $\varepsilon > 0$, l'inégalité (IV.2) implique

$$\begin{aligned} |Q_n(p) - f(p)| &\leq \mathbf{E}[|f(S_n/n) - f(p)|] \\ &\leq \mathbf{E}[|f(S_n/n) - f(p)| \mathbf{1}_{\{|\frac{1}{n}S_n - p| \leq \varepsilon\}}] + \mathbf{E}[|f(S_n/n) - f(p)| \mathbf{1}_{\{|\frac{1}{n}S_n - p| > \varepsilon\}}] \\ &\leq w(\varepsilon) + 2\|f\|_\infty \mathbf{P}(|\frac{1}{n}S_n - p| > \varepsilon) \leq w(\varepsilon) + 2\|f\|_\infty \frac{\mathbf{var}(X_1)}{n\varepsilon^2} . \end{aligned}$$

Or $\mathbf{var}(X_1) = p(1 - p) \leq 1/4$. On a donc $|Q_n(p) - f(p)| \leq w(\varepsilon) + \|f\|_\infty / (2\varepsilon^2 n)$, qui est une majoration indépendante de p . Cela implique donc

$$\forall n \geq 1, \forall \varepsilon > 0, \quad \|Q_n - f\|_\infty \leq w(\varepsilon) + \frac{\|f\|_\infty}{2\varepsilon^2 n} .$$

En prenant par exemple $\varepsilon = n^{-1/4}$, on voit que

$$\lim_{n \rightarrow \infty} \|Q_n - f\|_\infty = 0 ,$$

ce qui prouve le théorème de Weierstraß.

Théorème IV.3.2 (Théorème d'approximation de Weierstraß) *Les fonctions polynômiales de $[0, 1]$ dans \mathbb{R} sont denses pour la norme uniforme dans l'espace des fonctions continues de $[0, 1]$ dans \mathbb{R} .* \square

On peut améliorer le théorème IV.3.1 en affaiblissement les hypothèses de plusieurs manières : on peut supposer les v.a. non-corrélées. Les v.a. peuvent également avoir des lois différentes. En réalité, il s'agit d'une convergence L^2 , qui implique une convergence en probabilité. L'idée de base est le lemme élémentaire suivant.

Lemme IV.3.3 \blacktriangleleft *Soit $(S_n)_{n \geq 1}$, une suite de v.a. réelles admettant un moment d'ordre 2. On suppose que*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{var}(S_n)}{(\mathbf{E}[S_n])^2} = 0 \quad \implies \quad \lim_{n \rightarrow \infty} \frac{S_n}{\mathbf{E}[S_n]} = 1 \text{ en probabilité.}$$

\blacktriangleleft **Preuve :** on a $\mathbf{E}[(S_n/\mathbf{E}[S_n] - 1)^2] = \mathbf{var}(S_n)/(\mathbf{E}[S_n])^2 \rightarrow 0$. On a donc une convergence L^2 donc en probabilité. \blacksquare

Exemple IV.3.2 Le collectionneur d'images. Un enfant collectionne des images d'animaux qu'il trouve dans des paquets de céréales : il y a N sortes d'images et chaque paquet nouveau contient une image. La fréquence des images est la même si bien que l'image contenu dans un paquet a une probabilité $1/N$ d'être du type $k \in \{1, \dots, N\}$. On s'intéresse au nombre T_N de paquets que l'enfant a consommés pour obtenir la collection complète des N images. Plus généralement, pour tout $k \in \{1, \dots, N\}$, on note T_k le nombre de paquets consommés par l'enfant pour obtenir k images différentes.

Il est facile de voir que les variables $T_{k+1} - T_k$, $1 \leq k \leq N - 1$ sont indépendantes et que pour tout $n \geq 1$,

$$\mathbf{P}(T_{k+1} - T_k = n) = \frac{N-k}{N} \left(\frac{k}{N}\right)^{n-1} .$$

Autrement dit, $T_{k+1} - T_k - 1$ est une variable géométrique de paramètre de succès $\frac{N-k}{N}$. On a donc $\mathbf{E}[T_{k+1} - T_k] = \frac{N}{N-k}$. Comme $T_1 = 1$, on a

$$\mathbf{E}[T_N] = 1 + \sum_{1 \leq k \leq N-1} \mathbf{E}[T_{k+1} - T_k] = N \sum_{1 \leq k \leq N} \frac{1}{k} \sim N \log N .$$

Si $N = 20$, il faut consommer en *moyenne* $20 \log 20 \approx 26$ paquets pour obtenir la collection complète d'images.

Précisons la signification du terme « *en moyenne* ». On a montré que $T_N = 1 + G_1 + \dots + G_{N-1}$, où les variables G_k sont indépendantes de lois géométriques données par

$$\forall n \geq 1, \quad \mathbf{P}(G_k = n) = \frac{N-k}{N} \left(\frac{k}{N}\right)^{n-1} .$$

Par la proposition III.2.12, on a

$$\mathbf{var}(T_N) = \sum_{1 \leq k \leq N-1} \mathbf{var}(G_k) = \sum_{1 \leq k \leq N-1} \frac{Nk}{(N-k)^2} \leq N^2 \sum_{1 \leq k \leq N-1} \frac{1}{k^2} \leq \zeta(2)N^2 .$$

On en déduit que $\lim_N \mathbf{var}(T_N)/(\mathbf{E}[T_N])^2 = 0$ et donc

$$\lim_{N \rightarrow \infty} \frac{T_N}{N \log N} = 1 \text{ en probabilité.}$$

On peut bien entendu améliorer ce résultat par des techniques similaires. □

IV.4 Loi Forte des grands nombres.

IV.4.a Loi du 0-1 de Kolmogorov.

Sauf mention du contraire, l'espace de probabilité de référence est noté $(\Omega, \mathcal{F}, \mathbf{P})$.

Définition IV.4.1 ◀▶ (*Tribu P-triviale*) Soit \mathcal{T} une tribu sur Ω telle que $\mathcal{T} \subset \mathcal{F}$. On dit que \mathcal{T} est *P-triviale* si pour tout $A \in \mathcal{T}$, $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$. □

Lemme IV.4.1 ◀▶ Soit \mathcal{T} , une sous-tribu *P-triviale* de \mathcal{F} . Soit X , une v.a. réelle \mathcal{T} -mesurable. Alors, il existe $y_0 \in \mathbb{R}$ tel que p.s. $X = y_0$.

Preuve : pour tout $y \in \mathbb{R}$, $\mathbf{P}(X \leq y) = 0$ ou 1 . On note $I := \{y \in \mathbb{R} : \mathbf{P}(X \leq y) = 1\}$. Il est clair que c'est un intervalle infini à droite. On note y_0 son extrémité gauche. Alors $\mathbf{P}(X \leq y_0 + 2^{-n}) = 1$ et par monotonie séquentielle $\mathbf{P}(X \leq y) = \lim_{n \rightarrow \infty} \mathbf{P}(X \leq y_0 + 2^{-n}) = 1$. Donc $I = [y_0, \infty[$. Pour tout $y < y_0$, par définition de y_0 , on a $\mathbf{P}(X \leq y) = 0$ et par monotonie séquentielle $\mathbf{P}(X < y_0) = \lim_{n \rightarrow \infty} \mathbf{P}(X \leq y_0 - 2^{-n}) = 0$. Donc $\mathbf{P}(X = y_0) = \mathbf{P}(X \leq y_0) - \mathbf{P}(X < y_0) = 1$. ■

Définition IV.4.2 ◀▶ (*Tribu asymptotique*) Soient \mathcal{G}_n , $n \in \mathbb{N}$, une suite de sous-tribus de \mathcal{F} . On pose

$$\mathcal{T}_n = \sigma(\mathcal{G}_p, p > n) \quad \text{et} \quad \mathcal{T} = \bigcap_{n \geq 0} \mathcal{T}_n .$$

Alors \mathcal{T}_∞ est la *tribu asymptotique* des \mathcal{G}_n , $n \in \mathbb{N}$. La tribu asymptotique d'une suite de variables $(X_n)_{n \in \mathbb{N}}$ est la tribu asymptotique des tribus $\sigma(X_n)$, $n \in \mathbb{N}$. □

Exemple IV.4.1 Soit $(X_n)_{n \in \mathbb{N}}$, une suite de v.a. \mathcal{F} -mesurables. On note \mathcal{T} leur tribu asymptotique, c'est-à-dire $\mathcal{T} = \bigcap_{p \in \mathbb{N}} \sigma(X_n; n \geq p)$. Pour tout $n \in \mathbb{N}$, on pose $S_n = X_0 + \dots + X_n$. Alors

$$\left\{ \sup_{n \in \mathbb{N}} S_n = \infty \right\} = \left\{ \limsup_{n \rightarrow \infty} S_n = \infty \right\} \in \mathcal{T} .$$

Preuve : pour tout $n \in \mathbb{N}$, et tout $\omega \in \Omega$, $S_n(\omega) \neq \infty$. Donc

$$\forall p \in \mathbb{N}, \quad \left\{ \sup_{n \in \mathbb{N}} S_n = \infty \right\} = \left\{ \limsup_{n \rightarrow \infty} S_n = \infty \right\} = \left\{ \sup_{n > p} (S_n - S_p) = \infty \right\} \in \sigma(X_n; n > p) ,$$

ce qui montre bien le résultat voulu. \square

Théorème IV.4.2 ◀▶ (Loi du 0-1 de Kolmogorov) Soient \mathcal{G}_n , $n \in \mathbb{N}$, une suite de sous-tribus de \mathcal{F} qui sont supposés mutuellement indépendantes. Alors, leur tribu asymptotique est \mathbf{P} -triviale.

Preuve : on pose $\mathcal{F}_n := \sigma(\mathcal{G}_0, \dots, \mathcal{G}_n)$, $\mathcal{T}_n := \sigma(\mathcal{G}_p, p > n)$ et $\mathcal{T} := \bigcap_{n \geq 0} \mathcal{T}_n$. Par le corollaire III.1.4 page 99 (indépendance par paquets), la tribu \mathcal{F}_n est indépendant de \mathcal{T}_n . Comme $\mathcal{T} \subset \mathcal{T}_n$, la proposition III.1.1 page 98, on en déduit que \mathcal{F}_n est indépendant de \mathcal{T} . On note $\mathcal{P} := \bigcup_{n \geq 0} \mathcal{F}_n$. Il est clair que \mathcal{P} est indépendant de \mathcal{T} : en effet soit $A \in \mathcal{P}$ et $B \in \mathcal{T}$; il existe $n \in \mathbb{N}$ tel que $A \in \mathcal{F}_n$ et comme \mathcal{F}_n est indépendant de \mathcal{T} , on a $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On remarque ensuite que \mathcal{P} est un pi-système. Le théorème III.1.3 page 98 implique alors que $\sigma(\mathcal{P})$ est indépendant de \mathcal{T} . Mais $\sigma(\mathcal{P}) = \sigma(\mathcal{G}_n, n \in \mathbb{N})$. Donc \mathcal{T} est indépendant de $\sigma(\mathcal{G}_n, n \in \mathbb{N})$. Or clairement $\mathcal{T} \subset \sigma(\mathcal{G}_n, n \in \mathbb{N})$. La proposition III.1.1 page 98 implique donc que \mathcal{T} est indépendant de \mathcal{T} : pour tout $A \in \mathcal{T}$, $\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A)\mathbf{P}(A)$, ce qui implique que $\mathbf{P}(A) = 0$ ou 1. \blacksquare

Exemple IV.4.2 (Application à la marche aléatoire symétrique) Donnons une première application de la loi du 0-1 de Kolmogorov à la marche aléatoire simple symétrique $(S_n)_{n \geq 0}$ introduite à l'exemple III.3.1 page 116. On rappelle que $S_0 = 0$ et que pour tout $n \geq 1$, $S_n = \xi_1 + \dots + \xi_n$, où $(\xi_n)_{n \geq 1}$ est une suite de v.a. indépendantes telles que $\mathbf{P}(\xi_n = 1) = \mathbf{P}(\xi_n = -1) = 1/2$. On montre que la marche aléatoire oscille, c'est-à-dire que

$$\mathbf{P}\text{-p.s.} \quad \limsup_{n \rightarrow \infty} S_n = \infty \quad \text{et} \quad \liminf_{n \rightarrow \infty} S_n = -\infty . \quad (\text{IV.3})$$

Cela montre que p.s. pour tout $k \in \mathbb{Z}$, la suite $(S_n)_{n \geq 0}$ prend une infinité de fois la valeur k . On dit qu'elle est *récurrente*.

Prouvons (IV.3) : on pose $A := \{\limsup_n S_n = \infty\}$ et $B := \{\liminf_n S_n = -\infty\}$. On remarque d'abord que $(-\xi_n)_{n \geq 1}$ est une suite de variables indépendantes de même loi que $(\xi_n)_{n \geq 1}$. Comme $\limsup_n (-S_n) = -\liminf_n S_n$, cela implique que $\mathbf{P}(A) = \mathbf{P}(B)$. Par ailleurs $A \cup B = \{\limsup_n |S_n| = \infty\}$. Or on a montré à l'exemple III.17 page 116, que $\mathbf{P}(\limsup_n |S_n| = \infty) = 1$. Cela implique que $\mathbf{P}(A) = \mathbf{P}(B) > 0$. On note ensuite \mathcal{T} la tribu asymptotique des variables $(\xi_n)_{n \geq 1}$. Comme démontré dans l'exemple IV.4.1, $A, B \in \mathcal{T}$. Comme les v.a. ξ_n sont indépendantes, la loi du 0-1 de Kolmogorov implique que \mathcal{T} est \mathbf{P} -triviale. On doit donc avoir $\mathbf{P}(A) = \mathbf{P}(B) = 1$ et donc $\mathbf{P}(A \cap B) = 1$, ce qui implique immédiatement (IV.3). \square

IV.4.b La loi forte des grands nombres.

Sauf mention du contraire, l'espace de probabilité de référence est noté $(\Omega, \mathcal{F}, \mathbf{P})$. Cette section est consacrée à la preuve du théorème suivant.

Théorème IV.4.3 ◀▶ (Loi des grands nombres) Soit $(X_n)_{n \geq 1}$, une suite de v.a. réelles \mathcal{F} -mesurables. On les suppose indépendantes, de même loi et intégrables. Alors,

$$\mathbf{P}\text{-p.s.} \quad \frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[X_1].$$

Preuve : pour tout $n \geq 1$, on pose $S_n := X_1 + \dots + X_n$ et $S_n^* := X_2 + \dots + X_{n+1}$. On pose également, $S_0 = S_0^* = 0$. On fixe $a \in \mathbb{R}$ et on définit ensuite $M_n := \max_{0 \leq k \leq n} (S_k - ak)$ et $M_n^* := \max_{0 \leq k \leq n} (S_k^* - ak)$. On remarque que $0 \leq M_n \leq M_{n+1}$, et $0 \leq M_n^* \leq M_{n+1}^*$. Enfin, on pose

$$M_\infty := \lim_{n \rightarrow \infty} M_n = \sup_{n \geq 0} M_n \quad \text{et} \quad M_\infty^* := \lim_{n \rightarrow \infty} M_n^* = \sup_{n \geq 0} M_n^*.$$

On remarque ensuite que

$$\begin{aligned} M_{n+1} &= \max(0, X_1 - a, X_1 - a + S_1^* - a, X_1 - a + S_2^* - 2a, \dots, X_1 - a + S_n^* - na) \\ &= \max(0, X_1 - a + M_n^*) = M_n^* + X_1 + \max(-X_1 - M_n^*, -a) \\ &= M_n^* + X_1 - \min(X_1 + M_n^*, a). \end{aligned} \quad (\text{IV.4})$$

Il est clair que le max d'un nombre fini de variables intégrables est une variable intégrable (sa valeur absolue est majorée par la somme des valeurs absolues des variables intégrables). Donc M_{n+1} et M_n^* sont intégrables. De plus, il est clair que M_n et M_n^* ont même loi. Par conséquent, on peut écrire

$$\mathbf{E}[M_{n+1}] - \mathbf{E}[M_n^*] = \mathbf{E}[M_{n+1}] - \mathbf{E}[M_n] = \mathbf{E}[M_{n+1} - M_n] \geq 0,$$

et (IV.4) implique

$$\mathbf{E}[X_1] \geq \mathbf{E}[\min(X_1 + M_n^*, a)]. \quad (\text{IV.5})$$

Comme par définition $M_n^* \geq 0$, on voit que $|\min(X_1 + M_n^*, a)| \leq |X_1| + |a|$. La convergence dominée s'applique et en passant à la limite dans (IV.5), on obtient l'inégalité suivante :

$$\mathbf{E}[X_1] \geq \mathbf{E}[\min(X_1 + M_\infty^*, a)]. \quad (\text{IV.6})$$

L'exemple IV.4.1 montre que l'événement $\{M_\infty = \infty\}$ est dans la tribu asymptotique des variables X_n . Comme ces variables sont indépendantes, cette tribu asymptotique est \mathbf{P} -triviale. Par conséquent on sait que $\mathbf{P}(M_\infty = \infty)$ vaut 0 ou 1. Supposons que $\mathbf{P}(M_\infty = \infty) = 1$. Comme pour tout $n \in \mathbb{N}$, M_n et M_n^* ont même loi, leurs limites p.s. également, c'est-à-dire que M_∞ et M_∞^* ont même loi. Donc $\mathbf{P}(M_\infty^* = \infty) = 1$ et (IV.6) implique que $\mathbf{E}[X_1] \geq a$.

On a donc montré que si $a > \mathbf{E}[X_1]$, alors p.s. $M_\infty < \infty$, c'est-à-dire que presque sûrement pour tout $n \geq 1$, on a $S_n - an \leq M_\infty < \infty$. Donc pour tout $a > \mathbf{E}[X_1]$, p.s. $\limsup_n S_n/n \leq a$. On a donc presque sûrement pour tout $p \in \mathbb{N}$, $\limsup_n S_n/n \leq \mathbf{E}[X_1] + 2^{-p}$, ce qui entraîne

$$\mathbf{P}\text{-p.s.} \quad \limsup_{n \rightarrow \infty} S_n/n \leq \mathbf{E}[X_1].$$

On applique ce résultat à la suite $(-X_n)_{n \geq 1}$, qui satisfait les mêmes hypothèses que $(X_n)_{n \geq 1}$ et on obtient

$$\mathbf{P}\text{-p.s.} \quad -\liminf_{n \rightarrow \infty} S_n/n = \limsup_{n \rightarrow \infty} (-S_n)/n \leq -\mathbf{E}[X_1],$$

ce qui implique que p.s. $\liminf_n S_n/n = \limsup_n S_n/n = \mathbf{E}[X_1]$. ■

Mentionnons que la convergence de la loi des grands nombre a également lieu dans L^1 .

Proposition IV.4.4 *Sous les mêmes hypothèses que le théorème IV.4.3, on a également,*

$$\left\| \frac{1}{n}(X_1 + \dots + X_n) - \mathbf{E}[X_1] \right\|_1 \xrightarrow{n \rightarrow \infty} 0 .$$

Exemple IV.4.3 *Un contreexemple.* On se donne une suite $(X_n)_{n \geq 1}$ de v.a. réelles \mathcal{F} -mesurables indépendantes et de même loi qui est donnée par la densité $x \in \mathbb{R} \mapsto (\pi(1+x^2))^{-1}$. Autrement dit, les X_n suivent une loi de Cauchy. On rappelle le calcul fait à l'exemple (II.10.3) page 88 :

$$\forall u \in \mathbb{R}, \quad \mathbf{E}[e^{iuX_n}] = \frac{1}{\pi} \int_{\mathbb{R}} \frac{e^{iux}}{1+x^2} dx = e^{-|u|} .$$

Il est par ailleurs clair que $\mathbf{E}[|X_n|] = \infty$, c'est-à-dire que les v.a. X_n ne sont pas intégrables. On pose $S_n = X_1 + \dots + X_n$. On vérifie que

$$\mathbf{E}[e^{iuS_n/n}] = \mathbf{E}[e^{iuX_1/n}] \dots \mathbf{E}[e^{iuX_n/n}] = e^{-|u|} .$$

Donc, pour tout n , S_n/n a même loi que X_1 , c'est-à-dire suit une loi de Cauchy. Par conséquent S_n/n ne converge pas en probabilité : en effet, si c'était le cas, on note S , la v.a. limite ; on pourrait extraire une suite convergeant p.s. vers S qui serait mesurable par rapport à la tribu asymptotique des $(X_n)_{n \geq 1}$, qui est triviale par la loi du 0-1 de Kolmogorov. Donc la suite $(S_n/n)_{n \geq 1}$ devrait converger vers une constante s , or pour tout $\varepsilon > 0$, la suite $\mathbf{P}(|S_n/n - s| > \varepsilon)$ est constante et ne tend pas vers 0 lorsque $n \rightarrow \infty$, ce qui est contradictoire. Donc $(S_n/n)_{n \geq 1}$ ne converge pas en probabilité, elle ne converge donc ni p.s., ni en norme L^p : elle ne satisfait aucune loi des grands nombres. \square

Exemple IV.4.4 *Application au renouvellement.* Soit $(X_n)_{n \geq 1}$, une suite de v.a. \mathcal{F} -mesurables indépendantes et de même loi. On suppose que

$$\mathbf{P}(X_n > 0) = 1 \quad \text{et} \quad m = \mathbf{E}[X_n] < \infty .$$

On pose alors

$$T_0 = 0 \quad \text{et} \quad \forall n \geq 1, \quad T_n = X_1 + \dots + X_n .$$

On s'intéresse au comportement asymptotique des variables suivantes :

$$\forall t \in \mathbb{R}_+, \quad N_t = \sum_{n \geq 1} \mathbf{1}_{[0,t]}(T_n) .$$

Comme conséquence de la loi des grands nombres, nous montrons le théorème suivant.

