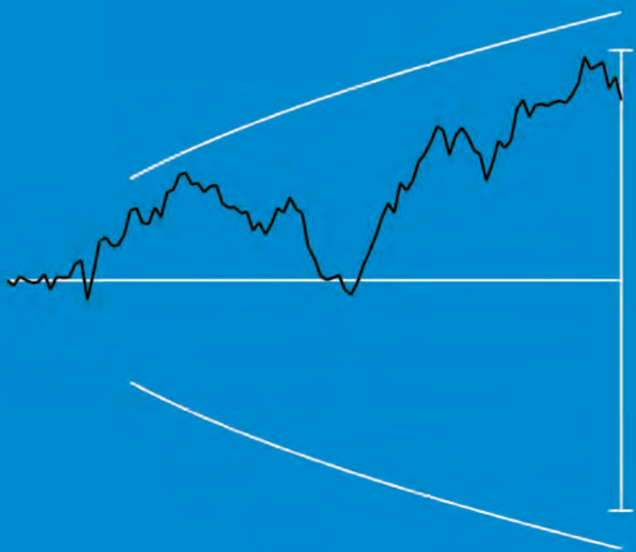
COLLECTION ENSEIGNEMENT SUP //// Mathématiques

L3M1

Probabilité

Philippe Barbe et Michel Ledoux



EDP SCIENCES

PROBABILITÉ

PROBABILITÉ

Philippe Barbe et Michel Ledoux

Collection dirigée par Daniel Guin

EDP SCIENCES

17, avenue du Hoggar
Parc d'activités de Courtabœuf, BP 112
91944 Les Ulis Cedex A, France

L'*illustration de couverture* représente une marche aléatoire centrée, linéairement interpolée ; les courbes supérieure et inférieure sont les bornes de la loi du logarithme itéré, et l'intervalle vertical atteint par la marche aléatoire illustre une application du théorème limite central.

Imprimé en France

ISBN : 978-2-86883-931-2

TABLE DES MATIÈRES

PRÉFACE

Le calcul des probabilités est une branche très vivante des mathématiques actuelles. Les premières formalisations de la notion de hasard au XVII^{e} siècle répondaient pour l'essentiel à diverses questions issues de la théorie des jeux. Au cours du XX^{e} siècle, le calcul des probabilités a trouvé avec A. N. Kolmogorov une axiomatique rigoureuse et efficace s'appuyant sur l'intégration de Lebesgue. L'intuition probabiliste est aujourd'hui un outil efficace dans diverses branches des mathématiques, de l'analyse et la théorie de la mesure jusqu'à la géométrie et même l'algèbre, et forme le support théorique des statistiques modernes.

Ce livre est consacré à l'exposition des notions de base du calcul des probabilités. Il s'appuie de façon essentielle sur la théorie de la mesure et de l'intégration de Lebesgue. (Mesures de probabilités discrètes ou à densité sont donc étudiées dans un même cadre, au titre d'exemples priviligiés les plus usuels.) Les deux premiers chapitres sont en fait un rappel des éléments de base de la théorie élémentaire de la mesure et de l'intégrale de Lebesgue. Ils ne peuvent cependant être considérés comme un traitement exhaustif. Le lecteur peut consulter le livre de J. Faraut, dans la même collection, pour un exposé plus complet. Le chapitre III introduit les premiers aspects des probabilités avec les notions de variables aléatoires et de leurs lois, illustrées par de nombreux exemples. Les fonctions caractéristiques (transformées de Fourier) y sont également étudiées. Le chapitre IV fait réellement entrer le lecteur dans les considérations probabilistes avec le concept d'indépendance. L'addition des variables aléatoires indépendantes y est interprétée comme la traduction fonctionnelle, à la riche intuition, du produit de convolution des mesures. Au chapitre V sont présentées les diverses notions de convergence de suites de variables aléatoires, convergence presque sûre, en probabilité, en loi. La loi des grands nombres et le théorème central limite constituent les exemples fondamentaux de ces divers modes de convergence. Le chapitre suivant est un exposé des notions de conditionnement (probabilités, espérances, lois), illustré par le modèle gaussien. Le chapitre VII est une brève introduction à la notion de martingale à temps discret où sont notamment établis le théorème d'arrêt et les théorèmes de convergence des martingales. Enfin, le dernier chapitre traite succintement de chaînes de Markov (mesures invariantes, convergences). Un appendice présentant les lois de probabilités usuelles avec leurs caractéristiques principales complète la rédaction.