Théorème IV.4.5 (Un théorème de renouvellement) *Avec les hypothèses et les notations ci-dessus, on a*

$$\mathbf{P}\text{-p.s.} \quad \lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{m} .$$

Preuve : on observe que $t \mapsto N_t$ est croissant et on pose $N_\infty := \sup_{t \geq 0} N_t = \lim_{\infty} \uparrow N_t$ qui est une v.a. à valeurs dans $[0, \infty]$ \mathcal{F} -mesurable. On observe que pour tout $k \in \mathbb{N}$ et tout $t \in \mathbb{R}_+$, on a $\mathbf{P}(N_t \geq k) = \mathbf{P}(T_k \leq t)$. Donc $\mathbf{P}(N_\infty \geq k) = 1$, pour tout $k \in \mathbb{N}$, ce qui implique que p.s. $N_\infty = \infty$. On a donc p.s. $\lim_{t \rightarrow \infty} N_t = \infty$.

La loi des grands nombres permet ensuite d'affirmer que p.s. $\lim_n T_n/n = m$. Or $T_{N_t} \leq t \leq T_{N_t+1}$ donc

$$\frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{N_t + 1}{N_t} \frac{T_{N_t+1}}{N_t + 1},$$

ce qui permet de conclure puisque p.s. $\lim_{t \rightarrow \infty} N_t = \infty$. ■

Comme pour le processus de Poisson, on peut imaginer que les T_n représentent les instants où les ampoules d'une lampe médicale grillent : l'ampoule de la lampe fonctionne pendant un temps de même loi que celle de X_1 , puis elle grille et est immédiatement remplacée par une autre ampoule de type similaire, etc. Alors, X_n représente la durée de fonctionnement de la $n^{\text{ième}}$ ampoule en fonctionnement et N_t est égal au nombre d'ampoules grillées au temps t . Le théorème précédent dit que N_t vaut approximativement t/m , ce qui n'est pas très surprenant. □

Exemple IV.4.5 *Aiguille de Buffon, méthode de Monte-Carlo.* Georges-Louis Leclerc de Buffon, est le grand naturaliste du XVIII^{ième} siècle, connu pour avoir donné une méthode raisonnée de description des animaux, pour avoir le premier énoncé une définition cohérente de l'espèce animale et pour avoir remis en question l'âge biblique de notre planète. Mais le premier travail l'ayant fait connaître (et immédiatement entrer à l'Académie des Sciences) est un mémoire mathématique écrit en 1733 sur une expérience aléatoire permettant d'estimer le nombre π . Cette expérience est connue sous le nom de l'expérience de *l'aiguille de Buffon* et se décrit comme suit.

Dans une grande pièce, dont le sol est recouvert d'un parquet aux lattes de même espacement $2a$, on lance n fois au hasard une aiguille de longueur a . On note S_n le nombre de fois où l'aiguille est tombée sur deux lattes à la fois. Alors n/S_n tend vers π .

Preuve : on suppose que l'observateur se place de façon à ce que les lattes soient orientées verticalement. Lorsqu'on lance l'aiguille pour la $k^{\text{ième}}$ fois, on note D_k la distance du centre de l'aiguille au bord gauche de la latte sur lequel il est tombé et θ_k l'angle que fait la droite portée par l'aiguille par rapport aux lattes. Il est raisonnable de penser que D_k est uniforme sur $[0, 2a]$, que θ_k est uniforme sur $[-\pi/2, \pi/2]$ et que D_k et θ_k sont indépendantes. On pose $X_k = 1$ si l'aiguille numero k tombe sur deux lattes et $X_k = 0$, sinon. Les v.a. $(X_k)_{k \geq 1}$ sont indépendantes et de même loi, et on a $S_n = X_1 + \dots + X_n$. Par ailleurs,

$$\mathbf{E}[X_k] = \mathbf{P}(X_k = 1) = 2\mathbf{P}(a|\sin \theta_k| \geq 2D_k) = \frac{2}{2a} \frac{a}{2} \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} |\sin \theta| d\theta = \frac{1}{\pi}$$

et la loi des grands nombres permet de conclure. ■

La méthode de Buffon vaut plus pour l'originalité du procédé que pour son efficacité car la convergence de n/S_n vers π est lente : $|n/S_n - \pi|$ est grossièrement de l'ordre de $n^{-1/2}$. Pour avoir 99% de chance de connaître les 4 premières décimales de π , il est nécessaire de lancer environ cent millions de fois l'aiguille ... Or les mathématiciens de l'époque de Buffon connaissaient les cent premières décimales de π par la formule de J. Machin a permis en 1706 d'en calculer les 100 premières décimales.

La méthode de Buffon est en réalité très robuste : imaginons que le sol d'une pièce de 10 m² soit recouvert par un linoléum rouge et bleu aux motifs très compliqués. Supposons que l'on veuille estimer la surface totale des motifs rouges. Comme les motifs sont extrêmement compliqués, une méthode consistant à trianguler les surfaces rouges est quasiment impossible, les erreurs de mesures risquant vraiment d'être trop grandes. La méthode de Buffon, même si elle converge plutôt lentement donne une réponse : si on lance aléatoirement une balle en caoutchouc n fois, que l'on note S_n le

nombre de fois où la balle termine sa course sur du rouge (ce qui est facile à vérifier), alors $10 \times S_n/n$ tend vers la surface totale des motifs rouges exprimée en m^2 .

La méthode de Buffon est tombée dans l'oubli durant longtemps, elle n'est mentionnée qu'à titre de curiosité mathématique au XIX^{ème} siècle. Elle a été redécouverte pour les besoins de calculs approchés d'intégrales de fonctions à plusieurs variables. Les physiciens nucléaires et les mathématiciens ayant développé cette méthode aléatoire l'ont baptisée, par dérision, *méthode de Monte-Carlo* car la source d'aléa utilisée à l'époque étaient des roulettes.

La méthode de Monte-Carlo exploite un autre aspect de l'expérience de Buffon, qu'il ne pouvait concevoir à l'époque : soit $f : [0, 1]^k \rightarrow \mathbb{R}$, une fonction mesurable bornée (continue bornée par exemple). On cherche à calculer

$$I = \int_{[0,1]^k} f(x_1, \dots, x_k) dx_1 \dots dx_k .$$

On suppose qu'il est relativement facile, étant donné $x \in [0, 1]^k$, de calculer $f(x)$. La méthode de Monte-Carlo consiste à générer une suite $(U_n)_{n \geq 1}$ de v.a. indépendantes de loi uniforme sur $[0, 1]^k$, et d'appliquer la loi des grands nombres qui affirme que

$$\text{p.s.} \quad \frac{f(U_1) + \dots + f(U_n)}{n} \xrightarrow[n \rightarrow \infty]{} I .$$

Les v.a. uniformes dans la pratique sont des nombres pseudo-aléatoires presque uniformes et presque indépendants fournis par un logiciel. Le fait que les v.a. U_n ne soient pas vraiment indépendantes n'est pas très gênant (voir le théorème ergodique plus loin dans le cours). Ce qui est plus ennuyeux est la lenteur de la convergence : l'erreur que l'on commet en assimilant I à $(f(U_1) + \dots + f(U_n))/n$ est de l'ordre de $Cn^{-1/2}$, où C dépend de f d'une manière que l'on peut estimer. Lorsque la dimension $k \leq 10$, et lorsque la fonction f est régulière (C^2 par exemple), les méthodes analytiques sont plus efficaces. En revanche la vitesse de convergence de la méthode de Monte-Carlo reste à peu près la même quelle que soit la dimension k . Elle est donc meilleure lorsque la dimension k est très grande : c'est précisément pour cette raison qu'elle a été et est toujours employée. \square

Quelques commentaires sur la lois des grands nombres. Commençons par un bref historique. Jakob Bernoulli semble être le premier à avoir prouvé formellement en 1713 la loi faible des grands nombres pour des variables à valeurs dans $\{0, 1\}$, qui ont été ultérieurement appelées variables de Bernoulli. Bienaymé en 1853, est le premier à avoir énoncé l'inégalité de Tchebychev, qui porte parfois le nom de Bienaymé-Tchebychev ; il est également le premier à avoir remarqué que la variance de la somme de n v.a. indépendantes est la somme de leur variance. Tchebychev en 1867 a corrigé les imperfections mathématiques des travaux de Bienaymé et il est le premier à avoir formulé et prouvé rigoureusement la loi faible des grands nombres pour des variables possédant un moment d'ordre 2. C'est Borel qui, dans des travaux de 1903 et de 1909, a démontré une loi forte des grands nombres pour des variables admettant un moment d'ordre 4. Il l'a fait à l'aide du lemme portant son nom. Cantelli en 1917 a amélioré les arguments de Borel. Kolmogorov en 1930 est le premier à avoir montré la loi des grands nombres en général, c'est-à-dire le théorème IV.4.4, et à avoir montré que cet énoncé ne pouvait être amélioré. Sa preuve est différente de celle qui est exposée dans le cours.

La loi des grands nombres est en quelque sorte le théorème qui permet de donner un contenu réel à la théorie des probabilités telle qu'elle est exposée dans ce cours. En effet, que signifie au fond la phrase suivante : "en lachant un verre de la marque Glouglou à un mètre du sol, il a 21,07% de

chance de se briser" ? Cela signifie que si on jette n verres du même type depuis un mètre du sol, la proportion des verres cassés tend vers 21,07%. On peut dire que c'est la loi des grands nombres qui permet d'ancrer dans la réalité (et de donner un certain pouvoir prédictif) à la notion de probabilité d'un événement. De même que signifie la phrase "en appliquant ma stratégie secrète, mon gain moyen au whist est de 17,87 euros par partie" ? Cela signifie que si je joue n fois au whist en appliquant ma stratégie secrète, la moyenne arithmétique de mes gains tend vers 17,87 euros lorsque n tend vers l'infini. La loi des grands nombres donne donc aussi un contenu réel à la notion d'espérance (ou de moyenne) d'une variable aléatoire.

La notion d'espérance mathématique ne doit pas s'interpréter de manière erronée comme le montre l'expérience imaginaire suivante. Supposons que l'on propose à monsieur A de jouer à un jeu d'argent où il a une probabilité de $1 - 10^{-9}$ de gagner 10^8 euros et une probabilité 10^{-9} de perdre 10^{20} euros. Monsieur A se demande s'il doit jouer : rassemblant ses souvenirs de Licence, il pose $X :=$ son gain algébrique à ce jeu et calcule

$$\mathbf{E}[X] = 10^8(1 - 10^{-9}) - 10^{20}10^{-9} .$$

Il constate que $\mathbf{E}[X]$ un nombre spectaculairement négatif et décide de ne pas jouer. Mais tout ce que monsieur A a prouvé en décidant de ne pas jouer, est que les études supérieures permettent d'être stupide de manière sophistiquée. Si monsieur A ne joue qu'une seule fois à ce jeu, il est quasiment sûr de gagner 100 millions d'euros. Il doit donc jouer ! (sauf s'il ne veut pas de tant d'argent, mais c'est un autre problème). Une bonne question est : *avec une probabilité de 9999/10000 combien de fois monsieur A peut jouer en gagnant à chaque fois ?* Cela revient à chercher n tel que $(1 - 10^{-9})^n \approx 1 - 10^{-4}$, c'est-à-dire

$$n = \frac{\log(1 - 10^{-4})}{\log(1 - 10^{-9})} = \frac{10^{-4} + 10^{-8}/2 + \dots}{10^{-9} + 10^{-18}/2 + \dots} = 10^5(1 + 10^{-4}/2 + \dots) .$$

Si monsieur A avait mieux compris son cours de probabilité de licence, il aurait pu jouer environ 100 000 fois à ce jeu en toute sécurité et empocher environ 10000000000000 euros.

Exemple IV.4.6 Donnons un exemple assez étrange où notre intuition du hasard, basée sur la loi des grands nombres, est un peu malmenée. Monsieur A possède un dé \mathcal{D} et monsieur B possède un dé \mathcal{D}' . Les deux dés sont cubiques. Sur chaque face de ces dés figure un nombre entier entre 1 et 6, mais il est possible que sur un dé, un même chiffre apparaisse plusieurs fois : par exemple, les chiffres figurant sur \mathcal{D} peuvent être tous égaux à 3 et ceux figurant sur \mathcal{D}' peuvent être (2, 2, 2, 2, 6, 6). Les joueurs décident de lancer leur dé n fois ; A et B déposent chacun $n/2$ euros pour créer une banque dont l'avoir total est de n euros ; un lancer se déroule comme suit : les joueurs lancent leur dé ; le gagnant est celui dont le dé montre sur sa face supérieure le chiffre le plus grand ; le gagnant reçoit 1 euro de la banque, le perdant ne reçoit rien ; en cas d'égalité les joueurs reçoivent chacun 0,5 euro. Le dé \mathcal{D} est dit *meilleur* que le dé \mathcal{D}' si le joueur A gagne en moyenne plus souvent avec le dé \mathcal{D} contre le joueur B avec le dé \mathcal{D}' .

Monsieur A calcule la valeur moyenne des faces de son dé \mathcal{D} : il obtient $18/6 = 3$ (ce n'est pas très dur à calculer). Monsieur B décide de faire de même, et il constate que la valeur moyenne des faces de son dé \mathcal{D}' est $20/6 = 10/3$. Se souvenant vaguement de son cours de probabilité, monsieur B exulte : il a hâte de commencer la partie. Mais il a tort : si on note X_n le gain du joueur A après n parties, et si on note Y_n le gain du joueur B, la loi des grands nombres montre que

$$\text{p.s.} \quad \lim_n X_n/n = \mathbf{P}(\mathcal{D} > \mathcal{D}') = \mathbf{P}(\mathcal{D}' = 2) = \frac{4}{6} = \frac{2}{3} .$$

On vérifie également que

$$\text{p.s.} \quad \lim_n Y_n/n = \mathbf{P}(\mathcal{D}' > \mathcal{D}) = \mathbf{P}(\mathcal{D}' = 6) = \frac{2}{6} = \frac{1}{3}.$$

On peut donc affirmer sur cet exemple que \mathcal{D} est meilleur que \mathcal{D}' . La valeur moyenne d'un dé n'est tout simplement pas la grandeur statistique pertinente pour prédire qui va avoir tendance à gagner à ce jeu.

Supposons maintenant que monsieur A apporte les 3 dés suivants :

$$\mathcal{D}_1 = (3, 3, 3, 3, 3, 6), \quad \mathcal{D}_2 = (2, 2, 2, 5, 5, 5), \quad \mathcal{D}_3 = (1, 4, 4, 4, 4, 4).$$

Monsieur A propose à monsieur B de choisir l'un de ces trois dés, ce qu'il fait. Puis monsieur A, choisit l'un des deux dés restants et sans plus changer de dés, ils effectuent n parties.

Monsieur B, peu méfiant, choisit le dé \mathcal{D}_1 ; alors monsieur A choisit le dé \mathcal{D}_3 et au bout de n parties, monsieur A gagne : c'était en effet probable car $\mathbf{P}(\mathcal{D}_3 > \mathcal{D}_1) = 25/36$, c'est-à-dire que \mathcal{D}_3 est meilleur que \mathcal{D}_1 . Monsieur A propose ensuite à monsieur B de refaire une partie : monsieur B choisit le dé \mathcal{D}_3 ; monsieur A choisit alors le dé \mathcal{D}_2 et il gagne au bout de n parties. Cela était probable car $\mathbf{P}(\mathcal{D}_2 > \mathcal{D}_3) = 21/36$. Monsieur A, qui veut édifier monsieur B, lui propose encore de rejouer : monsieur B qui croit enfin gagner choisit le dé \mathcal{D}_2 et monsieur A choisit alors le dé \mathcal{D}_1 et ... il gagne encore au bout de n parties, ce qui était probable puisque $\mathbf{P}(\mathcal{D}_1 > \mathcal{D}_2) = 21/36$. En fait,

$$\mathcal{D}_1 \text{ est meilleur que } \mathcal{D}_2, \mathcal{D}_2 \text{ est meilleur que } \mathcal{D}_3 \text{ et } \mathcal{D}_3 \text{ est meilleur que } \mathcal{D}_1,$$

ce qui semble assez paradoxal. Les dés ci-dessus sont appelés *Dés non-transitifs*. Pour lever ce paradoxe, il faut lancer les trois dés en même temps : nous laissons les détails aux lecteurs, qui, s'ils le souhaitent peuvent entamer une carrière d'escroc. \square

IV.5 Exercices.

Exercice IV.5.1 Soit $(\theta_n)_{n \geq 1}$ une suite de réels strictement positifs telle que $\lim_{n \rightarrow +\infty} \theta_n = +\infty$. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes telle que pour tout $n \geq 1$, X_n suive la loi exponentielle de paramètre θ_n .

a. Étudier la convergence en probabilités de la suite $(X_n)_{n \geq 1}$. Quelle hypothèse n'a-t-on pas utilisée ?

b. Reprendre la question précédente avec la convergence dans L^1 .

c. Étudier, dans le cas où $\theta_n = n$ puis dans le cas où $\theta_n = \log n$, la convergence presque sûre de la suite $(X_n)_{n \geq 1}$.

Exercice IV.5.2 Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités. Soit $(A_n)_{n \geq 1}$ une suite d'événements sur cet espace de probabilités. Soit $p \geq 1$ un réel. Déterminer pour chacune des convergences suivantes à quelle condition sur la suite $(A_n)_{n \geq 1}$ elle a lieu.

a. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge en probabilité vers 0.

b. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge en probabilité vers $\mathbf{1}_A$.

b. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge dans L^p vers 0.

d. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge dans L^p vers $\mathbf{1}_A$.

c. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge presque sûrement vers 0.

f. La suite $(\mathbf{1}_{A_n})_{n \geq 1}$ converge presque sûrement vers $\mathbf{1}_A$.

Exercice IV.5.3 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. Montrer que si la suite X_n converge simultanément vers deux variables aléatoires X et Y , alors $X = Y$ presque sûrement, et ceci quel que soit le mode de convergence vers X et quel que soit le mode de convergence vers Y , parmi : convergence presque sûre, convergence dans L^p avec $p \geq 1$, convergence en probabilité.

Exercice IV.5.4 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires qui toutes suivent la loi exponentielle de paramètre 1.

b. Montrer que $\mathbf{P} \left(\limsup \frac{X_n}{\log n} > 1 \right) = 0$.

On suppose désormais X_1, X_2, \dots indépendantes.

c. Montrer que $\mathbf{P} \left(\limsup \frac{X_n}{\log n} < 1 \right) = 0$. Montrer que ce résultat peut être faux sans l'hypothèse d'indépendance.

d. Montrer que $\limsup \frac{X_n}{\log n}$ est presque sûrement égale à une constante que l'on déterminera.

e. Déterminer une suite $(a_n)_{n \geq 1}$ de réels qui converge vers 0 et telle qu'on ait $X_n < a_n$ infiniment souvent avec probabilité 1, c'est-à-dire $\mathbf{P}(X_n < a_n \text{ infiniment souvent}) = 1$. En déduire ce que vaut $\liminf X_n$.

f. Déterminer une suite $(b_n)_{n \geq 1}$ de réels telle qu'on ait $\liminf (b_n X_n) = 1$ presque sûrement.

Exercice IV.5.5 a. Construire une suite de réels $(x_n)_{n \geq 1}$ telle que pour tout $n \geq 1$, on ait $x_n = 0$ ou $x_n = 1$, et telle que la suite $(m_n)_{n \geq 1}$ définie par

$$m_n = \frac{x_1 + \dots + x_n}{n}$$

ne soit pas convergente.

b. Construire une suite de variables aléatoires $(X_n)_{n \geq 1}$ telle que pour tout $n \geq 1$, la variable X_n suive la loi de Bernoulli de paramètre $\frac{1}{2}$ et telle qu'on ait

$$\mathbf{P} \left(\text{la suite} \left(\frac{X_1 + \dots + X_n}{n} \right)_{n \geq 1} \text{ est convergente} \right) = 0.$$

Exercice IV.5.6 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. Soit $p \geq 1$. Soit X une variable aléatoire telle que $\mathbf{E}[|X|^p] < +\infty$. On suppose que la série suivante converge :

$$\sum_{n \geq 1} \mathbf{E}[|X_n - X|^p] < +\infty.$$

On parle parfois de *convergence rapide* dans L^p .

Montrer que la suite $(X_n)_{n \geq 1}$ converge dans L^p vers X . Montrer que la suite $(X_n)_{n \geq 1}$ converge presque sûrement vers X .

Exercice IV.5.7 a. En écrivant $e^{-\frac{t^2}{2}} = \frac{te^{-\frac{t^2}{2}}}{t}$, montrer que pour tout $x > 0$, on a

$$\int_x^{+\infty} e^{-\frac{t^2}{2}} dt \leq \frac{1}{x} e^{-\frac{x^2}{2}}.$$

b. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$. Soit x un réel strictement positif. On pose

$$T = \min\{n \geq 1 : X_n \geq x\}.$$

Déterminer la loi de T et donner une borne inférieure explicite pour l'espérance de T . Quel est l'ordre de grandeur de cette borne lorsque $x = 5$? Lorsque $x = 10$? On pourra utiliser l'égalité approchée $\log 10 \simeq 2,3$.

c. Reprendre la question précédente lorsque les variables X_n suivent la loi de Cauchy standard.

Exercice IV.5.8 (Cet exercice est une reformulation de l'exercice 11 de la feuille précédente.)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes telle que pour tout $n \geq 1$, X_n suive la loi exponentielle de paramètre $\log n$.

a. Montrer qu'avec probabilité 1, la suite $(X_n)_{n \geq 1}$ prend une infinité de fois des valeurs supérieures à 1.

b. Montrer que pour tout $\varepsilon > 0$, avec probabilité 1, la suite $(X_n)_{n \geq 1}$ ne prend qu'un nombre fini de fois des valeurs supérieures à $1 + \varepsilon$.

c. Montrer que pour tout $\varepsilon > 0$, avec probabilité 1, la suite $(X_n)_{n \geq 1}$ prend une infinité de fois des valeurs inférieures à ε .

Exercice IV.5.9 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$. On fait les hypothèses suivantes :

(H_1) la suite $(X_n)_{n \geq 1}$ converge en probabilité vers une variable aléatoire X ,

(H_2) il existe un réel $M > 0$ tel que pour tout $n \geq 1$ on ait $\mathbf{E}[X_n^2] \leq M$.

a. Montrer, en utilisant le lemme de Fatou, que $\mathbf{E}[X^2] \leq M$.

b. Montrer que la suite $(X_n)_{n \geq 1}$ converge dans L^1 vers X .

c. Montrer par un exemple que les deux hypothèses peuvent être vérifiées sans qu'il existe une variable aléatoire positive et intégrable Y telle que pour tout $n \geq 1$, $|X_n| \leq Y$.

Autrement dit, le résultat ci-dessus assure parfois la convergence dans L^1 dans des conditions où le théorème de convergence dominée est inopérant.

Chapitre V

Convergence en loi. Théorème central limite.

V.1 Définition, premiers résultats.

On munit \mathbb{R}^d d'une norme $\|\cdot\|$ et de la tribu des boréliens $\mathcal{B}(\mathbb{R}^d)$ correspondante. On rappelle les notations suivantes.

- $C_b(\mathbb{R}^d)$ est l'espace des fonctions continues bornées sur \mathbb{R}^d .
- $\mathcal{M}_1(\mathbb{R}^d)$ est l'espace des mesures $\mu : \mathcal{B}(\mathbb{R}^d) \rightarrow \mathbb{R}_+$ telles que $\mu(\mathbb{R}^d) = 1$, c'est-à-dire l'espace des mesures de probabilités sur \mathbb{R}^d .

On rappelle la définition de la convergence étroite.

Définition V.1.1 ◀▶ (Convergence étroite, convergence en loi)

- (a) Soient $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R}^d)$, $n \in \mathbb{N}$. La suite de mesures $(\mu_n)_{n \in \mathbb{N}}$ converge en loi (on dit aussi *étroitement*) vers μ si

$$\forall f \in C_b(\mathbb{R}^d), \quad \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \int_{\mathbb{R}^d} f d\mu .$$

- (b) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient X et $X_n : \Omega \rightarrow \mathbb{R}^d$, $n \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. La suite de v.a. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si les lois $(\mu_{X_n})_{n \in \mathbb{N}}$ convergent en loi vers μ_X . Autrement dit, par le théorème de transfert, la suite de v.a. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si

$$\forall f \in C_b(\mathbb{R}^d), \quad \lim_{n \rightarrow \infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)] .$$

On remarque que la convergence en loi ne concerne pas directement les v.a. X_n mais leur lois. Mais par abus de langage, on écrit souvent dans la suite $\lim_n X_n = X$ en loi, ce qui est aussi parfois noté $X_n \Longrightarrow X$. □

On rappelle le lemme II.10.10 (page 90) : soient $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$.

$$\mu = \nu \quad \iff \quad \forall f \in C_c(\mathbb{R}^d), \quad \int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu . \quad (\text{V.1})$$

Il s'en suit immédiatement l'unicité de la limite étroite (ou en loi).

Proposition V.1.1 ◀▶ Soient $\mu_n, \mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$, $n \in \mathbb{N}$. On suppose que $\lim_n \mu_n = \mu$ en loi et $\lim_n \mu_n = \nu$ en loi. Alors $\mu = \nu$.

Mentionnons que la convergence étroite sur $\mathcal{M}_1(\mathbb{R}^d)$ est métrisable (ce n'est pas un résultat évident). La convergence étroite des suites suffit donc à déterminer une topologie. On montre ensuite le lemme suivant.

Lemme V.1.2 ◀▶ Soient $\mu, \mu_p \in \mathbb{R}^n$, $p \in \mathbb{N}$, des mesures finies. Alors, $\lim_p \mu_p = \mu$ étroitement ssi

$$\mu_p \xrightarrow[p \rightarrow \infty]{\text{loi}} \mu \iff \forall \psi \in C_c(\mathbb{R}^n), \int_{\mathbb{R}^n} \psi d\mu_p \xrightarrow[p \rightarrow \infty]{} \int_{\mathbb{R}^n} \psi d\mu. \quad (\text{V.2})$$

Preuve : l'implication \Rightarrow est évidente car $C_c(\mathbb{R}^n) \subset C_n(\mathbb{R}^n)$. Montrons la réciproque en supposant que pour toute fonction $\psi \in C_c(\mathbb{R}^n)$, on ait $\lim_p \int_{\mathbb{R}^n} \psi d\mu_p = \int_{\mathbb{R}^n} \psi d\mu$. On fixe $\varphi \in C_b(\mathbb{R}^n)$. Pour tout réel $r > 0$ on pose :

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad \phi_r(\mathbf{x}) = (1 - (\|\mathbf{x}\| - r)_+)_+.$$

Pour tous $r \geq r'$, on voit que $\mathbf{1}_{\overline{B}(0, r')} \leq \phi_{r'} \leq \phi_r \leq \mathbf{1}_{\overline{B}(0, r+1)}$ et $\lim_{r \rightarrow \infty} \phi_r = \mathbf{1}_{\mathbb{R}^n}$ ponctuellement. De plus, ϕ_r et $\varphi \phi_r$ sont dans $C_c(\mathbb{R}^n)$. On remarque ensuite que $|\varphi - \varphi \phi_r| \leq \|\varphi\|_\infty (1 - \phi_r)$. Pour tout $p \in \mathbb{N}$ et tout $r > 0$, on a donc

$$\left| \int_{\mathbb{R}^n} \varphi d\mu - \int_{\mathbb{R}^n} \varphi d\mu_p \right| \leq \left| \int_{\mathbb{R}^n} \varphi \phi_r d\mu - \int_{\mathbb{R}^n} \varphi \phi_r d\mu_p \right| + \|\varphi\|_\infty \left(2 - \int_{\mathbb{R}^n} \phi_r d\mu - \int_{\mathbb{R}^n} \phi_r d\mu_p \right).$$

Par (V.2), en passant à la limsup en p , on obtient pour tout $r > 0$,

$$\limsup_{p \rightarrow \infty} \left| \int_{\mathbb{R}^n} \varphi d\mu - \int_{\mathbb{R}^n} \varphi d\mu_p \right| \leq 2\|\varphi\|_\infty \left(1 - \int_{\mathbb{R}^n} \phi_r d\mu \right).$$

Or par convergence monotone $\lim_{r \rightarrow \infty} \int \phi_r d\mu = 1$, ce qui implique que $\limsup_p \left| \int \varphi d\mu - \int \varphi d\mu_p \right| = 0$ et donc que $\lim_p \int \varphi d\mu_p = \int \varphi d\mu$, pour toute fonction $\varphi \in C_b(\mathbb{R}^n)$. ■

Le théorème suivant est dû à D. Alexandroff. Il est connu sous le nom de *Portemanteau's Theorem*, qui est son appellation standard mais il s'agit d'un canular (ayant fonctionné à la perfection) de P. Billingsley qui dans son célèbre ouvrage *Convergence in Probability Measures*, l'attribue à un certain Jean-Pierre Portemanteau qui aurait publié ce résultat dans la revue *Annals of the University of Felletin* en 1915 dans l'article intitulé *Espoir pour l'ensemble vide*. La seule chose existant bel et bien dans le canular est la commune de Felletin (Creuse) qui compte 1867 habitants.

Théorème V.1.3 ◀▶ (**Jean-Pierre Portemanteau's Theorem d'Alexandroff**) Soient $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R}^d)$, $n \in \mathbb{N}$. Les assertions suivantes sont équivalentes.

- (i) $\lim_n \mu_n = \mu$ étroitement.
- (ii) $\forall F \subset \mathbb{R}^d$, fermé, $\limsup_n \mu_n(F) \leq \mu(F)$.
- (iii) $\forall O \subset \mathbb{R}^d$, ouvert, $\liminf_n \mu_n(O) \geq \mu(O)$.
- (iv) Pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ tel que $\mu(\overline{A}) = \mu(\overset{\circ}{A})$, $\lim_n \mu_n(A) = \mu(A)$.
- (v) Pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$, $\mathcal{B}(\mathbb{R}^d)$ -mesurable bornée telle que l'ensemble de ses points de discontinuité soit μ -négligeable, on a $\lim_n \int_E f d\mu_n = \int_E f d\mu$.

Preuve : on observe que $(v) \implies (i)$, trivialement. Par passage au complémentaire, $(ii) \iff (iii)$. Supposons que (ii) et (iii) soient vrais. On a donc

$$\mu(\mathring{A}) \leq \liminf_{n \rightarrow \infty} \mu_n(\mathring{A}) \leq \liminf_{n \rightarrow \infty} \mu_n(A) \leq \limsup_{n \rightarrow \infty} \mu_n(A) \leq \limsup_{n \rightarrow \infty} \mu_n(\bar{A}) \leq \mu(\bar{A}).$$

Si $\mu(\bar{A}) = \mu(\mathring{A})$, cela implique (iv) . Il reste donc à démontrer $(i) \implies (ii)$ et $(iv) \implies (v)$.

On suppose (i) . Pour tout fermé $F \subset \mathbb{R}^d$ et tout $p \in \mathbb{N}$, on pose $\phi_p(x) = (1 - pd(x, F))_+$, où $d(x, F) = \inf_{y \in F} \|x - y\|$ est la distance de x au fermé F . On rappelle que $d(\cdot, F)$ est 1-Lipschitzienne donc continue. On en déduit que $\phi_p \in C_b(\mathbb{R}^d)$. Par ailleurs, on observe que $\phi_p \geq \phi_{p+1} \geq \mathbf{1}_F$ et que $\lim_p \phi_p = \mathbf{1}_F$ ponctuellement. Donc, pour tout $p \in \mathbb{N}$

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}^d} \phi_p d\mu_n = \int_{\mathbb{R}^d} \phi_p d\mu.$$

Or par convergence dominée, $\lim_p \int_{\mathbb{R}^d} \phi_p d\mu = \mu(F)$, ce qui montre (ii) .

On suppose (iv) . Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, une fonction $\mathcal{B}(\mathbb{R}^d)$ -mesurable telle que $c := \|f\|_\infty < \infty$. On note D l'ensemble de ses points de discontinuité de f et $C = \mathbb{R}^d \setminus D$, l'ensemble des points de continuité de f . On a montré au lemme IV.1.6 page 129 que $C \in \mathcal{B}(\mathbb{R}^d)$ et donc $D \in \mathcal{B}(\mathbb{R}^d)$ également. On suppose que $\mu(D) = 0$ et donc que $\mu(C) = 1$.

On rappelle que ℓ désigne la mesure de Lebesgue sur \mathbb{R} . Pour tout $y \in [0, c]$, on pose $A_y = \{f \geq y\}$. Par Fubini positif, pour toute mesure $\nu \in \mathcal{M}_1(\mathbb{R}^d)$, on a

$$\int_{\mathbb{R}^d} f d\nu = \int_{\mathbb{R}^d} \nu(dx) \int_{[0, c]} \ell(dy) \mathbf{1}_{\{0 \leq y \leq f(x)\}} = \int_{[0, c]} \ell(dy) \int_{\mathbb{R}^d} \nu(dx) \mathbf{1}_{\{0 \leq y \leq f(x)\}} = \int_{[0, c]} \ell(dy) \nu(A_y). \quad (\text{V.3})$$

On pose également pour tout $y \in [0, c]$, $B_y = \{f > y\}$ et $s(y) = \mu(A_y)$. Il est facile de voir que s décroît. Pour toute suite $y_n \in]y, c]$, $n \in \mathbb{N}$, décroissant vers y , on a $\bigcup_{n \in \mathbb{N}} A_{y_n} = B_y$. Donc $s(y+) = \mu(B_y)$. Cela montre que $\{y \in [0, c] : \mu(A_y) > \mu(B_y)\} \subset \{y \in [0, c] : s(y) > s(y+)\}$. Comme s décroît, ses discontinuités sont dénombrables, donc Lebesgue-négligeables. Cela montre

$$\ell(\{y \in [0, c] : \mu(A_y) > \mu(B_y)\}) = 0. \quad (\text{V.4})$$

On vérifie ensuite que

$$B_y \cap C \subset \mathring{A}_y \subset A_y \subset \bar{A}_y \subset A_y \cup D. \quad (\text{V.5})$$

En effet, si $x \in B_y \cap C$, alors f est continue en x et $f(x) > y$, et par continuité, il existe $\varepsilon > 0$, tel que pour tout $z \in B(x, \varepsilon)$, $f(z) > y$ et donc $B(x, \varepsilon) \subset B_y \subset A_y$, ce qui montre la première inclusion de (V.5). Supposons ensuite que $x \in \bar{A}_y \setminus A_y$. Il existe donc $x_n \in A_y$, $n \in \mathbb{N}$, tendant vers x qui est tel que $f(x) < y$, car $x \notin A_y$. Comme $x_n \in A_y$, on a aussi $f(x_n) \geq y$, pour tout $n \in \mathbb{N}$. On voit que f ne peut être continue en x , et donc que $x \in D$, ce qui montre la dernière inclusions de (V.5).

Puisque $\mu(D) = 0$ et $\mu(C) = 1$, on a $\mu(B_y) = \mu(B_y \cap C)$ et $\mu(A_y \cup D) = \mu(A_y)$. Par (V.5), on a

$$\{y \in [0, c] : \mu(\bar{A}_y) > \mu(\mathring{A}_y)\} \subset \{y \in [0, c] : \mu(A_y) > \mu(B_y)\}$$

Par (V.4), $\{y \in [0, c] : \mu(\bar{A}_y) > \mu(\mathring{A}_y)\}$ est ℓ -négligeable et (iv) implique que pour ℓ -presque tout y , $\lim_n \mu_n(A_y) = \mu(A_y)$. Par convergence dominée et (V.3), on a donc

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d\mu_n = \lim_{n \rightarrow \infty} \int_{[0, c]} \ell(dy) \mu_n(A_y) = \int_{[0, c]} \ell(dy) \mu(A_y) = \int_{\mathbb{R}^d} f d\mu.$$

On obtient le même résultat pour toute fonction Borélienne bornée à discontinuités μ -négligeables, en raisonnant sur ses parties positive et négative, ce qui prouve (v). ■

On en déduit le théorème suivant.

Théorème V.1.4 ◀▶ Soit $\Phi : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$, une fonction mesurable. Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $X_n : \Omega \rightarrow \mathbb{R}^{d_1}$, $n \in \mathbb{N}$, et $X : \Omega \rightarrow \mathbb{R}^{d_1}$, des v.a. \mathcal{F} -mesurables. On fait les hypothèses suivantes.

(a) $X_n \Longrightarrow X$.

(b) \mathbf{P} -p.s. Φ est continue en X .

Alors, $\Phi(X_n) \Longrightarrow \Phi(X)$.

Preuve : soit $g \in C_b(\mathbb{R}^{d_2})$. Alors, (b) implique que $g \circ \Phi$ est une fonction $\mathcal{B}(\mathbb{R}^{d_1})$ -mesurable dont l'ensemble des points de discontinuité est μ_X -négligeable. On utilise le théorème V.1.3 de Portemanteau (v) \Rightarrow (i) pour montrer que

$$\lim_{n \rightarrow \infty} \mathbf{E}[g(\Phi(X_n))] = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^{d_1}} g \circ \Phi d\mu_{X_n} = \int_{\mathbb{R}^{d_1}} g \circ \Phi d\mu_X = \mathbf{E}[g(\Phi(X))],$$

ce qui implique le résultat voulu. ■

La proposition suivante montre que la convergence en probabilité implique la convergence en loi.

Proposition V.1.5 ◀▶ Soient $X_n : \Omega \rightarrow \mathbb{R}^d$, $n \in \mathbb{N}$, et $X : \Omega \rightarrow \mathbb{R}^d$, des v.a. \mathcal{F} -mesurables. On suppose que $\lim_n X_n = X$ en probabilité. Alors $X_n \Longrightarrow X$ en loi.

Preuve : soit $f \in C_c(\mathbb{R}^d)$. Pour tout $\varepsilon > 0$, on pose

$$w_f(\varepsilon) = \sup \{ |f(x) - f(y)|; x, y \in \mathbb{R}^d : \|x - y\| \leq \varepsilon \}$$

qui le module d'uniforme continuité de f . Comme f est continue à support compact, elle est uniformément continue et $\lim_\varepsilon w_f(\varepsilon) = 0$. On a ensuite des inégalité suivantes :

$$\begin{aligned} |\mathbf{E}[f(X_n)] - \mathbf{E}[f(X)]| &\leq \mathbf{E}[|f(X_n) - f(X)|] \\ &= \mathbf{E}[|f(X_n) - f(X)| \mathbf{1}_{\{\|X_n - X\| > \varepsilon\}}] + \mathbf{E}[|f(X_n) - f(X)| \mathbf{1}_{\{\|X_n - X\| \leq \varepsilon\}}] \\ &\leq 2\|f\|_\infty \mathbf{P}(\|X_n - X\| > \varepsilon) + w_f(\varepsilon), \end{aligned}$$

ce qui implique que $\limsup_n |\mathbf{E}[f(X_n)] - \mathbf{E}[f(X)]| \leq w_f(\varepsilon)$, pour tout réel $\varepsilon > 0$. Donc, pour toute fonction $f \in C_c(\mathbb{R}^d)$, $\lim_n \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)]$, ce qui entraîne $X_n \Longrightarrow X$ par le lemme V.1.2 (page V.1.2). ■

Remarque V.1.1 ◀▶ La convergence presque sûre entraîne la convergence en probabilité qui entraîne la convergence en loi. Dans \mathbb{C} , la convergence L^p entraîne la convergence en probabilité qui entraîne la convergence en loi. □

Lorsque la limite en loi est une constante, alors la convergence a lieu en probabilité.

Proposition V.1.6 ◀▶ Soient $X_n : \Omega \rightarrow \mathbb{R}^d$, $n \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. Soit $x \in \mathbb{R}^d$. On suppose que $X_n \Longrightarrow x$ en loi. Alors, $\lim_n X_n = x$ en probabilité.

Preuve : soit $\varepsilon > 0$. On applique le théorème V.1.3 au fermé $F = \mathbb{R}^d \setminus B(x, \varepsilon)$ et on a

$$\limsup_{n \rightarrow \infty} \mathbf{P}(\|X_n - x\| \geq \varepsilon) = \limsup_{n \rightarrow \infty} \mathbf{P}(X_n \in F) \leq \mathbf{P}(x \in F) = 0,$$

ce qui implique le résultat voulu ■

Lemme V.1.7 ◀▶ (Lemme de Slutsky) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $X_n : \Omega \rightarrow \mathbb{R}^{d_1}$, $n \in \mathbb{N}$, et $X : \Omega \rightarrow \mathbb{R}^{d_1}$, des v.a. \mathcal{F} -mesurables. Soient $Y_n : \Omega \rightarrow \mathbb{R}^{d_2}$, $n \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. Soit $y \in \mathbb{R}^{d_2}$. On fait les hypothèses suivantes.

(a) $X_n \Longrightarrow X$.

(b) $Y_n \Longrightarrow y$.