Ce livre est destiné à des étudiants de 3^{e} année de licence de mathématiques ayant suivi un cours de base de mesure et intégration, dont les éléments fondamentaux sont toutefois rappelés dans les deux premiers chapitres. Il ne suppose pas une connaissance préalable des notions de probabilités enseignées d'ordinaire dans les deux premières années de licence et habituellement axés sur les probabilités discrètes et les problèmes de combinatoire dont il n'est fait que très peu état dans cet ouvrage. Ce livre peut être utilisé comme support d'un cours de probabilité de L3, ou d'un premier semestre de master. Cet ouvrage contient en outre les prérequis nécessaires à l'épreuve écrite de mathématiques générales pour l'agrégation ainsi que pour les leçons spécialisées. Chaque chapitre est complété par une série d'exercices destinés à approfondir et à illustrer les éléments de la théorie venant d'être introduits.

Ce livre n'est pas la contribution des seuls auteurs, mais reflète en partie aussi l'enseignement des probabilités par l'équipe du laboratoire de statistique et probabilités de l'université Paul-Sabatier de Toulouse au cours de ces dernières années. Nous remercions ainsi D. Bakry, M. Benaïm, Ph. Carmona, L. Coutin, J.-L. Dunau, G. Letac, D. Michel et tous les membres du laboratoire pour nous avoir permis de puiser librement dans leurs notes de cours et leurs réserves d'exercices, et pour nous avoir conseillé et relu à divers moments de la préparation. Nous remercions tout particulièrement D. Michel et X. Milhaud pour avoir suppléé le chapitre VIII sur les chaînes de Markov, ainsi que pour leur soutien et leur aide. P. Lezaud a relu avec un soin extrême tout le manuscrit et a testé la plupart des exercices. Qu'il soit sincèrement remercié pour cette tâche bien ingrate. Un dernier mot enfin. Le temps passé à la rédaction de ce livre est très certainement insuffisant pour que cet ouvrage puisse prétendre à beaucoup d'originalité et pour que le résultat soit à la hauteur des espérances et de l'enthousiasme des premières lignes. Il ne saurait être aussi exempt d'imperfections et d'erreurs pour lesquels nous nous excusons par avance.

Un chapitre est numéroté par un chiffre romain, et une section de chapitre par un chiffre arabe. Un énoncé dans une section est désigné par le numéro de la section et le numéro d'ordre de cet énoncé dans la section. Ainsi, II.3.4 désigne l'énoncé 4 dans la section 3 du chapitre II.

Toulouse, septembre 1998 — Ph. Barbe, M. Ledoux

Préface à la seconde édition

Nous remercions les éditions EDP Sciences, ainsi que l'éditeur scientifique de la collection, D. Guin, de nous proposer de publier une nouvelle édition de notre ouvrage paru en 1998.

Le texte est pour l'essentiel identique à la version intiale. Celle-ci comporte un nombre trop important d'erreurs, mineures ou plus sérieuses, qui nuisent beaucoup à sa lisibilité. Nous avons essayé de corriger les principales erreurs et imperfections (sans toutefois pouvoir prétendre les avoir éliminées toutes). Plusieurs corrections nous ont été aimablement communiquées par divers collègues. Nous remercions tout particulièrement R. Ben David pour ses corrections et commentaires très minutieux (même si nous ne les avons pas tous suivis). Nous remercions aussi M. Arnaudon, Fr. Barthe, M. Benaïm, B. Bercu, Ph. Carmona, H. Carrieu, R. Chomienne, S. Cohen, Th. Delmotte, Th. Gallay, Ch. Leuridan, P. Lezaud et D. Robert.

H. Carrieu prépare actuellement un fascicule des exercices corrigés de ce livre. Nous le remercions bien vivement pour cet excellent complément.

Paris, Toulouse, septembre 2006 — Ph. Barbe, M. Ledoux

I

THÉORIE DE LA MESURE

L'objet de ce chapitre est de rappeler les éléments de théorie de la mesure qui seront indispensables au développement du calcul des probabilités dans les chapitres suivants. Une mesure abstraite sur un ensemble Ω généralise la notion de longueur, d'aire ou de volume, sur la droite, le plan ou l'espace. Intuitivement, le lien avec les probabilités est qu'une probabilité mesure la vraisemblance d'un événement.

Sur la droite (ou le plan, ou l'espace), la longueur (ou l'aire, ou le volume) est une fonction qui à un ensemble associe un nombre réel positif. Cette fonction est additive, au sens où appliquée à $A \cup B$, elle est la somme de la fonction appliquée en A et de la fonction appliquée en B, pourvu que A et B soient disjoints. On demandera à une mesure abstraite de vérifier cette additivité.