Alors $(X_n, Y_n) \Longrightarrow (X, y)$ en loi sur $\mathbb{R}^{d_1+d_2}$.

preuve : on munit \mathbb{R}^{d_1} d'une norme notée $\|\cdot\|_1$ et on munit \mathbb{R}^{d_2} d'une norme notée $\|\cdot\|_2$. On munit aussi $\mathbb{R}^{d_1+d_2} \cong \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ de la distance $\|(x, x')\| = \max(\|x\|_1, \|x'\|_2)$. Soit $f \in C_c(\mathbb{R}^{d_1+d_2})$. On note son module d'uniforme continuité par

$$w_f(\varepsilon) = \sup \{ |f(x, x') - f(y, y')|; (x, x'), (y, y') \in \mathbb{R}^{d_1+d_2} : \|(x, x') - (y, y')\| \leq \varepsilon \}$$

On pose

$$u_n = |\mathbf{E}[f(X_n, Y_n)] - \mathbf{E}[f(X, y)]| \quad \text{et} \quad v_n = |\mathbf{E}[f(X_n, y)] - \mathbf{E}[f(X, y)]|.$$

Soit $\varepsilon > 0$. On obtient les inégalités suivantes.

$$\begin{aligned} u_n &\leq \mathbf{E}[|f(X_n, Y_n) - f(X_n, y)|] + v_n \\ &\leq \mathbf{E}[|f(X_n, Y_n) - f(X_n, y)| \mathbf{1}_{\{\|Y_n - y\|_2 > \varepsilon\}}] + \mathbf{E}[|f(X_n, Y_n) - f(X_n, y)| \mathbf{1}_{\{\|Y_n - y\|_2 \leq \varepsilon\}}] + v_n \\ &\leq 2\|f\|_\infty \mathbf{P}(\|Y_n - y\|_2 > \varepsilon) + w_f(\varepsilon) + v_n. \end{aligned}$$

Par (a), on a $\lim_n v_n = 0$. Par la proposition V.1.5, on a $\limsup_n u_n \leq w_f(\varepsilon)$, ce qui montre que $\lim_n u_n = 0$, car ε peut être arbitrairement petit, ce qui implique le résultat voulu par le lemme V.1.2, page 148. ■

La remarque suivante montre qu'en général un vecteur aléatoire dont les composantes convergent en loi, individuellement, peut malgré cela ne pas converger en loi.

Remarque V.1.2 Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $\xi : \Omega \rightarrow \{-1, 1\}$, une v.a. telle que $\mathbf{P}(\xi = 1) = \mathbf{P}(\xi = -1) = 1/2$. Pour tout $n \in \mathbb{N}$, on pose $X_n = \xi$ et $Y_n = (-1)^n \xi$. Il est clair que $\mathbf{P}(Y_n = 1) = \mathbf{P}(Y_n = -1) = 1/2$. Il est clair que $X_n \Longrightarrow \xi$, $Y_n \Longrightarrow \xi$, car ces variables sont de loi constante, égale à celle de ξ . Il est également clair que les variables (X_n, Y_n) ne convergent pas en loi vers (ξ, ξ) : en fait ces variables ne convergent pas en loi. □

Le corollaire suivant découle du lemme de Slutsky (lemme V.1.7, page 151) et de la proposition V.1.5 (page 150).

Corollaire V.1.8 ◀▶ Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $X_n : \Omega \rightarrow \mathbb{R}^{d_1}$, $n \in \mathbb{N}$, et $X : \Omega \rightarrow \mathbb{R}^{d_1}$, des v.a. \mathcal{F} -mesurables. Soient $Y_n : \Omega \rightarrow \mathbb{R}^{d_2}$, $n \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. Soit $y \in \mathbb{R}^{d_2}$. On fait les hypothèses suivantes.

(a) $X_n \Longrightarrow X$.

(b) $Y_n \Longrightarrow y$.

Alors les assertions suivantes sont vérifiées.

(i) Si $d_1 = d_2$, alors $X_n + Y_n \Longrightarrow X + y$.

(ii) Si $d_2 = 1$, alors $Y_n \cdot X_n \Longrightarrow y \cdot X$.

(iii) Si $d_2 = 1$ et si $y \neq 0$, alors $Y_n^{-1} \cdot X_n \Longrightarrow y^{-1} \cdot X$.

(iv) Si $d_1 = d_2 = 1$, alors $\min(X_n, Y_n) \Longrightarrow \min(X, Y)$ et $\max(X_n, Y_n) \Longrightarrow \max(X, Y)$.

Preuve : on sait par le lemme de Slutsky (lemme V.1.7, page 151) que $(X_n, Y_n) \Longrightarrow (X, y)$ et on applique la proposition V.1.5 (page 150) à $\Phi(x, y) = x + y$ pour (i), à $\Phi(x, y) = y \cdot x$ pour (ii), à $\Phi(x, y) = y^{-1} \cdot x$ pour (iii) et à $\Phi(x, y) = \min(x, y)$ et $\Phi(x, y) = \max(x, y)$ pour (iv). ■

V.2 Convergence en loi pour des v.a. réelles.

V.2.a Reformulation en termes de la fonction de répartition.

On rappelle que les lois de v.a. réelles sont caractérisées par leur fonction de répartition. Dans cette section nous reformulons la convergence en loi en termes de fonction de répartition. Nous tirons également parti de cette approche pour obtenir des résultats approfondis sur la convergence en loi dans \mathbb{R} .

Définition V.2.1 On note $\mathbf{Rep}(\mathbb{R})$ l'ensemble des fonctions $F : \mathbb{R} \rightarrow [0, 1]$ qui sont croissantes continues à droite et telles que

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

L'ensemble $\mathbf{Rep}(\mathbb{R})$ est appelé l'ensemble des fonctions de répartition. □

On rappelle, sous forme de lemme, le résultat élémentaire suivant.

Lemme V.2.1 Pour toute mesure $\mu \in \mathcal{M}_1(\mathbb{R})$, on pose

$$\forall x \in \mathbb{R}, \quad F_\mu(x) = \mu(]-\infty, x]) .$$

Alors, $F_\mu \in \mathbf{Rep}(\mathbb{R})$

On rappelle également dans la proposition suivante la définition II.7.5, page 72, du pseudo-inverse continu à droite d'une fonction de répartition ainsi que les résultats de la proposition II.7.8, page 72.

Proposition V.2.2 ◀▶ Soit $F \in \mathbf{Rep}(\mathbb{R})$. On définit son pseudo-inverse continu à droite comme la fonction $G :]0, 1[\rightarrow \mathbb{R}$ donnée par

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

Alors, les assertions suivantes sont vérifiées.

(i) G est continue à droite croissante sur $]0, 1[$ et si F est strictement croissante continue alors G est la réciproque de F .

(ii) On note μ la mesure image de la mesure de Lebesgue sur $]0, 1[$ par G :

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mu(B) = \ell(G^{-1}(B)) .$$

Alors $F(x) = \mu(]-\infty, x])$, $x \in \mathbb{R}$. D'un point de vue probabiliste, cela peut être reformulé comme suit : soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité et $U : \Omega \rightarrow]0, 1[$, une v.a. \mathcal{F} -mesurable de loi uniforme. On pose $X = G(U)$, alors F est la fonction de répartition de X :

$$\forall x \in \mathbb{R}, \quad F(x) = \mathbf{P}(X \leq x) .$$

Preuve : voir la preuve de la proposition II.7.8, page 72. ■

On montre le lemme suivant discutant d'une forme de continuité pour les pseudo-inverses continus à droite.

Lemme V.2.3 Soient $F_n \in \mathbf{Rep}(\mathbb{R})$, $n \in \mathbb{N}$ et $F \in \mathbf{Rep}(\mathbb{R})$. On note G_n et G leurs pseudo-inverses continus à droite :

$$\forall n \in \mathbb{N}, \forall y \in]0, 1[, \quad G_n(y) = \inf \{x \in \mathbb{R} : F_n(x) > y\} \quad \text{et} \quad G(y) = \inf \{x \in \mathbb{R} : F(x) > y\} .$$

On suppose que

$$\forall x \in \mathbb{R} \text{ tel que } F(x) = F(x-), \quad \lim_{n \rightarrow \infty} F_n(x) = F(x) .$$

Autrement dit, on suppose que les F_n convergent vers F en tout point de continuité de F . Alors,

$$\forall y \in]0, 1[\text{ tel que } G(y) = G(y-), \quad \lim_{n \rightarrow \infty} G_n(y) = G(y) .$$

Autrement dit, les G_n convergent vers G en tout point de continuité de G .

Preuve : on fixe $y \in]0, 1[$. Soit x , un point de continuité de F tel que $x > G(y)$. Alors, $\lim_n F_n(x) = F(x) > y$ et pour tout n assez grand, $F_n(x) > y$; donc on a $x \geq G_n(y)$ pour tout n assez grand ce qui implique $x \geq \limsup_n G_n(y)$. Comme F n'a que des discontinuités dénombrables, x peut être choisi arbitrairement proche de $G(y)$ et ce qui précède entraîne que $G(y) \geq \limsup_n G_n(y)$.

Soient $0 < z < y < 1$, et soit x , un point de continuité de F tel que $x < G(z)$. On a donc $F(x) \leq z < y$ et donc, pour tout n assez grand, $F_n(x) < y$. Par conséquent, $x \leq G_n(y)$, pour tout n assez grand et on a $x \leq \liminf_n G_n(y)$. Comme F n'a que des discontinuités dénombrables, x peut être choisi arbitrairement proche de $G(z)$ et on obtient pour tout $z \in]0, y[$, $G(z) \leq \liminf_n G(y)$. Puisque G est croissante on a donc $G(y-) \leq \liminf_n G(y)$. En définitive, on a prouvé que

$$\forall y \in]0, 1[, \quad G(y-) \leq \liminf_{n \rightarrow \infty} G_n(y) \leq \limsup_{n \rightarrow \infty} G_n(y) \leq G(y) ,$$

ce qui implique le résultat voulu. ■

On en déduit le théorème principal de cette section qui reformule la convergence étroite en termes de fonction de répartition.

Théorème V.2.4 $\blacktriangleleft \blacktriangleright$ (Convergence en loi par les fonctions de répartition) Soient $\mu_n \in \mathcal{M}_1(\mathbb{R})$, $n \in \mathbb{N}$, et $\mu \in \mathcal{M}_1(\mathbb{R})$. Alors, les deux assertions suivantes sont équivalentes.

(a) $\lim_n \mu_n = \mu$ étroitement.

(b) Pour tout $x \in \mathbb{R}$ tel que $\mu(\{x\}) = 0$, on a $\lim_n \mu_n(]-\infty, x]) = \mu(]-\infty, x])$.

Ce résultat se reformule en termes probabilistes de la manière suivante : soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité ; soient $X_n : \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}$ et $X : \Omega \rightarrow \mathbb{R}$, des v.a. réelles \mathcal{F} -mesurables. Alors, les deux assertions suivantes sont équivalentes.

(c) $X_n \Longrightarrow X$.

(d) Pour tout $x \in \mathbb{R}$ où F_X est continue, on a $\lim_n F_{X_n}(x) = F_X(x)$, c'est-à-dire

$$\forall x \in \mathbb{R} \text{ tel que } \mathbf{P}(X = x) = 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}(X_n \leq x) = \mathbf{P}(X \leq x).$$

Preuve : on montre d'abord que (a) \implies (b). Soit $x \in \mathbb{R}$ tel que $\mu(\{x\}) = 0$. On pose $A =]-\infty, x]$. On a donc $\mu(\bar{A}) = \mu(\dot{A})$ et le théorème V.1.3 de Portemanteau (v) \implies (i) implique (b). On remarque ensuite que l'implication (a) \implies (b) est équivalente à l'implication (c) \implies (d).

Montrons que (b) \implies (a). Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et $U : \Omega \rightarrow [0, 1]$, une v.a. \mathcal{F} -mesurable de loi uniforme. On note F_n la fonction de répartition de μ_n et F , celle de μ . On note également G_n le pseudo-inverse continu à droite de F_n et G , celui de F . On note D l'ensemble des points de discontinuité de G : puisque G est croissante, D est dénombrable. L'hypothèse (b) permet d'appliquer le lemme V.2.3 qui implique que pour tout $y \notin D$, $\lim_n G_n(y) = G(y)$. Comme D est dénombrable, il est Lebesgue négligeable et donc \mathbf{P} -p.s. $U \notin D$, ce qui implique que \mathbf{P} -p.s. $\lim_n G_n(U) = G(U)$. Or, par la proposition V.2.2, la v.a. $G_n(U)$ a pour loi μ_n et la v.a. $G(U)$ a pour loi μ . Comme la convergence presque sûre entraîne la convergence en loi (proposition V.1.5, page 150), on en déduit (a). Cela montre (a) \iff (b), ce qui est équivalent à montrer (c) \iff (d). Cela termine la preuve. ■

V.2.b Application au théorème Central-Limite de De Moivre-Laplace

Appliquons le théorème précédent en étudiant l'écart à la moyenne de somme de Bernoulli indépendantes. Soit $(X_n)_{n \geq 1}$, une suite de v.a. de Bernoulli indépendantes et de même paramètre $p \in]0, 1[$. Pour tout $n \geq 1$, on pose

$$S_n = X_1 + \dots + X_n.$$

On sait que S_n suit une loi Binomiale de paramètres (n, p) . La moyenne de S_n est donc $\mathbf{E}[S_n] = np$, c'est-à-dire que $\frac{1}{n}S_n$ a pour moyenne p . Comme les variables X_n sont bornées (elles valent 0 ou 1), elles admettent un moment d'ordre 4 et la loi des grands nombres de Borel (proposition III.3.2, page 112) s'applique et on permet d'affirmer que

$$\mathbf{P}\text{-p.s.} \quad \frac{1}{n}S_n = \frac{X_1 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} p = \mathbf{E}[X_1].$$

On cherche ensuite à évaluer l'ordre de grandeur de l'écart $\frac{1}{n}S_n - p$.

L'écart de S_n à sa moyenne peut être mesuré partiellement par sa variance : la proposition III.2.12 page 107 implique que

$$\mathbf{E}[(S_n - np)^2] = \mathbf{var}(S_n) = \sum_{1 \leq k \leq n} \mathbf{var}(X_k) = np(1-p).$$

On a donc une convergence de $\frac{1}{n}S_n$ vers p en norme L^2 :

$$\left\| \frac{1}{n}S_n - p \right\|_2 = \sqrt{\frac{p(1-p)}{n}} \xrightarrow{n \rightarrow \infty} 0.$$

En fait on montre que $\sqrt{n}(\frac{1}{n}S_n - p)$ se répartit approximativement comme une Gaussienne, c'est-à-dire, informellement que

$$\frac{1}{n}S_n \approx p + \sqrt{\frac{p(1-p)}{n}} X \quad \text{où } X \sim \mathcal{N}(0, 1).$$

Plus précisément, on montre le théorème de De Moivre-Laplace.

Théorème V.2.5 (Théorème Central-Limite de De Moivre-Laplace) Soit $(X_n)_{n \geq 1}$, une suite de v.a. de Bernoulli indépendantes et de même paramètre $p \in]0, 1[$. Pour tout $n \geq 1$, on pose $S_n = X_1 + \dots + X_n$ et $\sigma = p(1-p)$. Alors, pour tous $a, b \in \mathbb{R}$ tels que $a < b$, on a

$$\mathbf{P}\left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n}S_n - p\right) \leq b\right) \xrightarrow[n \rightarrow \infty]{} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Preuve : pour tout $x \in \mathbb{R}$, on pose

$$x_n = \lfloor pn + \sigma\sqrt{n}x \rfloor \quad \text{et} \quad q_n(x) = \mathbf{P}(S_n = x_n).$$

On vérifie facilement que

$$\sigma\sqrt{n} \int_{a+\frac{1}{\sigma\sqrt{n}}}^b q_n(x) dx \leq \mathbf{P}\left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n}S_n - p\right) \leq b\right) \leq \sigma\sqrt{n} \int_a^b q_n(x) dx. \quad (\text{V.6})$$

On remarque ensuite que

$$p + \frac{\sigma}{\sqrt{n}}x - \frac{1}{n} \leq \frac{x_n}{n} \leq p + \frac{\sigma}{\sqrt{n}}x.$$

Donc, pour tout n assez grand, pour tout $x \in [a, b]$, on a $|x_n| \leq n$ et on peut écrire

$$q_n(x) = \binom{n}{x_n} p^{x_n} (1-p)^{n-x_n} = \frac{n!}{x_n!(n-x_n)!} \exp(x_n \log p + (n-x_n) \log(1-p)).$$

La formule de Stirling affirme que $n! = \sqrt{2\pi n}(n/e)^n u(n)$, avec $u(n) = 1 + \mathcal{O}(1/n)$. On a donc $q_n(x) = F_n(x)G_n(x)$ où

$$F_n(x) = \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{\sqrt{x_n(n-x_n)}} \frac{u(n)}{u(x_n)u(n-x_n)}$$

et

$$\log G_n(x) = n \log n - x_n \log x_n - (n-x_n) \log(n-x_n) + x_n \log p + (n-x_n) \log(1-p).$$

Il est d'abord facile de montrer qu'il existe $b_n : [a, b] \rightarrow \mathbb{R}$ telle que $\sup_{n \geq 0} \sup_{[a, b]} b_n < \infty$ et

$$\sigma\sqrt{n} F_n(x) = \frac{1}{\sqrt{2\pi}} \left(1 + \frac{b_n(x)}{\sqrt{n}}\right).$$

On remarque ensuite que si on pose $H(y) = y \log y$, on a

$$-\frac{1}{n} \log G_n(x) = pH\left(\frac{x_n}{np}\right) + (1-p)H\left(\frac{n-x_n}{n(1-p)}\right).$$

Un calcul long mais élémentaire montre qu'il existe $c_n : [a, b] \rightarrow \mathbb{R}$ telle que $\sup_{n \geq 0} \sup_{[a, b]} c_n < \infty$ et

$$\log G_n(x) = -\frac{x^2}{2} + \frac{c_n(x)}{\sqrt{n}}.$$

On en déduit qu'il existe une constante $C_{a,b,p}$ ne dépendant que de a, b et p telle que

$$\sup_{x \in [a, b]} |\sigma \sqrt{n} q_n(x) - (2\pi)^{-1/2} e^{-x^2/2}| \leq \frac{C_{a,b,p}}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} 0,$$

ce qui permet de conclure par (V.6). ■

V.3 convergence en loi sur \mathbb{N} .

V.3.a Premiers résultats.

Faisons quelques brefs rappels. Soit $\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ une loi de probabilité qui est concentrée sur les entiers, c'est-à-dire que $\mu(\mathbb{N}) = 1$. Elle est entièrement caractérisée par les nombres $\mu(\{n\})$, $n \in \mathbb{N}$. En effet, on vérifie que pour tout $A \subset \mathbb{N}$, A est la réunion dénombrable des singletons le constituant : $A = \bigcup_{n \in A} \{n\}$. Comme toute famille de singletons est deux-à-deux disjointe, on a

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

Autrement dit, μ est purement atomique : c'est la somme pondérée de masses de Dirac suivante :

$$\mu = \sum_{n \in \mathbb{N}} \mu(\{n\}) \delta_n.$$

De même pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée :

$$\int f d\mu = \sum_{n \in \mathbb{N}} f(n) \mu(\{n\}). \quad (\text{V.7})$$

Réciproquement on se donne une suite a_n , $n \in \mathbb{N}$ telle que $a_n \geq 0$ pour tout $n \in \mathbb{N}$ et $\sum_{n \in \mathbb{N}} a_n = 1$. Alors pour tout $A \subset \mathbb{N}$, on pose $\nu(A) = \sum_{n \in A} a_n$. Alors on vérifie facilement que ν est une mesure sur la tribu $\mathcal{P}(\mathbb{N})$ de tous les sous-ensembles de \mathbb{N} et que c'est une mesure de probabilité sur \mathbb{N} .

Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité et soit $X : \Omega \rightarrow \mathbb{N}$. Alors elle est \mathcal{F} -mesurable si et seulement si les événements $\{X = n\} \in \mathcal{F}$. Ce qui précède montre que la loi de X est alors déterminée entièrement par la suite $(\mathbf{P}(X = n))_{n \in \mathbb{N}}$ car par définition $\mu_X(\{n\}) = \mathbf{P}(X = n)$ et on a

$$\forall A \subset \mathbb{N}, \quad \mathbf{P}(X \in A) = \sum_{n \in A} \mathbf{P}(X = n).$$

On note $\mathcal{M}_1(\mathbb{N})$ les mesures de probabilité sur \mathbb{N} . On montre tout d'abord le résultat suivant.

Théorème V.3.1 \blacktriangleleft Soient $\mu_p, \mu \in \mathcal{M}_1(\mathbb{N})$, $p \in \mathbb{N}$. Alors les deux assertions suivantes sont équivalentes.

(a) $\mu_p \rightarrow \mu$ en loi.

(b) Pour tout $n \in \mathbb{N}$, $\lim_{p \rightarrow \infty} \mu_p(\{n\}) = \mu(\{n\})$.

Ce résultat s'interprète de manière probabiliste comme suit : soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient $X_p, X : \Omega \rightarrow \mathbb{N}$, $p \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. Alors les assertions suivantes sont équivalentes.

(c) $X_p \Longrightarrow X$.

(d) Pour tout $n \in \mathbb{N}$, $\lim_{p \rightarrow \infty} \mathbf{P}(X_p = n) = \mathbf{P}(X = n)$.

◀► **Preuve** : tout d'abord, rappelons que l'on voit les mesures de probabilité sur \mathbb{N} comme des mesures de probabilité sur \mathbb{R} particulières. On suppose tout d'abord (a) : alors pour toute fonction $f \in C_b(\mathbb{R})$, et par (V.7), on a

$$\int f d\mu_p = \sum_{n \in \mathbb{N}} f(n) \mu_p(\{n\}) \xrightarrow{p \rightarrow \infty} \sum_{n \in \mathbb{N}} f(n) \mu(\{n\}) = \int f d\mu. \quad (\text{V.8})$$

On fixe $n_0 \in \mathbb{N}$ et on applique (V.8) à une fonction continue qui est nulle en dehors de l'intervalle $[n_0 - \frac{1}{2}, n_0 + \frac{1}{2}]$ et telle $f(n_0) = 1$ (il en existe évidemment : exercice). On a $\int f d\mu_p = \mu_p(\{n_0\})$ et $\int f d\mu = \mu(\{n_0\})$. Donc (V.8) implique que $\lim_p \mu_p(\{n_0\}) = \mu(\{n_0\})$. Comme cela est vérifié pour tout $n_0 \in \mathbb{N}$, cela implique (b).

Réciproquement, supposons (b). Soit $\psi \in C_c(\mathbb{R})$. Comme ψ est à support compact il existe $n_0 \in \mathbb{N}$ tel que ψ est nulle en dehors de l'intervalle $[-n_0, n_0]$. On en déduit donc que

$$\int \psi d\mu_p = \sum_{n \in \mathbb{N}} \psi(n) \mu_p(\{n\}) = \psi(0) \mu_p(\{0\}) + \dots + \psi(n_0) \mu_p(\{n_0\}).$$

Comme la somme est finie (b) implique que

$$\int \psi d\mu_p = \psi(0) \mu_p(\{0\}) + \dots + \psi(n_0) \mu_p(\{n_0\}) \xrightarrow{p \rightarrow \infty} \psi(0) \mu(\{0\}) + \dots + \psi(n_0) \mu(\{n_0\}) = \int \psi d\mu.$$

On a donc montré que pour toute fonction $\psi \in C_c(\mathbb{R})$, $\lim_p \int \psi d\mu_p = \int \psi d\mu$. Le lemme V.1.2 (page 148) entraîne que $\mu_p \rightarrow \mu$ en loi, c'est-à-dire (a). Cela termine la preuve de (a) \iff (b). La reformulation en termes probabiliste est immédiate. ■

On introduit ensuite la définition suivante.

Définition V.3.1 (Distance en variation) Pour toutes $\mu, \nu \in \mathcal{M}_1(\mathbb{N})$, on pose

$$d_{\text{var}}(\mu, \nu) = \sum_{n \in \mathbb{N}} |\mu(\{n\}) - \nu(\{n\})|.$$

Il est facile de vérifier que d_{var} est une distance sur $\mathcal{M}_1(\mathbb{N})$. □

Proposition V.3.2 Soient $\mu_p, \mu \in \mathcal{M}_1(\mathbb{N})$, $p \in \mathbb{N}$. Alors, les deux assertions suivantes sont équivalentes.

(a) $\mu_p \rightarrow \mu$ en loi.

(b) $d_{\text{var}}(\mu_p, \mu) \rightarrow 0$.

Preuve : montrons tout d'abord l'implication (b) \implies (a). On observe que pour tout $n_0 \in \mathbb{N}$, on a $|\mu_p(\{n_0\}) - \mu(\{n_0\})| \leq d_{\text{var}}(\mu_p, \mu)$. Cela montre que pour tout $n_0 \in \mathbb{N}$, on a $\lim_{p \rightarrow \infty} \mu_p(\{n_0\}) = \mu(\{n_0\})$, ce qui par le théorème V.3.1 ((b) \implies (a)) entraîne que $\mu_p \rightarrow \mu$ en loi, c'est-à-dire (a).

Réciproquement, supposons (a). Soit $n_0 \in \mathbb{N}$ qui est spécifié plus loin. On observe tout d'abord que l'on a les inégalités suivantes.

$$\begin{aligned} d_{\text{var}}(\mu_p, \mu) &= \sum_{0 \leq n \leq n_0} |\mu_p(\{n\}) - \mu(\{n\})| + \sum_{n > n_0} |\mu_p(\{n\}) - \mu(\{n\})| \\ &\leq \sum_{0 \leq n \leq n_0} |\mu_p(\{n\}) - \mu(\{n\})| + \sum_{n > n_0} \mu_p(\{n\}) + \sum_{n > n_0} \mu(\{n\}) \\ &\leq \sum_{0 \leq n \leq n_0} |\mu_p(\{n\}) - \mu(\{n\})| + 1 - \sum_{0 \leq n \leq n_0} \mu_p(\{n\}) + 1 - \sum_{0 \leq n \leq n_0} \mu(\{n\}) \\ &\leq \sum_{0 \leq n \leq n_0} |\mu_p(\{n\}) - \mu(\{n\})| + 2 - \sum_{0 \leq n \leq n_0} \mu_p(\{n\}) - \sum_{0 \leq n \leq n_0} \mu(\{n\}). \end{aligned}$$

En utilisant le théorème V.3.1 ((a) \implies (b)), on voit que pour tout $n \in \mathbb{N}$, $\lim_p \mu_p(\{n\}) = \mu(\{n\})$. Comme il s'agit de sommes finies, cela implique que

$$\lim_{p \rightarrow \infty} \sum_{0 \leq n \leq n_0} |\mu_p(\{n\}) - \mu(\{n\})| \quad \text{et} \quad \lim_{p \rightarrow \infty} \sum_{0 \leq n \leq n_0} \mu_p(\{n\}) = \sum_{0 \leq n \leq n_0} \mu(\{n\}).$$

Donc, pour tout $n_0 \in \mathbb{N}$,

$$\limsup_{p \rightarrow \infty} d_{\text{var}}(\mu_p, \mu) \leq 2 \left(1 - \sum_{0 \leq n \leq n_0} \mu(\{n\}) \right).$$

Or $\lim_{n_0 \rightarrow \infty} \sum_{0 \leq n \leq n_0} \mu(\{n\}) = 1$, donc ce qui précède implique que $\limsup_{p \rightarrow \infty} d_{\text{var}}(\mu_p, \mu) = 0$, ce qui implique (b). Cela termine la preuve de la proposition. \blacksquare

V.3.b Application aux variables de Poisson et aux événements rares.

On rappelle la définition d'une loi de Poisson.

Définition V.3.2 (*Loi de Poisson*) Une v.a. $X : \Omega \rightarrow \mathbb{N}$ est (de loi) de Poisson de paramètre $\theta \in]0, \infty[$ si $\mathbf{P}(X = k) = e^{-\theta} \theta^k / k!$, pour tout $k \in \mathbb{N}$. \square

On rappelle que si X est de Poisson de paramètre θ , alors

$$\mathbf{E}[r^X] = \exp(-\theta(1-r)), \quad r \in [0, 1]$$

et

$$\mathbf{E}[X] = \theta, \quad \mathbf{E}[X^2] = \theta + \theta^2 \quad \text{et donc} \quad \text{var}(X) = \theta.$$

Cela justifie la définition suivante des lois de Poisson dégénérées.

Définition V.3.3 (*Loi de Poisson dégénérée*) Une v.a. X suit une loi de Poisson de paramètre infini (resp. nul) si $\mathbf{P}(X = \infty) = 1$ (resp. si $\mathbf{P}(X = 0) = 1$). \square

Lemme V.3.3 Soient X_1, \dots, X_n , des v.a. de Poisson de paramètres respectifs $\theta_1, \dots, \theta_n$. On les suppose indépendantes. Alors $X_1 + \dots + X_n$ est une v.a. de Poisson de paramètre $\theta_1 + \dots + \theta_n$.

Plus généralement, soit $(X_n)_{n \in \mathbb{N}}$ des v.a. de Poisson indépendantes. On note θ_n le paramètre de X_n . Alors $X := \sum_{n \in \mathbb{N}} X_n$ est une v.a. de Poisson (éventuellement dégénérée) de paramètre $\theta := \sum_{n \in \mathbb{N}} \theta_n$.

Preuve : on montre d'abord le premier point. Pour cela, on observe que

$$\mathbf{E}[r^{X_1 + \dots + X_n}] = \mathbf{E}[r^{X_1}] \dots \mathbf{E}[r^{X_n}] = \exp(-(\theta_1 + \dots + \theta_n)(1 - r)) ,$$

ce qui permet de conclure puisque les fonctions génératrices caractérisent les lois sur \mathbb{N} .

Montrons le second point du lemme. Pour tout $n \in \mathbb{N}$, on pose $S_n = X_0 + \dots + X_n$ et $t_n = \theta_0 + \dots + \theta_n$. On suppose d'abord que $\theta = \sum_{n \in \mathbb{N}} \theta_n < \infty$. On remarque que $S_n \leq S_{n+1}$ et que $\lim_n S_n = X$. Par convergence dominée et d'après le premier point du lemme, on a donc pour tout $r \in [0, 1]$,

$$\mathbf{E}[r^X] = \lim_{n \rightarrow \infty} \mathbf{E}[r^{S_n}] = \lim_{n \rightarrow \infty} \exp(-t_n(1 - r)) = \exp(-\theta(1 - r)) ,$$

ce qui implique que X suit une loi de Poisson de paramètre θ .

On suppose ensuite que $\theta = \infty$. S'il existe $n \in \mathbb{N}$ tel que $\theta_n = \infty$, alors $X_n = \infty$ p.s. et donc $X = \infty$, ce qui montre le résultat voulu. Il reste à examiner le cas où pour tout $n \in \mathbb{N}$, $\theta_n < \infty$. Pour tout $x \in \mathbb{R}_+$ et tout $n \in \mathbb{N}$, une inégalité de Markov implique que

$$\mathbf{P}(S_n \leq x) \leq e^x \mathbf{E}[e^{-S_n}] = e^{x - t_n(1 - e^{-1})} .$$

Donc $\lim_n \mathbf{P}(S_n \leq x) = 0$. Or $\{S < x\} \subset \bigcap_{n \in \mathbb{N}} \{S_n \leq x\}$, donc $\mathbf{P}(X < x) = 0$, pour tout $x \in \mathbb{R}_+$, ce qui implique que p.s. $X = \infty$, ce qui permet de conclure. ■

Approximation des lois de Poisson par des loi binomiales. Les lois de Poisson apparaissent comme des loi binomiales dégénérées : plus précisément, on fixe $\theta > 0$ et pour tout $n \in \mathbb{N}^*$ tel que $n > \theta$, on considère $Y_{n,1}, \dots, Y_{n,n}$, des variables de Bernoulli indépendantes de même paramètre $p_n = \theta/n$. On pose

$$X_n = \sum_{1 \leq k \leq n} Y_{n,k} .$$

On constate que $\mathbf{P}(Y_{n,k} = 1) = p_n$ devient très petite lorsque $n \rightarrow \infty$. En revanche on voit que $\mathbf{E}[X_n] = np_n = \theta$, qui ne varie pas avec n . La loi X_n est une binomiale de paramètres (n, p_n) . Pour tout $k \in \mathbb{N}$, et tout $n \geq k$, on a

$$\mathbf{P}(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\theta^k}{k!} \left(1 - \frac{1}{n}\theta\right)^{n-k} \prod_{1 \leq j \leq k-1} \left(1 - \frac{j}{n}\right)$$

Cela montre que

$$\forall k \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} \mathbf{P}(X_n = k) = e^{-\theta} \theta^k / k! ,$$

Par le théorème V.3.1 (page 156) cela implique le résultat suivant.

Proposition V.3.4 Soit $\theta \in]0, \infty[$. On note X , une variable suivant une loi de Poisson de paramètre θ . Pour tout $n \in \mathbb{N}^*$ tel que $n > \theta$, on considère $Y_{n,1}, \dots, Y_{n,n}$, des variables de Bernoulli indépendantes de même paramètre $p_n = \theta/n$. Alors

$$\sum_{1 \leq k \leq n} Y_{n,k} \implies X .$$

Le théorème suivant donne un résultat plus précis.

Théorème V.3.5 (Approximation Binomiale-Poisson) *Pour tout $n \geq 1$, on se donne une suite $(Y_{n,k})_{1 \leq k \leq n}$ de variables indépendantes de Bernoulli de paramètres respectifs $p_{n,k}$, c'est-à-dire $\mathbf{P}(Y_{n,k} = 1) = p_{n,k}$. On pose*

$$X_n = \sum_{1 \leq k \leq n} Y_{n,k} \quad \text{et} \quad \theta_n = \sum_{1 \leq k \leq n} p_{n,k} .$$

Alors pour tout $\theta > 0$,

$$\sum_{k \in \mathbb{N}} \left| \mathbf{P}(X_n = k) - e^{-\theta} \frac{\theta^k}{k!} \right| \leq 2|\theta - \theta_n| + 2 \sum_{1 \leq k \leq n} p_{n,k}^2 .$$

Si $\lim_n \theta_n = \theta$ et si $\lim_n \max_{1 \leq k \leq n} p_{n,k} = 0$, alors $\lim_n \sum_{1 \leq k \leq n} p_{n,k}^2 = 0$, et

$$\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} \left| \mathbf{P}(X_n = k) - e^{-\theta} \frac{\theta^k}{k!} \right| = 0 .$$

Avant de prouver ce théorème, voyons un exemple d'application de cette approximation : les statistiques nous disent que 3,2 piétons sont écrasés à Paris chaque mois (c'est un chiffre fantaisiste). Quelle est la probabilité que, le mois prochain, exactement 5 piétons soient écrasés à Paris ? Pour répondre rapidement à cette question on peut établir le modèle (grossier) suivant : il y a (environ) $n \simeq 2.10^6$ parisiens qui ont une "chance" très petite de se faire écraser dans le mois suivant, et ce, indépendamment les uns des autres. On numérote les parisiens de 1 à n et on note Y_k la variable aléatoire qui vaut 1 si le parisien numero k se fait écraser le mois prochain et qui vaut 0 sinon. D'après nos hypothèses, on peut considérer $(Y_k)_{1 \leq k \leq n}$ comme une suite i.i.d. de variables de Bernoulli de paramètre p . Par conséquent $X = Y_1 + \dots + Y_n$ suit une loi Binomiale de paramètres (n, p) . Sa moyenne vaut $np = 3,2$. En appliquant la proposition précédente avec $np = \theta = 3,2$, on obtient

$$\sum_{k \in \mathbb{N}} \left| \mathbf{P}(X = k) - e^{-3,2} \frac{(3,2)^k}{k!} \right| \leq (3,2)^2/n .$$

Or $(3,2)^2/n$ est de l'ordre de 10^{-5} . On peut donc assimiler la probabilité que le mois prochain exactement 5 piétons soient écrasés à Paris à la quantité $e^{-3,2} (3,2)^5 / 5!$.

Preuve du théorème V.3.5.

Pour démontrer le théorème V.3.5, on cherche à comparer les mesures de probabilités sur \mathbb{Z} . Une mesure de probabilité μ sur \mathbb{Z} est simplement donnée par $\mu = (\mu(k))_{k \in \mathbb{Z}}$ où

$$\forall k \in \mathbb{Z}, \quad \mu(k) \geq 0 \quad \text{et} \quad \sum_{k \in \mathbb{Z}} \mu(k) = 1 .$$

On note $\mathcal{M}_1(\mathbb{Z})$ l'ensemble des mesures de probabilités sur \mathbb{Z} .

Définition V.3.4 (*Distance en variation pour les lois sur \mathbb{Z}*) Pour toutes mesures $\mu, \nu \in \mathcal{M}_1(\mathbb{Z})$, on pose

$$d_{\text{var}}(\mu, \nu) = \sum_{k \in \mathbb{Z}} |\mu(k) - \nu(k)| .$$

Cette quantité est appelée *distance en variation de μ et ν* . □

On peut voir $\mathcal{M}_1(\mathbb{Z})$ comme un sous-ensemble de l'espace des suite $\ell^1(\mathbb{Z})$ et d_{var} n'est autre que la norme L^1 . Il est facile de vérifier $\mathcal{M}_1(\mathbb{Z})$ est un fermé de $\ell^1(\mathbb{Z})$ ce qui implique que $(\mathcal{M}_1(\mathbb{Z}), d_{\text{var}})$ est un espace métrique séparable complet. Soient $\mu, \nu \in \mathcal{M}_1(\mathbb{Z})$, on rappelle que $\mu * \nu$ est la convolée de μ avec ν . Il est clair que $\mu * \nu \in \mathcal{M}_1(\mathbb{Z})$: précisément, on a

$$\forall k \in \mathbb{Z}, \quad \mu * \nu(k) = \sum_{j \in \mathbb{N}} \mu(k-j)\nu(j) .$$

On rappelle que si X et Y sont deux v.a. à valeurs dans \mathbb{Z} , \mathcal{F} -mesurables et indépendantes, alors la loi de $X + Y$ est $\mu_X * \mu_Y$.

Lemme V.3.6 Soient $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}_1(\mathbb{Z})$. Alors

$$d_{\text{var}}(\mu_1 * \mu_2, \nu_1 * \nu_2) \leq d_{\text{var}}(\mu_1, \nu_1) + d_{\text{var}}(\mu_2, \nu_2) .$$

Preuve : on a

$$\begin{aligned} d_{\text{var}}(\mu_1 * \mu_2, \nu_1 * \nu_2) &\leq \sum_{j, k \in \mathbb{Z}} |\mu_2(k-j)\mu_1(j) - \nu_2(k-j)\nu_1(j)| = \sum_{j, k \in \mathbb{Z}} |\mu_2(k)\mu_1(j) - \nu_2(k)\nu_1(j)| \\ &\leq \sum_{j, k \in \mathbb{Z}} |\mu_2(k)\mu_1(j) - \mu_2(k)\nu_1(j)| + \sum_{j, k \in \mathbb{Z}} |\mu_2(k)\nu_1(j) - \nu_2(k)\nu_1(j)| \\ &= \sum_{k \in \mathbb{Z}} \mu_2(k) d_{\text{var}}(\mu_1, \nu_1) + \sum_{j \in \mathbb{Z}} \nu_1(j) d_{\text{var}}(\mu_2, \nu_2) = d_{\text{var}}(\mu_1, \nu_1) + d_{\text{var}}(\mu_2, \nu_2), \end{aligned}$$

ce qui termine la preuve. ■

On introduit la notation suivante

$$\forall p \in [0, 1], \quad \mu_p = p\delta_1 + (1-p)\delta_0 \quad \text{et} \quad \forall \theta \in]0, \infty[\quad \nu_\theta = \sum_{k \in \mathbb{N}} \frac{1}{k!} e^{-\theta} \theta^k \delta_k . \quad (\text{V.9})$$

Autrement dit μ_p est la loi de Bernoulli de paramètre p et ν_θ la loi de Poisson de paramètre θ .

Lemme V.3.7 Pour tout $p \in]0, 1]$, on a $d_{\text{var}}(\mu_p, \nu_p) \leq 2p^2$. Pour tous $\theta, \theta' \in]0, \infty[$, on a également $d_{\text{var}}(\nu_\theta, \nu_{\theta'}) \leq 2|\theta - \theta'|$.

Preuve : par convexité on rappelle que $0 \leq 1 - e^{-p} \leq p$. Or

$$\begin{aligned} d_{\text{var}}(\mu_p, \nu_p) &= (e^{-p} - 1 + p) + (p - pe^{-p}) + \sum_{k \geq 2} \frac{p^k}{k!} e^{-p} \\ &= (e^{-p} - 1 + p) + (p - pe^{-p}) + (1 - e^{-p} - pe^{-p}) = 2p(1 - e^{-p}) \leq 2p^2, \end{aligned}$$

ce qui montre la première inégalité. Montrons la seconde : sans perte de généralité on suppose que $\theta < \theta'$. On pose $\delta = \theta' - \theta$. Le lemme ?? implique $\nu_{\theta'} = \nu_{\theta} * \nu_{\delta}$. Le lemma V.3.6 implique que

$$d_{\text{var}}(\nu_{\theta}, \nu_{\theta'}) = d_{\text{var}}(\nu_{\theta} * \delta_0, \nu_{\theta} * \nu_{\delta}) \leq d_{\text{var}}(\nu_{\theta}, \nu_{\theta}) + d_{\text{var}}(\delta_0, \nu_{\delta}) = d_{\text{var}}(\delta_0, \nu_{\delta}) .$$

Or $d_{\text{var}}(\delta_0, \nu_{\theta' - \theta}) = 1 - e^{-\delta} + \sum_{k \geq 1} \nu_{\delta}(k) = 2(1 - e^{-\delta}) \leq 2\delta$, ce qui permet de conclure. ■

Preuve du théorème V.3.5 : on fixe n . Avec les notations de (V.9), $\mu_{X_n} = \mu_{p_{n,1}} * \dots * \mu_{p_{n,n}}$. Le lemme ?? implique également que $\nu_{\theta_n} = \nu_{p_{n,1}} * \dots * \nu_{p_{n,n}}$. Par les lemmes précédents, on obtient les inégalité suivantes :

$$\begin{aligned} d_{\text{var}}(\nu_{\theta}, \mu_{X_n}) &\leq d_{\text{var}}(\nu_{\theta}, \nu_{\theta_n}) + d_{\text{var}}(\nu_{\theta_n}, \mu_{X_n}) \leq 2|\theta - \theta_n| + d_{\text{var}}(\mu_{p_{n,1}} * \dots * \mu_{p_{n,n}}, \nu_{p_{n,1}} * \dots * \nu_{p_{n,n}}) \\ &= 2|\theta - \theta_n| + \sum_{1 \leq k \leq n} 2p_{n,k}^2 , \end{aligned}$$

ce qui permet de conclure. ■

V.4 Convergence en loi et fonction caractéristique.