Un fait peu intuitif est qu'il existe des sous-ensembles de la droite (ou du plan, ou de l'espace) pour lesquels on ne peut pas définir leur longueur (ou aire, ou volume) (*cf.* exercice I.6). Il convient donc, dans un premier temps, de définir la classe d'ensembles que l'on veut (et peut) mesurer. Compte tenu de la propriété d'additivité décrite au paragraphe précédent, on imposera par exemple que cette classe soit stable par réunion finie.

I.1. Algèbre, tribu

Soit Ω un ensemble.

Exemples I.1.1. (i) Ω pourra être $\mathbb{R}$ ou $\mathbb{R}^d$, un espace métrique, ou plus généralement topologique.

(ii) On joue au dé en le lançant une fois. L'ensemble Ω peut être pris comme l'ensemble des faces du dé, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Lorsque l'on lance le dé au hasard, cela revient à choisir (« au hasard ») un élément de Ω.

Il convient de remarquer que l'on peut toujours ajouter des points à Ω. Dans l'exemple I.1.1.ii nous pourrions tout aussi bien prendre $\Omega = \{1, 2, 3, 4, 5, 6, 7\}$. Mais intuitivement, 7 a une probabilité nulle d'être réalisé.

On considère $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω. Un sous-ensemble $\mathcal{C}$ de $\mathcal{P}(\Omega)$ est un ensemble de parties de Ω.

Définition I.1.2. Un sous-ensemble $\mathcal{C}$ de $\mathcal{P}(\Omega)$ est une algèbre (de Boole) sur Ω si

(i) $\Omega \in \mathcal{C}$,

(ii) $\mathcal{C}$ est stable par passage au complémentaire (*i.e.* $A \in \mathcal{C} \Rightarrow \Omega \setminus A \in \mathcal{C}$),

(iii) $\mathcal{C}$ est stable par réunion finie (*i.e.* $A_1, \ldots, A_k \in \mathcal{C} \Rightarrow A_1 \cup \cdots \cup A_k \in \mathcal{C}$).

Dans l'axiome (iii) de la définition I.1.2, on pourrait se contenter de $k = 2$, le cas général s'en déduisant par récurrence. Par passage au complémentaire, une algèbre est aussi stable par intersection finie.

Définition I.1.3. Un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ est une tribu sur Ω si

(i) $\Omega \in \mathcal{A}$,

(ii) $\mathcal{A}$ est stable par passage au complémentaire (*i.e.* $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$),

(iii) $\mathcal{A}$ est stable par réunion dénombrable (*i.e.* $A_i \in \mathcal{A}$, $i \in \mathbb{N} \Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$).

On dit aussi que $\mathcal{A}$ est une σ-algèbre. Le couple $(\Omega, \mathcal{A})$ formé d'un ensemble Ω et d'une tribu $\mathcal{A}$ sera appelé un espace mesurable. Les éléments de $\mathcal{A}$ sont appelés ensembles mesurables.

Toute tribu est une algèbre.

Expliquons le sens de ces deux définitions. Tout d'abord le « σ » de σ-algèbre est un acronyme de « dénombrable » par référence à l'axiome (iii) dans la définition d'une tribu.

Exemples I.1.4. (i) $\mathcal{P}(\Omega)$ est toujours une algèbre et une tribu.

(ii) Le sous-ensemble $\{\emptyset, \Omega\}$ de $\mathcal{P}(\Omega)$, composé de la partie vide et de Ω, est une algèbre et une tribu, appelée algèbre ou tribu triviale.

(iii) L'ensemble des ouverts de $\mathbb{R}^d$ n'est pas une algèbre (et donc n'est pas une tribu) car le complémentaire d'un ouvert n'est pas nécessairement ouvert.

(iv) Une réunion de deux algèbres n'est pas une algèbre en général. Considérer par exemple $\Omega = \{0,1,2\}$, les algèbres $\mathcal{C}_1 = \{\emptyset, \{0,1,2\}, \{0\}, \{1,2\}\}$ et $\mathcal{C}_2 = \{\emptyset, \{0,1,2\}, \{1\}, \{0,2\}\}$, puis remarquer que la réunion de $\{0\}$ et $\{1\}$ n'appartient pas à $\mathcal{C}_1 \cup \mathcal{C}_2$.

(v) Une intersection d'un nombre quelconque d'algèbres (resp. de tribus) est une algèbre (resp. une tribu).

Certains auteurs définissent les algèbres comme étant stables par réunion et intersection finies.

En général, il est difficile d'expliciter tous les éléments d'une tribu. Les algèbres et les tribus se décrivent le plus souvent par leurs éléments générateurs.