V.4.a Le théorème de Paul Lévy.

On munit \mathbb{R}^n du produit scalaire canonique $\langle \cdot, \cdot \rangle$ et de la norme euclidienne associée. On rappelle pour toute mesure $\mu \in \mathcal{M}_1(\mathbb{R}^n)$, sa transformée de Fourier est donnée par

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \widehat{\mu}(\mathbf{u}) := \int_{\mathbb{R}^n} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} \mu(d\mathbf{x}) .$$

$\widehat{\mu} : \mathbb{R}^n \rightarrow \mathbb{C}$ est une fonction continue borné par 1 et $\widehat{\mu}(0) = 1$. On rappelle le théorème d'injection de la transformée de Fourier qui affirme que pour toutes $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^n)$, on a

$$\mu = \nu \quad \iff \quad \widehat{\mu} = \widehat{\nu} .$$

On montre le théorème suivant dû à Paul Lévy.

Théorème V.4.1 ◀▶ (Th. de continuité de Lévy, version faible) Soient $\mu, \mu_p : \mathcal{B}(\mathbb{R}^n) \rightarrow \mathbb{R}_+$, $p \in \mathbb{N}$, des mesures finies. Alors,

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \lim_{p \rightarrow \infty} \widehat{\mu}_p(\mathbf{u}) = \widehat{\mu}(\mathbf{u}) \quad \iff \quad \lim_{p \rightarrow \infty} \mu_p = \mu \text{ en loi.} \quad (\text{V.10})$$

Ce résultat se reformule de manière probabiliste comme suit : soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soient X et $X_p : \Omega \rightarrow \mathbb{R}^n$, $p \in \mathbb{N}$, des v.a. \mathcal{F} -mesurables. Alors

$$\forall \mathbf{u} \in \mathbb{R}^n, \quad \lim_{p \rightarrow \infty} \mathbf{E}[e^{i\langle \mathbf{u}, X_p \rangle}] = \mathbf{E}[e^{i\langle \mathbf{u}, X \rangle}] \quad \iff \quad X_p \implies X .$$

La preuve est longue et elle utilise plusieurs lemmes déjà démontrés dans la preuve du théorème d'injectivité. On rappelle les notations suivantes : on a montré dans l'exercice II.10.1 que pour tout $u \in \mathbb{R}$,

$$\int_{\mathbb{R}} \exp(iu - \frac{1}{2}x^2) = \sqrt{2\pi} \exp(-\frac{1}{2}u^2) dx .$$

Pour tout $\sigma > 0$, on pose

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad g_\sigma(\mathbf{x}) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2},$$

qui est la *densité Gaussienne centrée de variance σ^2* . On remarque que

$$g_\sigma(\mathbf{x}) = \sigma^{-n} g_1(\sigma^{-1}\mathbf{x}) \quad \text{et} \quad \int_{\mathbb{R}^n} g_\sigma d\ell_n = 1.$$

Un changement de variable linéaire combiné au théorème de Fubini permet également de montrer que

$$\widehat{g}_\sigma(\mathbf{u}) = \exp\left(-\frac{\sigma^2}{2}\|\mathbf{u}\|^2\right) = \left(\frac{\sqrt{2\pi}}{\sigma}\right)^n g_{\frac{1}{\sigma}}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^n. \quad (\text{V.11})$$

Autrement dit les fonctions du type g_σ sont laissées invariantes, à un facteur multiplicatif près, par la transformée de Fourier.

Pour toute $\mu \in \mathcal{M}_1(\mathbb{R}^n)$ et tout $\sigma > 0$, on pose

$$(g_\sigma * \mu)(\mathbf{x}) = \int_{\mathbb{R}^n} g_\sigma(\mathbf{x} - \mathbf{y}) d\mu(\mathbf{y}). \quad (\text{V.12})$$

Par le théorème II.10.1 de continuité des intégrales à paramètre $g_\sigma * \mu$ est continue. Par ailleurs on vérifie immédiatement que

$$\|g_\sigma * \mu\|_\infty \leq (\sigma\sqrt{2\pi})^{-n} < \infty. \quad (\text{V.13})$$

On rappelle ensuite le lemme II.10.8 (page 89) qui affirme que :

$$(g_\sigma * \mu)(\mathbf{x}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-\frac{\sigma^2}{2}\|\mathbf{u}\|^2} e^{i\langle \mathbf{x}, \mathbf{u} \rangle} \widehat{\mu}(-\mathbf{u}) d\ell_n(\mathbf{u}). \quad (\text{V.14})$$

Soit $\psi \in C_c(\mathbb{R}^n)$. Elle est uniformément continue et on note son *module d'uniforme continuité* par

$$\forall \delta \in \mathbb{R}_+, \quad \omega_\psi(\delta) = \sup \{ |\psi(\mathbf{x}) - \psi(\mathbf{y})|; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} - \mathbf{y}\| \leq \delta \}.$$

Clairement $\delta \mapsto \omega_\psi(\delta)$ est une application croissante donc mesurable et on vérifie que

$$\forall \psi \in C_c(\mathbb{R}^n), \quad \forall \delta \in \mathbb{R}_+, \quad \omega_\psi(\delta) \leq 2\|\psi\|_\infty \quad \text{et} \quad \lim_{\delta \rightarrow 0} \omega_\psi(\delta) = 0. \quad (\text{V.15})$$

On rappelle ensuite le lemme II.10.9, (page 90) qui affirme ce qui suit : soit $\mu \in \mathcal{M}_1(\mathbb{R}^n)$, soit $\psi \in C_c(\mathbb{R}^n)$ et soit $\sigma > 0$; alors,

$$\left| \int_{\mathbb{R}^n} \psi d\mu - \int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu) d\ell_n \right| \leq \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma\|\mathbf{z}\|) d\ell_n(\mathbf{z}) \xrightarrow{\sigma \rightarrow 0} 0. \quad (\text{V.16})$$

preuve du théorème V.4.1. Montrons tout d'abord que si $\mu_p \rightarrow \mu$ en loi alors les transformées de Fourier convergent. On fixe $\mathbf{u} \in \mathbb{R}^n$ et on pose $f(\mathbf{x}) = \cos(\langle \mathbf{u}, \mathbf{x} \rangle)$ et $h(\mathbf{x}) = \sin(\langle \mathbf{u}, \mathbf{x} \rangle)$. Il est clair que $f, h \in C_b(\mathbb{R}^n)$. Donc $\lim_p \int f d\mu_p = \int f d\mu$ et $\lim_p \int h d\mu_p = \int h d\mu$. Donc

$$\widehat{\mu}_p(\mathbf{u}) = \int f d\mu_p + i \int h d\mu_p \xrightarrow{p \rightarrow \infty} \int f d\mu + i \int h d\mu = \widehat{\mu}(\mathbf{u}),$$

cela pour tout $\mathbf{u} \in \mathbb{R}^n$.

Réciproquement on suppose que pour tout $\mathbf{u} \in \mathbb{R}^n$, on a $\lim_{p \rightarrow \infty} \widehat{\mu}_p(\mathbf{u}) = \widehat{\mu}(\mathbf{u})$. Soit $\psi \in C_c(\mathbb{R}^n)$. Par (V.16), pour tout $p \in \mathbb{N}$ et tout $\sigma > 0$, on a

$$\left| \int_{\mathbb{R}^n} \psi d\mu - \int_{\mathbb{R}^n} \psi d\mu_p \right| \leq \left| \int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu) d\ell_n - \int_{\mathbb{R}^n} \psi \cdot (g_\sigma * \mu_p) d\ell_n \right| + 2 \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}). \quad (\text{V.17})$$

Par (V.14) et par convergence dominée, pour tout $\mathbf{x} \in \mathbb{R}^n$ $\lim_p g_\sigma * \mu_p(\mathbf{x}) = g_\sigma * \mu(\mathbf{x})$. Par (V.13), on a de plus :

$$\forall p \in \mathbb{N}, \quad |\psi \cdot (g_\sigma * \mu_p)| \leq (\sigma \sqrt{2\pi})^{-n} |\psi|.$$

Or $\int |\psi| d\ell_n < \infty$, car $\psi \in C_c(\mathbb{R}^n)$. Donc par convergence dominée on a

$$\lim_{p \rightarrow \infty} \int \psi \cdot (g_\sigma * \mu_p) d\ell_n = \int \psi \cdot (g_\sigma * \mu) d\ell_n$$

et (V.17) implique que

$$\forall \sigma > 0, \quad \limsup_{p \rightarrow \infty} \left| \int_{\mathbb{R}^n} \psi d\mu - \int_{\mathbb{R}^n} \psi d\mu_p \right| \leq 2 \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}).$$

Or par convergence dominée $\lim_{\sigma \rightarrow 0} \int_{\mathbb{R}^n} g_1(\mathbf{z}) \omega_\psi(\sigma \|\mathbf{z}\|) d\ell_n(\mathbf{z}) = 0$. Donc $\lim_p \int \psi d\mu_p = \int \psi d\mu$. Cela est vrai pour toute fonction $\psi \in C_c(\mathbb{R}^n)$. La proposition V.1.2 (page 148) entraîne alors que $\mu_p \rightarrow \mu$ en loi. Cela montre donc l'équivalence (V.10). La formulation probabiliste est immédiate. Cela termine la preuve du théorème V.4.1. ■

V.4.b Le théorème Central-Limite sur \mathbb{R} , intervalles de confiance.

Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $X_n : \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, une suite de v.a. \mathcal{F} -mesurables. On fait les deux hypothèses suivantes

- Les variables X_n , $n \in \mathbb{N}$, sont indépendantes.
- Elles ont la même loi et admettent un moment d'ordre deux : $\mathbf{E}[X_n^2] < \infty$.

En particulier elles sont intégrable et on pose

$$S_n := X_1 + \dots + X_n, \quad m := \mathbf{E}[X_n] \quad \text{et} \quad \sigma^2 = \text{var}(X_n).$$

La loi vote des grands nombres permet d'affirmer que \mathbf{P} -p.s. $\frac{1}{n} S_n \rightarrow m$. On cherche ensuite à évaluer l'erreur $\frac{1}{n} S_n - m$. Une première façon de procéder est de calculer la norme L^2 de cette variable. Pour cela on remarque que

$$\mathbf{E}[S_n] = \mathbf{E}[X_1] + \dots + \mathbf{E}[X_n] = n.m.$$

Donc

$$\begin{aligned} \mathbf{E}\left[\left(\frac{1}{n} S_n - m\right)^2\right] &= \frac{1}{n^2} \mathbf{E}[(S_n - n.m)^2] \\ &= \frac{1}{n^2} \text{var}(S_n) \\ &= \frac{1}{n^2} (\text{var}(X_1) + \dots + \text{var}(X_n)) = \frac{1}{n^2} n\sigma^2 \\ &= \sigma^2/n. \end{aligned}$$

On en déduit

$$\|\frac{1}{n}S_n - m\|_2 = \sigma/\sqrt{n}.$$

Cela suggère que $\frac{\sqrt{n}}{\sigma}(\frac{1}{n}S_n - m)$ converge (en un certain sens) vers « quelque chose ». Le théorème Central-Limite de De Moivre-Laplace (théorème V.2.5, page 155) montre que lorsque les variables X_n sont des variables de Bernoulli, la convergence a lieu en loi vers une loi gaussienne standard : $\frac{\sqrt{n}}{\sigma}(\frac{1}{n}S_n - m) \implies \mathcal{N}(0, 1)$. Le théorème Central-Limite dont l'énoncé suit montre que c'est encore le cas la situation générale. La chose remarquable est que **la limite en loi de $\frac{\sqrt{n}}{\sigma}(\frac{1}{n}S_n - m)$ ne dépend pas de la loi des variables X_n .**

Théorème V.4.2 ◀▶ (Th. Central-Limite en dim. 1) Soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité. Soit $X_n : \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, une suite de v.a. \mathcal{F} -mesurables. On suppose les variables indépendantes et de même loi. De plus on suppose qu'elles ont toutes un moment d'ordre 2. On pose

$$S_n := X_1 + \dots + X_n, \quad m := \mathbf{E}[X_n] \quad \text{et} \quad \sigma^2 = \text{var}(X_n).$$

Alors, on a la convergence en loi suivante,

$$\frac{\sqrt{n}}{\sigma}(\frac{1}{n}S_n - m) \implies \mathcal{N}(0, 1).$$

Par le théorème V.2.4 (page 153) reformulant la convergence en loi avec la fonction de répartition, la convergence précédente est équivalente à

$$\forall x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy. \quad (\text{V.18})$$

Preuve : on utilise le théorème de P. Lévy (théorème V.4.1) le développement de Taylor de la fonction caractéristique donné dans la proposition suivante (l'énoncé est en dimension n car il est utilisé ainsi plus loin).

Proposition V.4.3 Soit $\mathbf{X} = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$, un vecteur aléatoire \mathcal{F} -mesurable admettant un moment d'ordre 2, c'est-à-dire que $\mathbf{E}[\|\mathbf{X}\|^2] < \infty$. Alors, sa fonction caractéristique est C^2 est on a le développement de Taylor suivant au voisinage de 0 :

$$\forall \mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n, \quad \widehat{\mu}_{\mathbf{X}}(\mathbf{u}) = 1 + i \sum_{1 \leq k \leq n} \mathbf{E}[X_k] u_k - \frac{1}{2} \sum_{1 \leq j, k \leq n} \mathbf{E}[X_j X_k] u_j u_k + o(\|\mathbf{u}\|^2).$$

Preuve : les X_k ont des moments d'ordre 2. Ce sont des v.a. intégrables. Comme déjà mentionné, $X_j X_k$ sont aussi intégrables. Or on observe que

$$\frac{\partial}{\partial u_k} e^{i\langle \mathbf{u}, \mathbf{X} \rangle} = i u_k X_k e^{i\langle \mathbf{u}, \mathbf{X} \rangle} \quad \text{et} \quad \frac{\partial^2}{\partial u_j \partial u_k} e^{i\langle \mathbf{u}, \mathbf{X} \rangle} = -u_j u_k X_j X_k e^{i\langle \mathbf{u}, \mathbf{X} \rangle}.$$

Le théorème de dérivation sous l'intégrale (appliqué deux fois) montre alors que

$$\frac{\partial \widehat{\mu}_{\mathbf{X}}}{\partial u_k}(\mathbf{u}) = i u_k \mathbf{E}[X_k e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] \quad \text{et} \quad \frac{\partial^2 \widehat{\mu}_{\mathbf{X}}}{\partial u_j \partial u_k}(\mathbf{u}) = -u_j u_k \mathbf{E}[X_j X_k e^{i\langle \mathbf{u}, \mathbf{X} \rangle}],$$

ce qui implique le résultat voulu. ■

Fin de la preuve du théorème Central-Limite V.4.2 On fixe $u \in \mathbb{R}$. On pose

$$Z_n = \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - m \right) = \frac{X_1 - m}{\sigma\sqrt{n}} + \dots + \frac{X_n - m}{\sigma\sqrt{n}}.$$

Puisque les variables ont la même loi et sont indépendantes, on a

$$\begin{aligned} \mathbf{E}[e^{iuZ_n}] &= \mathbf{E}[e^{iu(X_1-m)/\sigma\sqrt{n}} \dots e^{iu(X_n-m)/\sigma\sqrt{n}}] \\ &= \mathbf{E}[e^{iu(X_1-m)/\sigma\sqrt{n}}] \dots \mathbf{E}[e^{iu(X_n-m)/\sigma\sqrt{n}}] \\ &= (\mathbf{E}[e^{iu(X_1-m)/\sigma\sqrt{n}}])^n \end{aligned}$$

Par la proposition V.4.3 qui précède, on a, pour $u \in \mathbb{R}$ fixé et lorsque n tend vers l'infini :

$$\begin{aligned} g_n(u) &:= \mathbf{E}[e^{iu(X_1-m)/\sigma\sqrt{n}}] \\ &= e^{-ium/\sigma\sqrt{n}} \widehat{\mu}_{X_1}(u/\sigma\sqrt{n}) \\ &= e^{-ium/\sigma\sqrt{n}} \left(1 + \frac{ium}{\sigma\sqrt{n}} - \frac{u^2 \mathbf{E}[X_1^2]}{2\sigma^2 n} + o(n^{-1}) \right) \\ &= \left(1 - \frac{ium}{\sigma\sqrt{n}} - \frac{u^2 m^2}{2\sigma^2 n} + o(n^{-1}) \right) \left(1 + \frac{ium}{\sigma\sqrt{n}} - \frac{u^2 \mathbf{E}[X_1^2]}{2\sigma^2 n} + o(n^{-1}) \right) \\ &= 1 - \frac{u^2 (\mathbf{E}[X_1^2] - m^2)}{2\sigma^2 n} + o(n^{-1}) \\ &= 1 - \frac{u^2}{2n} + o(n^{-1}). \end{aligned}$$

On a donc (en utilisant la détermination habituelle du logarithme complexe)

$$\begin{aligned} \mathbf{E}[e^{iuZ_n}] &= g_n(u)^n = \left(1 - \frac{u^2}{2n} + o(n^{-1}) \right)^n \\ &= \exp \left(n \log \left(1 - \frac{u^2}{2n} + o(n^{-1}) \right) \right) \\ &= \exp \left(-u^2/2 + o(1) \right). \end{aligned}$$

Si on note Z une v.a. de loi $\mathcal{N}(0, 1)$, on a montré pour tout $u \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbf{E}[e^{iuZ_n}] = \exp(-u^2/2) = \mathbf{E}[e^{iuZ}]$$

et le théorème de Paul Lévy (théorème V.4.1) entraîne que $Z_n \Rightarrow Z$, qui est le résultat voulu. ■

◀▶ Intervalles de confiance asymptotiques Le théorème V.2.4 permet une reformulation concrète du théorème central-limite en dimension 1 : soit $(\Omega, \mathcal{F}, \mathbf{P})$, un espace de probabilité ; soient $(X_n)_{n \in \mathbb{N}}$, une suite de v.a. réelles indépendantes et de même loi ; on suppose qu'elles admettent un moment d'ordre 2 et on pose

$$m := \mathbf{E}[X_1] \quad \text{et} \quad \sigma^2 = \mathbf{var}(X_1).$$

Alors, le théorème central-limite s'énonce comme suit :

$$\forall x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy. \quad (\text{V.19})$$

Supposons que l'on cherche à estimer m à partir de l'échantillon (X_1, \dots, X_n) , qui sont des données obtenues par une expérience. La loi des grands nombres montre que lorsque n est grand la moyenne empirique $M_n = (X_1 + \dots + X_n)/n$ est proche de m . Le théorème central-limite permet d'avoir une idée de l'ordre de grandeur de l'erreur que l'on commet lorsqu'on assimile M_n à m : on a en effet

$$\forall x \in \mathbb{R}_+, \quad \lim_{n \rightarrow \infty} \mathbf{P} \left(m \in [M_n - n^{-1/2} \sigma x, M_n + n^{-1/2} \sigma x] \right) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-y^2/2} dy .$$

Si on choisit $x = 1,96$, les tables numériques sur la loi normale nous disent que

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(m \in [M_n - 1,96n^{-1/2} \sigma, M_n + 1,96n^{-1/2} \sigma] \right) = \frac{1}{\sqrt{2\pi}} \int_{-1,96}^{1,96} e^{-y^2/2} dy = 95\% .$$

En assimilant, pour n grand, les probabilité à leur limite (ce qui n'est pas rigoureux) on peut affirmer le fait suivant.

$$\text{Avec une probabilité de } 95\%, |M_n - m| \leq 1,96n^{-1/2} \sigma .$$

Si on veut une précision de 0,01, sur l'estimation, il faut donc que $1,96n^{-1/2} \sigma \approx 0,01$, c'est-à-dire qu'il faut un échantillon assez important de l'ordre de $(1,96\sigma/(0,01))^2$. Même si ce résultat n'est pas tout à fait rigoureux, il se dégage en tout cas, le principe statistique suivant : l'échantillon est en gros proportionnel à l'inverse du carré de la précision.

Illustrons ce fait avec un problème de sondage : supposons qu'une élection ait lieu ; deux candidats sont en lice (candidat A et candidat B) ; des sondages de sortie des urnes sont effectués : n personnes ont été interrogées ; on note $X_k = 1$ si la $k^{\text{ième}}$ personne interrogée a voté pour A et $X_k = 0$, sinon. Donc $M_n = (X_1 + \dots + X_n)/n$ représente la proportion des n personnes interrogées ayant voté pour le candidat A.