Définition I.1.5. Soit $\mathcal{E}$ un sous-ensemble de $\mathcal{P}(\Omega)$.

(i) L'algèbre $\mathcal{C}(\mathcal{E})$ engendrée par $\mathcal{E}$ est l'intersection de toutes les algèbres contenant $\mathcal{E}$.

(ii) La tribu $\sigma(\mathcal{E})$ engendrée par $\mathcal{E}$ est l'intersection de toutes les tribus contenant $\mathcal{E}$.

Compte tenu de la définition I.1.5, on peut parler de la tribu engendrée par deux tribus $\mathcal{A}_1$ et $\mathcal{A}_2$, que l'on note $\mathcal{A}_1 \vee \mathcal{A}_2$ ou aussi $\sigma(\mathcal{A}_1 \cup \mathcal{A}_2)$, ou encore $\sigma(\mathcal{A}_1, \mathcal{A}_2)$. On prendra bien soin de remarquer, d'après l'exemple I.1.4.iv, que $\mathcal{A}_1 \vee \mathcal{A}_2$ est en général différent de $\mathcal{A}_1 \cup \mathcal{A}_2$.

Exemples I.1.6. (i) Soit A une partie de Ω. L'algèbre $\mathcal{C}(\{A\})$ et la tribu $\sigma(\{A\})$ sont $\{\emptyset, \Omega, A, A^{\mathrm{c}}\}$.

(ii) Plus généralement, si $\mathcal{S} = \{S_1, \ldots, S_n\}$ est une partition finie de Ω, c'est-à-dire $\Omega = \bigcup_{1\leq i\leq n} S_i$ et $S_i \cap S_j = \emptyset$ pour $i \neq j$, alors

$$\mathcal{C}(\mathcal{S}) = \Big\{ \bigcup_{i\in T} S_i \,:\, T \subset \{1,\ldots,n\} \Big\},$$

où T parcourt l'ensemble des parties de $\{1,\ldots,n\}$, l'ensemble vide compris. En particulier, $\mathcal{C}(\mathcal{S})$ est en bijection avec l'ensemble des parties de $\{1,\ldots,n\}$ et se compose de 2^n éléments.

(iii) Si $\mathcal{S} = \{S_i \,:\, i \in \mathbb{N}\}$ est une partition de Ω, alors

$$\sigma(\mathcal{S}) = \Big\{ \bigcup_{i\in T} S_i \,:\, T \subset \mathbb{N} \Big\}.$$

Définition I.1.7. Si Ω est un espace topologique, on appelle tribu borélienne, notée $\mathcal{B}(\Omega)$, la tribu engendrée par les ouverts de Ω. Un borélien est un ensemble appartenant à la tribu borélienne.

La tribu borélienne est aussi engendrée par les fermés puisque la tribu est stable par passage au complémentaire.

Exemple I.1.8. Sur $\mathbb{R}$, la tribu borélienne coïncide avec la tribu engendrée par les intervalles $]\,a,b\,[$, $-\infty \leq a < b \leq \infty$. Elle coïncide aussi avec la tribu engendrée par les intervalles $[\,a,b\,]$, ou $]\,a,b\,]$, ou $[\,a,b\,[$.

On prendra bien soin de constater que si les éléments d'une famille génératrice sont explicites, il n'en est rien en général des éléments de la tribu (la plupart des boréliens de $\mathbb{R}$ ne sont pas des intervalles !).

Dans la suite, lorsque Ω est $\mathbb{R}^d$ (ou un espace topologique), il sera toujours muni de sa tribu borélienne. Si Ω est discret, on le munira de la tribu de ses parties.

Lorsque l'on a deux ensembles Ω_1 et Ω_2, on définit leur produit $\Omega_1 \times \Omega_2$, sur lequel on peut éventuellement définir des structures produits (topologie produit, groupe produit, etc). Lorsque l'on a des espaces mesurables $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, on souhaite faire de l'espace produit $\Omega_1 \times \Omega_2$ un espace mesurable.

Définition I.1.9. Soient $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, deux espaces mesurables. On appelle ensemble élémentaire de $\Omega = \Omega_1 \times \Omega_2$ une réunion finie de pavés $A_1 \times A_2$, avec $A_i \in \mathcal{A}_i$, $i = 1, 2$. La tribu produit $\mathcal{A}_1 \otimes \mathcal{A}_2$ sur Ω est la tribu engendrée par les ensembles élémentaires.

Exemples I.1.10. (i) Les ensembles élémentaires forment une algèbre.

(ii) En utilisant le fait que tout ouvert de $\mathbb{R}^2$ peut s'écrire comme une réunion dénombrable de pavés d'intervalles ouverts, on montre que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. On montre de même que la tribu sur $\mathbb{R}^d$ engendrée par d copies de $\mathcal{B}(\mathbb{R})$ est $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$.

De façon générale, en mathématique, lorsqu'une structure est définie sur un espace, on souhaite pouvoir la transporter sur d'autres espaces par des fonctions. En général, on utilise d'ailleurs les images réciproques par les fonctions. Par exemple, sur $\mathbb{R}$, la structure d'ordre est préservée par la réciproque d'une application croissante (*i.e.* si $x < y$ sont dans l'image de $\mathbb{R}$ par une fonction f croissante, alors

$f^{-1}(x) < f^{-1}(y)$). De même, la structure topologique est préservée par application de la réciproque d'une application continue (*i.e.* f est continue si $f^{-1}(U)$ est ouvert pour tout ouvert U). La notion analogue dans le contexte de la théorie de la mesure est celle de mesurabilité.

Si f est une application de Ω dans E et si B est une partie de E, on notera

$$f^{-1}(B) = \{\, \omega \in \Omega \,:\, f(\omega) \in B \,\}\,.$$

Si $\mathcal{B}$ est une famille de parties de E, on notera

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

Noter que si $\mathcal{B}$ est une algèbre (resp. tribu), $f^{-1}(\mathcal{B})$ est une algèbre (resp. tribu) d'après les propriétés de l'image réciproque ensembliste f^{-1}.

Définition I.1.11. (i) Soient $(\Omega, \mathcal{A})$ et $(E, \mathcal{B})$, deux espaces mesurables. Soit f une fonction de Ω dans E. On dit que f est mesurable (pour $\mathcal{A}$ et $\mathcal{B}$) si $f^{-1}(\mathcal{B}) \subset \mathcal{A}$; c'est-à-dire, $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$.

(ii) Si f est une fonction de Ω dans $(E, \mathcal{B})$, on appelle tribu engendrée par f, notée $\sigma(f)$, la plus petite tribu (sur Ω) qui rend f mesurable ; autrement dit, $\sigma(f) = f^{-1}(\mathcal{B})$.

(iii) Plus généralement, si $\mathcal{F}$ est une famille de fonctions d'un ensemble Ω à valeurs dans $(E, \mathcal{B})$, on appelle tribu engendrée par $\mathcal{F}$ la plus petite tribu (sur Ω) qui rend mesurable toute fonction de $\mathcal{F}$ (*i.e.* la tribu engendrée par les ensembles de la forme $f^{-1}(B)$ pour $B \in \mathcal{B}$ et $f \in \mathcal{F}$). On la note $\sigma(\mathcal{F})$.

Avec les notations de cette définition, dire que f est mesurable de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{B})$ revient à dire que $\sigma(f) \subset \mathcal{A}$.

Exemples I.1.12. (i) Si A est une partie de Ω, on définit la fonction indicatrice de A par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbb{1}_A(\omega) = 0$ si $\omega \notin A$. Soit $\mathcal{A}$ une tribu sur Ω. En tant que fonction à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la fonction $\mathbb{1}_A$ est mesurable pour $\mathcal{A}$ si et seulement si $A \in \mathcal{A}$.

(ii) Soit $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$ et soit Π_1 la projection de $\mathbb{R} \times \mathbb{R}$ sur sa première composante $\mathbb{R}$ définie par $\Pi_1(x, y) = x$. La tribu engendrée par Π_1 est formée des ensembles $B \times \mathbb{R}$ où B décrit les boréliens de $\mathbb{R}$. Cette tribu est différente de la tribu borélienne de $\mathbb{R}^2$. On notera que Π_1 est mesurable de $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bien que $\sigma(\Pi_1)$ ne coïncide pas avec la tribu borélienne de $\mathbb{R}^2$.

(iii) La tribu borélienne de $\mathbb{R}^2$ est engendrée par les projections Π_1 et Π_2 sur les coordonnées. En effet, $\Pi_1^{-1}(A) \cap \Pi_2^{-1}(B) = (A \times \Omega) \cap (\Omega \times B) = A \times B$, et les rectangles engendrent la tribu $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ (*cf.* I.1.9 et I.1.10).

Définition I.1.13. Une fonction mesurable de $(\Omega, \mathcal{A})$ dans un espace topologique muni de sa tribu borélienne $(E, \mathcal{B}(E))$ est dite borélienne.