En première approximation, on suppose que les personnes interrogées sont "tirées" aléatoirement parmi tous les électeurs (ce qui est raisonnable pour des sondages de sorties des urnes s'effectuant sur une journée) : on peut modéliser les v.a. $(X_k)_{1 \leq k \leq n}$ comme indépendantes et de même loi donnée par $\mathbf{P}(X_k = 1) = p$ et $\mathbf{P}(X_k = 0) = 1 - p$, où p est la proportion de tous les électeurs ayant voté pour le candidat A. Ce que les instituts de sondages vendent aux journaux et télévisions est leur estimation de p . Ici $\sigma = p(1-p) \leq 1/4$. Pour une prédiction sûre à 95%, on peut affirmer le fait suivant.

$$\text{Avec une probabilité de } 95\%, |M_n - p| \leq 1,96n^{-1/2} / 2 .$$

En général p est proche de 1/2 et les scores aux second tour des présidentielles peuvent être serrés, une précision de l'ordre de 0,5% est raisonnable pour être quasi-certain d'annoncer le vainqueur. On voit qu'il faut interroger de l'ordre de $n = (1,96/(2 \times 0,005))^2 \approx (200)^2 = 40\,000$ personnes, ce qui est assez important et coûteux.

Au delà de l'approximation consistant à assimiler, pour n grand, le membre de gauche de (V.19) à sa limite, il faut faire face à un problème pratique : l'erreur que l'on commet en assimilant M_n à m dépend de σ , qui est en principe tout aussi inconnu que m . Pour lever ce problème, on estime σ^2 en posant

$$\Sigma_n^2 = \frac{1}{n-1} \sum_{1 \leq k \leq n} (X_k - M_n)^2 .$$

On montre facilement que $\mathbf{E}[\Sigma_n^2] = \sigma^2$, en développant les carrés et en utilisant l'indépendance supposée des X_k . On montre le lemme suivant.

Lemme V.4.4 $\lim_{n \rightarrow \infty} \Sigma_n = \sigma$ presque sûrement.

Preuve : en développant les carrés, on obtient

$$\Sigma_n^2 = \frac{1}{n-1} \sum_{1 \leq k \leq n} X_k^2 - \frac{2n}{n-1} M_n^2 + \frac{n}{n-1} M_n^2 = \frac{1}{n-1} \sum_{1 \leq k \leq n} X_k^2 - \frac{n}{n-1} M_n^2 .$$

Par la loi des grands nombres, on a donc

$$\mathbf{P}\text{-p.s.} \quad \lim_{n \rightarrow \infty} \Sigma_n = \mathbf{E}[X_1^2] - (\mathbf{E}[X_1])^2 = \sigma^2 ,$$

ce qui permet de conclure. ■

Le lemme V.1.7 de Slutsky page 151 implique que

$$\left(\frac{\sqrt{n}}{\sigma} (M_n - m) ; \Sigma_n / \sigma \right) \Longrightarrow (X, 1) ,$$

où $X \sim \mathcal{N}(0, 1)$. On remarque ensuite que $\Phi(x, y) = x/y$ est p.s. continue en (X, σ) . Le théorème V.1.4 page 150 implique que

$$\frac{\sqrt{n}}{\Sigma_n} (M_n - m) \Longrightarrow X ,$$

c'est-à-dire qu'on obtient les intervalles de confiance asymptotiques suivants

$$\forall x \in \mathbb{R}_+, \quad \lim_{n \rightarrow \infty} \mathbf{P} \left(m \in [M_n - n^{-1/2} \Sigma_n x, M_n + n^{-1/2} \Sigma_n x] \right) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-y^2/2} dy .$$

Bien entendu, il est nécessaire de connaître la vitesse de convergence de la limite précédente pour évaluer l'erreur que l'on commet lorsqu'on assimile $\mathbf{P} (m \in [M_n - n^{-1/2} \Sigma_n x, M_n + n^{-1/2} \Sigma_n x])$ à sa limite : ce problème est délicat en général. Le cas des variable de Bernoulli a été examiné cependant avec suffisamment précision et des expériences numériques permettent d'établir des bornes pour ces erreurs en ce qui concerne les lois usuelles. Nous citons ici, sans preuve le théorème de Berry-Esseen qui répond précisément à cette question.

Théorème V.4.5 (Théorème de Berry-Esseen) $(X_n)_{n \geq 1}$, une suite de v.a. réelles \mathcal{F} -mesurables indépendantes et de même loi. On suppose qu'elles admettent un moment d'ordre 3 : $\mathbf{E}[|X_1|^3] < \infty$. On pose $m := \mathbf{E}[X_1]$, $\sigma^2 := \mathbf{var}(X_1)$. Alors pour tout $n \geq 1$, on a

$$\sup_{x \in \mathbb{R}} \left| \mathbf{P} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \leq x \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \right| \leq \frac{3\mathbf{E}[|X_1 - m|^3]}{\sigma^3 \sqrt{n}} .$$

La preuve de ce résultat est technique, bien qu'elle n'utilise que des notions présentes dans ce cours. On observe que ce résultat est vrai pour tout n mais puisque $\mathbf{E}[|X_1 - m|^3] \geq \sigma^2$, par Jensen, il n'est significatif que pour $n \geq 10$. De plus cet écart entre la fonction de répartition de la moyenne empirique et la fonction de répartition Gaussienne est en général d'ordre $n^{-1/2}$: pour cela, voir le cas des variables de Bernoulli, comme le montre une rapide inspection de la preuve du théorème de De Moivre-Laplace.

V.4.c Vecteurs gaussiens et TCL sur \mathbb{R}^d .

On cherche à généraliser le théorème Central-Limite aux vecteurs. Pour cela on introduit d'abord les vecteur gaussiens qui généralisent les variables gaussiennes réelles.

Rappel sur les variables gaussiennes réelles.

Définition V.4.1 ◀▶ (Lois Gaussiennes (normales) en dimension 1) Soient $m \in \mathbb{R}$ et $\sigma \in]0, \infty[$. Soit $X : \Omega \rightarrow \mathbb{R}$, une variable \mathcal{F} -mesurable. Cette variable est dite (de loi) Gaussienne, ou encore normale, de moyenne m et de variance σ^2 si sa loi admet la densité $g_{m,\sigma}$ donnée par

$$g_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in \mathbb{R}. \quad (\text{V.20})$$

On note cela $X \sim \mathcal{N}(m, \sigma^2)$. Si $m = 0$, X est dite centrée et une variable Gaussienne $\mathcal{N}(0, 1)$ est appelée Gaussienne standard.

Par commodité, on dit également que X est une loi Gaussienne de moyenne m et de variance nulle si X est presque sûrement égal à m . On continue de noter cela $X \sim \mathcal{N}(m, 0)$. On parle alors de variable Gaussienne dégénérée. \square

On vérifie facilement que si $X \sim \mathcal{N}(m, \sigma^2)$,

$$\mathbf{E}[X] = \int_{\mathbb{R}} x g_{m,\sigma}(x) dx = m \quad \text{et} \quad \mathbf{var}(X) = \int_{\mathbb{R}} (x - m)^2 g_{m,\sigma}(x) dx = \sigma^2.$$

Autrement dit la loi de X est entièrement déterminée par sa moyenne et sa variance. On rappelle que $\widehat{g_{0,\sigma}}(u) = \exp(-\frac{1}{2}\sigma^2 u^2)$. Cela entraîne facilement que si $X \sim \mathcal{N}(m, \sigma)$,

$$\forall u \in \mathbb{R}, \quad \mathbf{E}[e^{iuX}] = \widehat{\mu}_X(u) = \widehat{g}_{m,\sigma}(u) = e^{imu - \frac{1}{2}\sigma^2 u^2}.$$

Il est facile de montrer que la fonction caractéristique de X s'étend à \mathbb{C} , c'est-à-dire que

$$\forall z \in \mathbb{C}, \quad \mathbf{E}[e^{zX}] = e^{mz + \frac{1}{2}\sigma^2 z^2}.$$

Le membre de droite est une fonction analytique sur \mathbb{C} dont le rayon de convergence est infini (c'est une fonction entière). Si on suppose pour simplifier que $X \sim \mathcal{N}(0, \sigma)$, on vérifie alors

$$\mathbf{E}[e^{zX}] = \sum_{n \geq 0} \frac{\mathbf{E}[X^n]}{n!} z^n = \sum_{n \geq 0} \frac{\sigma^{2n}}{2^n n!} z^{2n}.$$

Cela implique que si $X \sim \mathcal{N}(0, \sigma)$, pour tout $n \in \mathbb{N}$,

$$\mathbf{E}[X^{2n+1}] = 0 \quad \text{et} \quad \mathbf{E}[X^{2n}] = \frac{(2n)!}{n! 2^n} \sigma^{2n} = \sigma^{2n} (2n-1) \times (2n-3) \times \dots \times 5 \times 3 \times 1.$$

Lemme V.4.6 ◀▶ Soient $X_1 \sim \mathcal{N}(m_1, \sigma_1^2), \dots, X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ des variables Gaussiennes indépendantes. Soient $a_1, \dots, a_n \in \mathbb{R}$. Alors $a_1 X_1 + \dots + a_n X_n \sim \mathcal{N}(m, \sigma^2)$ où

$$m = a_1 m_1 + \dots + a_n m_n \quad \text{et} \quad \sigma^2 = (a_1 \sigma_1)^2 + \dots + (a_n \sigma_n)^2.$$

Preuve : on pose $X = a_1 X_1 + \dots + a_n X_n$ et on a pour tout $u \in \mathbb{R}$,

$$\mathbf{E}[e^{iuX}] = \mathbf{E}[e^{ia_1 u X_1}] \dots \mathbf{E}[e^{ia_n u X_n}] = e^{imu - \frac{1}{2}\sigma^2 u^2},$$

et l'injectivité de la transformée de Fourier permet de conclure. \blacksquare

Les vecteurs gaussiens.

Définition V.4.2 ◀▶ (Vecteurs Gaussiens) Soit $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$, un vecteur \mathcal{F} -mesurable. C'est un *vecteur Gaussien* si pour tout $\mathbf{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$, $\langle \mathbf{u}, \mathbf{X} \rangle = u_1 X_1 + \dots + u_d X_d$ est une variable Gaussienne réelle (possiblement dégénérée). \square

Le lemme suivant découle immédiatement de la définition.

Lemme V.4.7 Soit $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$, un vecteur Gaussien. Soit M , une matrice réelle carrée de taille $d \times d$. Alors, le vecteur aléatoire $M\mathbf{X}$ est Gaussien.

Exemple V.4.1 Soient X_1, \dots, X_d , des v.a. réelles indépendantes suivant une loi gaussienne standard $\mathcal{N}(0, 1)$. Alors, $\mathbf{X} := (X_1, \dots, X_d)$ est un vecteur Gaussien. En effet, On fixe $\mathbf{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$ et $v \in \mathbb{R}$. L'indépendance implique que

$$\widehat{\mu}_{\langle \mathbf{u}, \mathbf{X} \rangle}(v) = \prod_{1 \leq k \leq d} \mathbf{E}[e^{iv u_k X_k}] = \exp\left(-\frac{1}{2}v^2(u_1^2 + \dots + u_d^2)\right) = \exp\left(-\frac{1}{2}v^2\|\mathbf{u}\|^2\right).$$

Cela montre que pour tout $\mathbf{u} \in \mathbb{R}^d$, $\langle \mathbf{u}, \mathbf{X} \rangle$ est une variable Gaussienne $\mathcal{N}(0, \|\mathbf{u}\|^2)$ par injectivité de la fonction caractéristique. \square

On rappelle les définitions du vecteur moyenne et de la matrice de covariance, ainsi que des formules élémentaires de la proposition II.9.10 page 82 qui entraînent immédiatement le lemme suivant.

Lemme V.4.8 ◀▶ Soit $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$, un vecteur Gaussien \mathcal{F} -mesurable. On note \mathbf{v} , le vecteur moyenne de \mathbf{X} et Γ , sa matrice de covariance. Alors pour tout $\mathbf{u} \in \mathbb{R}^d$,

$$\langle \mathbf{u}, \mathbf{X} \rangle \text{ est une v.a. réelle Gaussienne de moyenne } \langle \mathbf{u}, \mathbf{v} \rangle \text{ et de variance } \langle \mathbf{u}, \Gamma \mathbf{u} \rangle.$$

et donc

$$\widehat{\mu}_{\mathbf{X}}(\mathbf{u}) = \mathbf{E}[e^{i\langle \mathbf{u}, \mathbf{X} \rangle}] = \widehat{\mu}_{\langle \mathbf{u}, \mathbf{X} \rangle}(1) = \exp\left(i\langle \mathbf{u}, \mathbf{v} \rangle - \frac{1}{2}\langle \mathbf{u}, \Gamma \mathbf{u} \rangle\right). \quad (\text{V.21})$$

Définition V.4.3 ◀▶ Le lemme précédent montre que la loi d'un vecteur Gaussien \mathbf{X} est entièrement caractérisée par son vecteur moyenne \mathbf{v} et sa matrice de covariance Γ . On note le fait que \mathbf{X} est un vecteur Gaussien de vecteur moyen \mathbf{v} et de matrice de covariance Γ par $\mathbf{X} \sim \mathcal{N}(\mathbf{v}, \Gamma)$. \square

Soit $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$, un vecteur Gaussien \mathcal{F} -mesurable de vecteur moyen \mathbf{v} et de matrice de covariance Γ . On note r le rang de Γ et $\lambda_1, \dots, \lambda_r$ ses valeurs propres non-nulles. Comme Γ est symétrique positive, $\lambda_1, \dots, \lambda_r > 0$. On note Δ et $\widetilde{\Delta}$ les matrices $d \times d$, diagonales données par

$$\Delta = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \sqrt{\lambda_r} & \ddots & & \vdots \\ \vdots & & \ddots & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 0 & 0 \end{pmatrix} \quad \text{et} \quad \widetilde{\Delta} = \begin{pmatrix} 1/\sqrt{\lambda_1} & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & 1/\sqrt{\lambda_r} & \ddots & & \vdots \\ \vdots & & \ddots & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & 0 \end{pmatrix}$$

$\Delta \widetilde{\Delta}$ est donc la matrice \mathbf{I}_r qui est diagonale et dont les r premiers coefficients diagonaux sont égaux à 1 et les $d-r$ suivants sont nuls. Comme Γ est symétrique, il existe une matrice orthogonale O telle que $O^* \Gamma O = \Delta^2$, où O^* désigne la transposée de O .

Proposition V.4.9 On reprend les notations ci-dessus. On pose $\mathbf{Y} := \tilde{\Delta}O^*(\mathbf{X} - \mathbf{v})$. Alors

$$\mathbf{Y} = (Y_1, \dots, Y_r, 0, \dots, 0),$$

où Y_1, \dots, Y_r sont des v.a. réelles indépendantes de loi $\mathcal{N}(0, 1)$, qui sont appelées les composantes indépendantes de \mathbf{X} . De plus $\mathbf{X} = \mathbf{v} + O\Delta\mathbf{Y}$.

Preuve : \mathbf{Y} est un vecteur Gaussien de vecteur moyen nul. Par la proposition II.9.10 (ii), page 82, sa matrice de covariance Γ' est donnée par

$$\Gamma' = \tilde{\Delta}O^*\Gamma(\tilde{\Delta}O^*)^* = \tilde{\Delta}O^*\Gamma O\tilde{\Delta} = \tilde{\Delta}\Delta^2\tilde{\Delta} = \mathbf{I}_r.$$

On a donc $\mathbf{Y} \sim \mathcal{N}(0, \mathbf{I}_r)$, ce qui montre la proposition. ■

Théorème Central-Limite sur \mathbb{R}^d . Le théorème suivant étend en dimension quelconque le théorème Central-Limite V.4.2, page 165.

Théorème V.4.10 ◀▶ (Théorème central-limite) Soient $\mathbf{X}_n = (X_{n,1}, \dots, X_{n,d}) : \Omega \rightarrow \mathbb{R}^d$, $n \in \mathbb{N}$, des vecteurs aléatoires. On les suppose indépendants et de même loi. On suppose également qu'ils admettent un moment d'ordre deux, c'est-à-dire que $\mathbf{E}[\|\mathbf{X}_n\|^2] < \infty$. Ces vecteurs ont donc même vecteur moyenne et même matrice de covariance que l'on note comme suit :

$$\mathbf{v} := (\mathbf{E}[X_{n,1}], \dots, \mathbf{E}[X_{n,d}]) \quad \text{et} \quad \Gamma := (\text{cov}(X_{n,i}, X_{n,j}))_{1 \leq i, j \leq d}.$$

Alors,

$$\sqrt{n} \left(\frac{1}{n} (\mathbf{X}_1 + \dots + \mathbf{X}_n) - \mathbf{v} \right) \implies \mathcal{N}(0, \Gamma),$$

où on rappelle que $\mathcal{N}(0, \Gamma)$ désigne la loi Gaussienne sur \mathbb{R}^d de vecteur moyen nul et de matrice de covariance Γ .

Preuve : on fixe $\mathbf{u} \in \mathbb{R}^d$. Pour tout $n \in \mathbb{N}$, on pose $Y_n = \langle \mathbf{u}, \mathbf{X}_n \rangle - \langle \mathbf{u}, \mathbf{v} \rangle$. Les v.a. réelles $(Y_n)_{n \in \mathbb{N}}$ sont donc indépendantes et de même loi ; elles admettent un moment d'ordre 2 et on vérifie que

$$\mathbf{E}[Y_n] = 0 \quad \text{et} \quad \sigma^2 := \text{var}(Y_n) = \mathbf{E}[Y_n^2] = \langle \mathbf{u}, \Gamma \mathbf{u} \rangle.$$

On note ensuite $\mathbf{V}_n = \sqrt{n} \left(\frac{1}{n} (\mathbf{X}_1 + \dots + \mathbf{X}_n) - \mathbf{v} \right)$. On a donc $\langle \mathbf{u}, \mathbf{V}_n \rangle = \frac{1}{\sqrt{n}} (Y_1 + \dots + Y_n)$. On vérifie immédiatement que

$$\hat{\mu}_{\mathbf{V}_n}(\mathbf{u}) = \prod_{1 \leq k \leq n} \mathbf{E}[e^{in^{-1/2}(Y_k)}] = (\hat{\mu}_{Y_1}(n^{-1/2}))^n,$$

par la proposition III.2.17 page 109. Or la proposition V.4.3 page 165 en dimension 1 implique que

$$\hat{\mu}_{Y_1}(n^{-1/2}) = 1 - \frac{\sigma^2}{2n} + o(n^{-1}),$$

ce qui implique donc que

$$\lim_{n \rightarrow \infty} \hat{\mu}_{\mathbf{V}_n}(\mathbf{u}) = \exp\left(-\frac{1}{2}\sigma^2\right) = \exp\left(-\frac{1}{2}\langle \mathbf{u}, \Gamma \mathbf{u} \rangle\right),$$

et on conclut grâce au théorème V.4.1 de P. Lévy, page 162. ■

En application du théorème central-limite, donnons une caractérisation des lois Gaussiennes.

Théorème V.4.11 Soit $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$, un vecteur aléatoire \mathcal{F} -mesurable. On suppose que $\mathbf{E}[\|\mathbf{X}\|^2] < \infty$, que son vecteur moyenne est nul : $\mathbf{E}[\mathbf{X}] = 0$. On note Γ la matrice de covariance de \mathbf{X} . Soient $\mathbf{X}', \mathbf{X}'' : \Omega \rightarrow \mathbb{R}^d$, deux vecteurs aléatoires indépendants de même loi que \mathbf{X} . On suppose que

$$\mathbf{X} \stackrel{(loi)}{=} \frac{1}{\sqrt{2}}(\mathbf{X}' + \mathbf{X}'').$$

Alors $\mathbf{X} \sim \mathcal{N}(0, \Gamma)$.

Preuve : on se donne une suite $(\mathbf{X}_n)_{n \geq 1}$ de vecteurs aléatoires \mathcal{F} -mesurables indépendants et de même loi que \mathbf{X} . En itérant l'identité en loi de l'énoncé, on montre facilement

$$\mathbf{X} \stackrel{(loi)}{=} \frac{1}{\sqrt{2^n}}(\mathbf{X}_1 + \dots + \mathbf{X}_{2^n}).$$

Or le membre de droite converge en loi vers la loi normale $\mathcal{N}(0, \Gamma)$, par le théorème central-limite, ce qui permet de conclure. ■

V.5 Exercices

Exercice V.5.1 a. Construire une suite de variables aléatoires intégrables $(X_n)_{n \geq 1}$ et une variable aléatoire intégrable X telles qu'on ait

$$X_n \xrightarrow[n \rightarrow \infty]{loi} X \quad \text{et} \quad \lim_{n \rightarrow \infty} \mathbf{E}[X_n] \neq \mathbf{E}[X].$$

b. Montrer que si une suite de variables aléatoires $(X_n)_{n \geq 1}$ est telle qu'il existe une constante C telle qu'on ait $|X_n| \leq C$ presque sûrement pour tout $n \geq 1$, et si la suite $(X_n)_{n \geq 1}$ converge en loi vers X , alors $|X| \leq C$ presque sûrement et $\lim_{n \rightarrow \infty} \mathbf{E}[X_n] = \mathbf{E}[X]$.