Puisque nous munirons toujours $\mathbb{R}$ ou $\mathbb{R}^d$ de sa tribu borélienne, les fonctions mesurables à valeurs réelles sont boréliennes.

En pratique les tribus étant le plus souvent définies par une partie génératrice, la définition I.1.11 est difficile à vérifier. La proposition suivante montre que pour qu'une fonction soit mesurable, il suffit de vérifier sa propriété caractéristique sur une famille génératrice de la tribu d'arrivée.

Proposition I.1.14. *Soient Ω et E deux ensembles. Soit $\mathcal{E} \subset \mathcal{P}(E)$ et soit $\mathcal{B} = \sigma(\mathcal{E})$. La tribu engendrée par une fonction f de Ω dans $(E, \mathcal{B})$ est $\sigma(f) = \sigma(f^{-1}(\mathcal{E})) = \sigma(\{ f^{-1}(C) : C \in \mathcal{E} \})$.*

Plus généralement, si $\mathcal{F}$ est une famille de fonctions de Ω dans $(E, \mathcal{B})$, alors $\sigma(\mathcal{F}) = \sigma(\{ f^{-1}(C) : C \in \mathcal{E} \,;\, f \in \mathcal{F} \})$.

En particulier, pour qu'une fonction f de $(\Omega, \mathcal{A})$ dans $(E, \sigma(\mathcal{E}))$ soit mesurable, il suffit que $f^{-1}(\mathcal{E})$ soit inclus dans $\mathcal{A}$.

Démonstration. Soit

$$\mathcal{T} = \Big\{ B \subset E : f^{-1}(B) \in \sigma\big(f^{-1}(\mathcal{E})\big) \Big\} .$$

Il est aisé de vérifier que $\mathcal{T}$ est une tribu qui contient $\mathcal{E}$. Donc $\mathcal{T}$ contient $\sigma(\mathcal{E})$. Soit à présent $A \in \sigma(f)$. Par définition, $A = f^{-1}(B)$ pour un certain $B \in \sigma(\mathcal{E})$. Il s'ensuit $B \in \mathcal{T}$ et par construction de $\mathcal{T}$, $A = f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))$. Ainsi, $\sigma(f) \subset \sigma(f^{-1}(\mathcal{E}))$. L'inclusion réciproque est évidente.

Le cas d'une famille quelconque se traite de la même façon.

Enfin, si $f^{-1}(\mathcal{E}) \subset \mathcal{A}$, alors $\sigma(f^{-1}(\mathcal{E})) \subset \mathcal{A}$. Comme $\sigma(f^{-1}(\mathcal{E})) = \sigma(f)$ par le premier point, la conclusion s'ensuit. □

I.2. Ensembles de fonctions mesurables

Nous rassemblons ici quelques faits sur les fonctions mesurables, montrant que c'est une classe assez naturelle de fonctions.

Proposition I.2.1. *La composée de deux fonctions mesurables est mesurable.*

Démonstration. Soient $f_i : (\Omega_i, \mathcal{A}_i) \to (\Omega_{i+1}, \mathcal{A}_{i+1})$, $i = 1,2$, mesurables. Soit $A \in \mathcal{A}_3$. On a $(f_1 \circ f_2)^{-1}(A) = f^{-1}(f_2^{-1}(A))$. Puisque f_2 est mesurable, $f_2^{-1}(A) \in \mathcal{A}_2$, et puisque f_1 est mesurable, $f_1^{-1}(f_2^{-1}(A)) \in \mathcal{A}_1$. □

Lemme I.2.2. *Si f, g sont des fonctions mesurables de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors $\omega \in \Omega \mapsto (f(\omega), g(\omega)) \in \mathbb{R}^2$ est mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.*

Démonstration. Soit $A \times B$ un rectangle dans $\mathcal{B}(\mathbb{R}^2)$, et $h(\omega) = (f(\omega), g(\omega))$. Alors, $h^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B) \in \mathcal{A}$. Puisque les rectangles engendrent $\mathcal{B}(\mathbb{R}^2)$, on conclut grâce à la proposition I.1.14. □

Les fonctions mesurables par rapport à une tribu borélienne forment une classe plus vaste que les fonctions continues :

Proposition I.2.3. *Soient Ω_1, Ω_2 deux espaces topologiques munis de leur tribu borélienne. Toute fonction continue de Ω_1 dans Ω_2 est mesurable (ou borélienne ici).*

Démonstration. Remarquer que si U est ouvert dans Ω_2 et f est une fonction continue, $f^{-1}(U)$ est ouvert. Puis appliquer la proposition I.1.14. □

Si x et y sont deux nombres réels, on note $x \vee y$ leur maximum.

Corollaire I.2.4. *L'espace des fonctions mesurables (boréliennes) de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est stable pour les opérations de multiplication par une constante $(\lambda f)(\omega) = \lambda f(\omega)$ ($\lambda \in \mathbb{R}$), d'addition $(f+g)(\omega) = f(\omega) + g(\omega)$, de multiplication $(fg)(\omega) = f(\omega)g(\omega)$, et du maximum $(f \vee g)(\omega) = f(\omega) \vee g(\omega)$*

Démonstration. La fonction $\omega \mapsto \lambda f(\omega)$ est la composée de la fonction mesurable f et de la fonction continue $x \mapsto \lambda x$. De même $f+g$ (resp. fg, resp. $f \vee g$) est la composée de la fonction mesurable $\omega \mapsto (f(\omega), g(\omega))$ (en vertu du lemme I.2.2), et de la fonction continue $(x,y) \mapsto x+y$ (resp. $(x,y) \mapsto xy$, resp. $(x,y) \mapsto x \vee y$). □

Il est facile de voir qu'une limite ponctuelle de fonctions croissantes est croissante, mais qu'une limite ponctuelle de fonctions continues n'est pas nécessairement continue. La classe des fonctions mesurables est stable par limite simple.

Théorème I.2.5. *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables de* $(\Omega, \mathcal{A})$ *dans un espace métrique* (E, d) *muni de sa tribu borélienne. Si* f_n *converge ponctuellement vers* f *(i.e. pour tout* $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$*), alors* f *est mesurable.*

Démonstration. D'après la proposition I.1.14, il suffit de montrer que si U est ouvert dans E, alors $f^{-1}(U) \in \mathcal{A}$. Posons $U_r = \{\, x \in U \ : \ d(x, E \setminus U) > 1/r \,\}$, $r \geq 1$ entier. L'ensemble U_r est ouvert, donc est un borélien de E. Ainsi,

$$f^{-1}(U) = \bigcup_{r,m} \bigcap_{n\geq m} f_n^{-1}(U_r)$$

est un borélien. □

On peut approcher toute fonction mesurable par des fonctions mesurables plus simples.

Définition I.2.6. Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle fonction étagée (à valeurs dans $\mathbb{R}^d$) une fonction de la forme $f(\omega) = \sum_{1\leq i\leq k} a_i \mathbb{1}_{A_i}(\omega)$ où les A_i sont des éléments disjoints de $\mathcal{A}$, et où les coefficients a_i appartiennent à $\mathbb{R}^d$.

Proposition I.2.7. *Toute fonction* f *mesurable de* $(\Omega, \mathcal{A})$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *est limite simple de fonctions étagées. Si* f *est positive, la limite peut être choisie croissante.*

Démonstration. Prenons d'abord f positive. Définissons pour $n, k \geq 1$,

$$A_{n,k} = \left\{ \omega \ : \ \frac{k-1}{2^n} \leq f(\omega) < \frac{k}{2^n} \right\}.$$

Les $A_{n,k}$ sont éléments de $\mathcal{A}$ en tant qu'images réciproques par la fonction mesurable f d'intervalles. La suite

$$f_n(\omega) = \sum_{1\leq k\leq 2^{n^2}} \frac{k-1}{2^n} \mathbb{1}_{A_{n,k}}(\omega)$$

converge en croissant vers f.

Si f est quelconque, écrivons $f = f^+ - f^-$ avec $f^+ = f \vee 0$ et $f^- = (-f) \vee 0$, et approximons les fonctions positives f^+ et f^- par la méthode précédente. □

I.3. Classes monotones

Nous souhaitons finalement pouvoir mesurer les éléments d'une tribu, c'est-à-dire définir une fonction qui à chaque ensemble de la tribu associe un réel positif, et qui vérifie un certain nombre d'axiomes. Une des difficultés a priori est qu'une tribu peut contenir beaucoup d'ensembles. On souhaite donc pouvoir définir la mesure sur une classe plus restreinte d'ensembles et avoir un procédé d'extension permettant alors de la définir sur toute la tribu. Le but de cette section est de construire le bon outil pour réaliser le procédé d'extension. Son intérêt apparaîtra clairement dans la suite du cours.

Définition I.3.1. Une famille $\mathcal{M}$ de parties de Ω est appelée une classe monotone si
(i) $\Omega \in \mathcal{M}$,
(ii) si $A, B \in \mathcal{M}$ et $B \subset A$, alors $A \setminus B \in \mathcal{M}$,
(iii) $\mathcal{M}$ est stable par réunion monotone croissante (*i.e.* $A_i \in \mathcal{M}$, $i \in \mathbb{N}$, $A_i \subset A_{i+1} \Rightarrow \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$).