Exercice V.5.2 Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi de Cauchy standard. Énoncer et démontrer un résultat pour une suite de variables aléatoires de la forme $\frac{X_1 + \dots + X_n}{n^\alpha}$. On rappelle que $\hat{\mu}_{X_1}(t) = e^{-|t|}$.

Exercice V.5.3 Soit X une variable aléatoire réelle. Déterminer à quelle condition sur X la fonction caractéristique de X ne prend que des valeurs réelles.

Exercice V.5.4 Soit $\phi : \mathbb{R} \rightarrow \mathbb{C}$ la fonction caractéristique d'une variable aléatoire. Montrer que les fonctions

$$t \mapsto \overline{\phi(t)}, \quad t \mapsto |\phi(t)|^2, \quad t \mapsto \operatorname{Re}(\phi(t))$$

sont des fonctions caractéristiques. La fonction $t \mapsto \operatorname{Im}(\phi(t))$ est-elle une fonction caractéristique ?

Exercice V.5.5 Soient X et Y deux variables aléatoires qui admettent chacune un moment d'ordre 2.

a. On suppose que pour tout t réel, $\hat{\mu}_{X+Y}(t) = \hat{\mu}_X(t)\hat{\mu}_Y(t)$. Montrer que la covariance de X et Y est nulle.

b. Préciser les liens logiques entre les assertions suivantes :

- (i) X et Y sont indépendantes,
- (ii) $\hat{\mu}_{X+Y} = \hat{\mu}_X \hat{\mu}_Y$,
- (iii) $\operatorname{cov}(X, Y) = 0$.

Exercice V.5.6 Montrer que si une suite de variables aléatoires converge en loi et si chaque terme de la suite a une loi exponentielle, alors la loi limite est exponentielle ou la masse de Dirac en 0.

Exercice V.5.7 Montrer que si une suite de variables aléatoires converge en loi et si chaque terme de la suite a une loi constante ou gaussienne, alors la loi limite est constante ou gaussienne.

Exercice V.5.8 a. Soit Z une variable aléatoire qui admet un moment d'ordre 2. Soit $a > 0$ un réel. Montrer que

$$\mathbf{E}[|Z| - \min(|Z|, a)] \leq \frac{\mathbf{E}[Z^2]}{a}.$$

b. Soit $(X_n)_{n \geq 1}$ une suite de v.a.i.i.d. qui admettent un moment d'ordre 2 et qui sont centrées. Montrer que

$$\mathbf{E} \left[\left| \frac{X_1 + \dots + X_n}{\sqrt{n}} \right| \right] \xrightarrow{n \rightarrow \infty} \sqrt{\frac{2}{\pi}}.$$

En déduire la limite lorsque n tend vers l'infini de

$$\mathbf{E} \left[\frac{X_1 + \dots + X_n}{\sqrt{n}} \mathbf{1}_{\frac{X_1 + \dots + X_n}{\sqrt{n}} \geq 0} \right].$$

c. Appliquer ce résultat à une suite de variables aléatoires i.i.d. de loi égale à celle de $Y - 1$, où Y a la loi de Poisson de paramètre 1.

Exercice V.5.9 Un écrivain qui commet en moyenne une faute d'orthographe toutes les 10 pages vient d'écrire un roman de 400 pages. Quelle est la probabilité qu'il ait commis plus de 50 fautes ?

Exercice V.5.10 La somme des résultats de 10000 lancers d'un même dé est 35487. Pensez-vous que ce dé soit truqué ?

Exercice V.5.11 La loi de Benford est une loi semi-empirique qui affirme que dans les nombres que l'on rencontre dans la vie courante, le premier chiffre significatif n'est pas équiréparti dans l'ensemble $\{1, \dots, 9\}$, mais qu'il est distribué selon la loi suivante :

$$p_i = \mathbf{P}(\text{premier chiffre significatif égal à } i) = \log_{10} \frac{i+1}{i},$$

où \log_{10} est le logarithme décimal, défini par $\log_{10}(x) = \frac{\log(x)}{\log(10)}$. En particulier,

$$p_5 \simeq 7,9\% \text{ et } p_6 \simeq 6,7\%.$$

Au début des années 1990, un économiste étasunien a remarqué que lorsque les gens falsifiaient les comptes d'une société, ils avaient tendance à utiliser trop de 5 et de 6 comme premiers chiffres significatifs, c'est-à-dire plus que n'en prévoit la loi de Benford, qui par ailleurs est bien vérifiée pour ce genre de données numériques.

Mettez au point une méthode qui tire parti de cette observation pour aider à détecter des fraudes.

Exercice V.5.12 a. Durant l'année qui vient de s'écouler, 10000 enfants sont nés dans une certaine ville. Parmi ces enfants, 5242 sont des filles et 4758 sont des garçons. Vous paraît-il raisonnable de soutenir l'hypothèse qu'un enfant qui naît dans cette ville a exactement autant de chances d'être un garçon que d'être une fille ?

b. Durant l'année qui vient de s'écouler, dans un pays donné, le quotient du nombre de filles qui sont nées par le nombre de garçons qui sont nés est de 1,2 (on dit parfois que le *sex ratio* de ce pays est de 120%). Vous paraît-il raisonnable de soutenir l'hypothèse qu'un enfant qui naît dans ce pays a exactement autant de chances d'être un garçon que d'être une fille ?

Exercice V.5.13 Il y a actuellement dans le monde environ 400 réacteurs nucléaires civils en fonctionnement. L'âge moyen de ces réacteurs est d'environ 25 ans. Depuis le début de l'exploitation de l'énergie nucléaire, on recense au moins 6 accidents nucléaires graves.

On fait l'hypothèse qu'il existe un nombre $p \in [0, 1]$ tel que pour chaque réacteur, chaque année, un accident grave survienne avec probabilité p . La valeur officielle de p , donnée par Institut de Radioprotection et de Sûreté Nucléaire (IRSN) est $p = 10^{-5}$.

En supposant que $p = 10^{-5}$, calculer les bornes d'un intervalle qui devrait contenir le nombre d'accidents recensés avec une probabilité de 90%. Aurait-on pu obtenir ce résultat d'une manière directe ? Quelle valeur de p estimeriez-vous plus correcte ?

Exercice V.5.14 On procède à une élection dans une ville de 100 000 habitants. Les habitants ont le choix entre A et B. Lors du dépouillement, on commet des erreurs : cela arrive pour chaque bulletin, indépendamment des autres, avec une probabilité de $\frac{1}{1000}$. Combien de voix doit-on avoir compté au minimum pour A à l'issue du dépouillement pour pouvoir affirmer avec moins de 1% de chances de se tromper que A l'a emporté ?

Exercice V.5.15 Calculer $\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$.

Exercice V.5.16 Soient $f : \mathbb{R} \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions continues et bornées.

a. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Énoncer, en s'assurant que leurs hypothèses sont vérifiées, la loi forte des grands nombres et le théorème central limite appliqués à la suite $(f(X_n))_{n \geq 1}$.

b. On suppose que la fonction f est impaire. Déterminer, si elle existe, la limite

$$\lim_{n \rightarrow \infty} \frac{1}{(\sqrt{2\pi})^n} \int_{\mathbb{R}^n} g \left(\frac{1}{n} \sum_{k=1}^n f(t_k) \right) e^{-\frac{1}{2}(t_1^2 + \dots + t_n^2)} dt_1 \dots dt_n.$$

c. On suppose de plus que $\int_{-1}^1 f(t)^2 dt = 2$. Déterminer, si elle existe, la limite

$$\lim_{n \rightarrow \infty} \frac{1}{2^n} \int_{[-1,1]^n} g \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n f(t_k) \right) dt_1 \dots dt_n.$$

Annexe A

Table de la loi gaussienne standard.

La table suivante donne les valeurs de la fonction de répartition $\Phi(z)$ de la loi normale standard $\mathcal{N}(0, 1)$, pour $0 \leq z \leq 3.5$. Rappelons que

$$\Phi(z) := \int_{-\infty}^z \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx.$$

Les valeurs de $\Phi(z)$ pour $z < 0$ peuvent être retrouvées grâce à la formule $\Phi(z) = 1 - \Phi(-z)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Index

- $(\Omega, \mathcal{F}, \mathbf{P})$: espace de probabilité 12
 $\mathcal{L}_{\mathbb{R}^d}(\Omega, \mathcal{F})$ 125
 $\hat{\mu}_{\mathbf{X}}$: fonction caractéristique de \mathbf{X} 85
 $\text{cov}(X, Y)$: covariance de X et Y 81
 $\text{var}(X)$: variance de X 57
 F_X : fonction de répartition de X 69
 F_μ : fonction de répartition de μ 69
 L_X : trans. de Laplace 92
 $X_n \implies X$: conv. en loi 147
 $\Gamma_{\mathbf{X}}$: mat. de cov. de \mathbf{X} 82
 $\mathbf{E}[X]$: espérance de X 45
 \mathbf{P} -p.s. : \mathbf{P} -p.p 45
 $\mathbf{P}(A; B)$ 12
 $\mathbf{P}(B|A)$: proba de B sachant A 18
 $\mathcal{B}(E)$: Boréliens de E 4
 $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$: Tribu produit 74
 $\mathcal{N}(m, \sigma^2)$: loi normale 64, 169
 $\mathcal{P}(E)$: tous les sous-ensembles de E 2
 ℓ : mesure de Lebesgue sur \mathbb{R} 10
 ℓ_n : mesure de Lebesgue en dim. n 10, 77
 $L^p(\Omega, \mathcal{F}, \mathbf{P})$ 132
 $\lambda(\mathcal{R})$: classe monotone engendrée 6
 $\liminf_{n \rightarrow \infty} A_n$ 112
 $\limsup_{n \rightarrow \infty} A_n$ 112
 $\mu * \nu$ 107
 μ_X : loi de X 52
 $\sigma(X)$: écart-type de X 57
 $\sigma(\mathcal{R})$: tribu engendrée 3
 $\hat{\mu}$: trans. de Fourier de la mes. μ 85
 dF : mesure de Stieltjes de F 10
 g_σ : densité Gaussienne 86, 162
 $\mathbf{Rep}(\mathbb{R})$: espace des fonc. de répartition .., 152
 \mathbf{S}_+ : fonctions étagées positives 36
Additivité 14
Aiguille de Buffon 140
Alexandroff (théorème) 148
Approximation binomiale-Gaussienne, 155
Approximation Binomiale-Poisson 160
Approximation par des fonctions étagées....., 32
Asymptotique (tribu) 136
Atome 9, 66
Atomique (tribu) 36
Atomique : mesure purement 67
Bayes (fomule) 19
Bernoulli (loi) 49
Bernstein (polynôme) 134
Berry-Esseen (théorème) 168
Binomiale (approximation Gaussienne), 155
Binomiale (loi) 49
Binomiale-Poisson (approximation) 160
Boréliens 4
Borel (lemme) 113
Borel : lemme de Borel-Cantelli 112
Borel : loi des grands nombres 112
Buffon (aiguille) 140
Cantelli : lemme de Borel-Cantelli 112
Canular 148
Caractéristique (fonction) 85
Caractéristique : injectivité de la fonction, 88
Cauchy (densité) 89
Cauchy (loi) 139
Central-limite (théorème) 165, 171
Changement de variable abstrait 51
Changement de variable géométrique 79
Changement de variable : polaire 80
Classe monotone engendrée 6
Classe monotone : version fonctionnelle, 33
Classes monotones 2, 5
Collectionneur d'images 135
Conjugués (exposants) 55
Continuité en général 29
Convergence L^p 132
Convergence étroite 147
Convergence dominée 44

Convergence en loi	147	Fonction de répartition et conv. en loi	153
Convergence en loi et fonc. de répartition ..	153	Fonction génératrice	58
Convergence en probabilité	128	Fonction indicatrice	31
Convergence faible (des mesures)	147	Fonctions étagées	31
Convergence monotone	39	Fonctions continues en général	29
Convergence presque sûre	125	Fonctions mesurables	28
Convolution des mes. finies	107	Formule d'inversion de Fourier	88
Covariance (matrice)	81	Formule de Bayes	19
Covariance	81	Formule de la densité	62, 63, 66
Crible (formule)	15	Formule du crible	15
Croissance séquentielle	8, 14	Fourier (inversion)	88
Décroissance séquentielle	8, 14	Fourier : injectivité de la transformée	87
Dés non-transitifs	142	Fourier : trans. des mes. finies	85
De Moivres : th. de De Moivres-Laplace	155	Friteuses	115
Densité	61	Fubini (théorème positif)	75
Densité d'une variable aléatoire	63, 65	Fubini intégrable (théorème de)	75
Densité de Cauchy	89	Fubini-Lebesgue (théorème de)	75
Densité : formule de la	62, 63, 66	Fubini-Tonelli (théorème)	75
Densités Gaussiennes	86, 162	Génératrice (fonction)	58
Difféomorphisme	79	Géométrique (loi)	49
Différentiabilité	79	Gaussienne (approximation binomiale)	155
Diffuse (mesure)	9, 66	Gaussienne (densité)	86, 162
Discrète (tribu)	36	Gaussienne (loi, dim. 1)	64, 169
discrète : v.a.	46	Gaussienne standard	64
Distance en variation	157	Gaussiens (vecteurs)	170
Distance en variation, lois sur \mathbb{Z}	161	Hölder (inégalité)	55
Dynkin : th. d'unicité du prolongement	11	Homéomorphisme	78
Ecart-type	57	Identité de König	57
Ecrasés (piétons)	160	Inégalité de Hölder	55
Espérance d'une variable aléatoire	45	Inégalité de Jensen	54
Espace $L^p(\Omega, \mathcal{F}, \mathbf{P})$	132	Inégalité de Markov	56
Espace de probabilité	9, 12	Inégalité de Minkowski	55
Espace mesuré	7	Inégalité de Tchebychev	57
Etagées (fonctions)	31	Indépendance d'événements	20, 22, 97
Etroite (convergence)	147	Indépendance de classe d'événements	97
Événement	12	Indépendance de variables aléatoires	100
Existence de Lebesgue en dim. n	77	Indépendance par paquets	99
Exponentielle (loi)	64	Indépendance : construction de suite de v.a. ,	116
Exposants conjugués	55	Indicatrice (fonction)	31
Faible (convergence des mesures)	147	Injectivité de la fonction caractéristique	88
Fatou (lemme de)	41	Injectivité de la trans. de Fourier	87
Fermé d'une topologie	3	Injectivité de la trans. de Laplace	94
Fonction caractéristique	85	Intégrale à paramètre : continuité	83
Fonction de répartition	69	Intégrale à paramètre : dérivation	83

Intervalles de confiance	166	Méthode de Monte-Carlo	140
Interversion L^1	44	Marche aléatoire symétrique	116, 137
Interversion \sum/f positive	40	Markov (inégalité)	56
Inversion de Fourier L^1	88	Masse (d'une mesure)	9
Inversion globale	79	Matrice de covariance	81
Jensen (inégalité)	54	Mesurabilité des fonctions : simplification ..	28
König (identité)	57	Mesurabilité des v.a. discrètes	46
Kochen-Stone (lemme)	114	mesurabilité par rapport à une v.a.	35
Kolmogorov : loi du zéro-un	137	Mesure (de masse finie)	9
Lévy (théorème de continuité)	162	Mesure de Lebesgue en dim. n	10, 77
Lambda-système	5	Mesure de Lebesgue sur \mathbb{R}	10
Laplace (injectivité de la transformée)	94	Mesure de probabilité	9
Laplace (transformée)	92	Mesure de Stieltjes	10
Laplace : th. de De Moivre-Laplace	155	Mesure diffuse	9, 66
Lebesgue : mesure en dim. n	10, 77	Mesure image	51
Lebesgue : mesure sur \mathbb{R}	10	Mesure positive	7
Lemme de Borel	113	Mesure purement atomique	67
Lemme de Borel-Cantelli	112	Mesure sigma-finie	9
Lemme de Fatou	41	Minkowski (inégalité)	55
Lemme de Kochen-Stone	114	Moments d'une variable aléatoire	54
Lemme de Slutsky	151	Monte-Carlo	140
Lemme des réveils	105	Nombres normaux	119
Limite sup./inf. d'événements	112	Normal : nombre	119
Loi (convergence en)	147	Normale (loi, dim. 1)	64, 169
Loi	9	Normale standard	64
Loi binomiale	49	Oubli, propriété de v.a. exponentielles	71
Loi d'une variable aléatoire	52	Oubli, propriété de v.a. géométriques	50
Loi de Bernoulli	49	Ouverts d'une topologie	3
Loi de Cauchy	88, 139	p.s. : presque sûr	45
Loi de Poisson	50, 158	Pavés mesurables	74
Loi de Poisson dégénérée	50, 158	Pi-système	5
Loi des grands nombres (faible)	134	Piétons écrasés	160
Loi des grands nombres (forte)	138, 139	Poisson (loi dégénérée)	50, 158
Loi des grands nombres de Borel	112	Poisson (loi)	50, 158
Loi du tout ou rien de Kolmogorov	137	Poisson : approximation Binomiale-Poisson ...	160
Loi du zéro-un de Kolmogorov	137	Polaire : changement de variable	80
Loi exponentielle	64	Polynôme de Bernstein	134
Loi géométrique	49	Portemanteau (Jean-Pierre et théorème)	148
Loi Gaussienne (dim. 1)	64, 169	Pré-image : stabilité des tribus	27
Loi gaussienne standard	64	Presque partout	41
Loi normale (dim. 1)	64, 169	Presque sûr (événement)	13
Loi normale standard	64	Presque sûr : presque partout	45
Loi uniforme	53	Presque sûre (convergence)	125
Loi : convergence et fonc. de répartition ...	153		

Probabilité (convergence en)	128	Théorème de De Moivres-Laplace	155
Probabilité (espace de)	12	Théorème de Fubini intégrable	75
Probabilité	9	Théorème de Fubini positif	75
Probabilité conditionnelle	18	Théorème de Fubini-Lebesgue	75
Produit de convolution des mesures	107	Théorème de Fubini-Tonelli	75
Prolongement (unicité pour les mesures)	11	Théorème de Portemanteau	148
Propriété d'oubli (lois exponentielles)	71	Théorème de Radon-Nikodym (mes. complexes)	63
Propriété d'oubli (lois géométriques)	50	Théorème de renouvellement	139
Pseudo-inverse continu à droite	152	Théorème de transfert	51, 52
Pseudo-réciproque	72	Théorème de Weierstraß	134, 135
Purement atomique : mesure	67	Topologie	3
Répartition (fonction de)	69	Tout ou rien : loi de Kolmogorov	137
Répartition : fonctions et conv. en loi	153	Transfert (théorème de)	51, 52
Réveils (lemme)	105	Transformée de Fourier des mes. finies	85
Radon-Nikodym (mes. complexes)	63	Transformée de Laplace	92
Renouvellement (un théorème)	139	Transformée de Laplace : injectivité	94
Renouvellement	139	Tribu \mathbf{P} -triviale	136
Représentation (mesurabilité par rapport à une v.a.)	35	Tribu	2
		Tribu asymptotique	136
		Tribu atomique	36
		Tribu discrète	36
		Tribu engendrée	3
		Tribu engendrée par des v.a.	34
		Tribu produit	74
		Tribus : stabilité par pré-image	27
		Triviale : tribu \mathbf{P} -triviale	136
		Unicité du prolongement des mesures	11
		v.a. : variable aléatoire	33
		V.a. discrètes : critère de mesurabilité	46
		Variable aléatoire	33
		Variable aléatoire à densité	63, 65
		Variable aléatoire discrète	46
		Variable de loi uniforme	53
		Variance	57
		Variation (distance pour les lois sur \mathbb{Z})	161
		Variation : distance en	157
		Vecteur Gaussien : fonction caractéristique	170
		Vecteurs aléatoires	81
		Vecteurs Gaussiens	170
		Weierstraß (théorème de)	134, 135
		Zéro-un : loi de Kolmogorov	137
Sigma-additivité	7, 14		
Sigma-algèbre	2		
Sigma-fini (mesure)	9		
Sigma-sous-additivité	8, 15		
Simple (fonctions)	31		
Simplification de la mesurabilité	28		
Slutsky (lemme)	151		
Stabilité des tribus par pré-image	27		
Statistique d'ordre	103		
Stieltjes (mesure de)	10		
Stone (lemme de Kochen-Stone)	114		
Suite de v.a. indép. : construction	116		
Tchebychev (inégalité)	57		
Théorème central-limite	165, 171		
Théorème central-limite : De Moivres-Laplace ..	155		
Théorème d'Alexandroff	148		
Théorème d'inversion globale	79		
Théorème d'unicité du prolongement des mesures	11		
Théorème de Berry-Esseen	168		
Théorème de classe monotone fonctionnel ...	33		
Théorème de continuité P. Lévy	162		
Théorème de convergence dominée	44		
Théorème de convergence monotone	39		