Si $\mathcal{E} \subset \mathcal{P}(\Omega)$, on note $\mathcal{M}(\mathcal{E})$ la classe monotone engendrée par $\mathcal{E}$, c'est-à-dire l'intersection de toute les classes monotones contenant $\mathcal{E}$.

Exemples I.3.2. (i) Une tribu est une classe monotone.
(ii) Une classe monotone $\mathcal{M}$, stable par intersection finie, est une tribu. En effet, $\mathcal{M}$ est aussi stable par réunion finie en vertu de I.3.1.ii, et toute réunion peut s'écrire comme une réunion croissante ($\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}}(\bigcup_{j\leq i} A_j)$ pour toute famille A_i, $i \in \mathbb{N}$).

Pour que la définition d'une classe monotone engendrée par $\mathcal{E}$ ait un sens, il faut vérifier que l'intersection de deux, ou d'un nombre quelconque, de classes monotones est une classe monotone.

Le théorème important suivant affirme que la classe monotone engendrée par une famille de parties de Ω stable par intersection finie coïncide avec la tribu engendrée par cette famille.

***Théorème I.3.3* (des classes monotones).** *Soit $\mathcal{E}$ une famille de parties de Ω, stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.*

Démonstration. En vertu de l'exemple I.3.2.i, $\sigma(\mathcal{E})$ est une classe monotone qui contient $\mathcal{E}$ et donc $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Pour démontrer l'inclusion inverse, nous montrons que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie. Alors, d'après I.3.2.ii, $\mathcal{M}(\mathcal{E})$

sera une tribu contenant $\mathcal{E}$, et donc $\sigma(\mathcal{E}) \subset \mathcal{M}(\mathcal{E})$. Il suffit de prouver que si $A, B \in \mathcal{M}(\mathcal{E})$, alors $A \cap B \in \mathcal{M}(\mathcal{E})$. Soit

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

L'ensemble $\mathcal{M}_1$ est une classe monotone qui contient $\mathcal{E}$, donc $\mathcal{M}(\mathcal{E})$. Soit à présent

$$\mathcal{M}_2 = \{\, B \in \mathcal{M}(\mathcal{E}) \ : \ \forall\, C \in \mathcal{M}(\mathcal{E})\,,\, B \cap C \in \mathcal{M}(\mathcal{E})\,\}\,.$$

L'ensemble $\mathcal{M}_2$ est une classe monotone. Montrons qu'il contient $\mathcal{E}$. Il faut démontrer pour cela que si $B \in \mathcal{E}$, alors

$$\forall\, C \in \mathcal{M}(\mathcal{E})\,,\, B \cap C \in \mathcal{M}(\mathcal{E})\,.$$

Or $C \in \mathcal{M}(\mathcal{E}) \subset \mathcal{M}_1$, et donc, puisque $B \in \mathcal{E}$, $B \cap C = C \cap B \in \mathcal{M}(\mathcal{E})$. Ainsi, $\mathcal{M}_2 \supset \mathcal{E}$, donc $\mathcal{M}_2 \supset \mathcal{M}(\mathcal{E})$, ce qui montre que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie. Le théorème est établi. □

Il existe dans la littérature différentes définitions d'une classe monotone donnant lieu à différentes versions du théorème des classes monotones. Par exemple, on peut supprimer l'axiome (ii) de la définition I.3.1 et imposer dans le théorème I.3.3 que la classe $\mathcal{E}$ soit en outre stable par passage au complémentaire, c'est-à-dire qu'elle soit une algèbre. La version présentée est la mieux adaptée à l'étude de l'indépendance dans le chapitre IV.

Nous étudions à présent la version fonctionnelle du théorème des classes monotones. Pour cela, si f est une fonction définie sur un espace Ω, à valeurs réelles et bornée, nous notons $\|f\|_\infty = \sup\{|f(\omega)| \ : \ \omega \in \Omega\,\}$ sa norme uniforme.

Rappelons qu'une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ sur Ω est croissante si pour tout ω dans Ω, la suite $f_n(\omega)$ est croissante. De plus, la suite est bornée s'il existe une constante positive C telle que $|f_n(\omega)| \leq C$ pour tout n et tout ω ; autrement dit, si $\sup_{n\in\mathbb{N}} \|f_n\|_\infty$ est fini. Clairement, une suite croissante bornée converge.

Définition I.3.4. (i) Un ensemble $\mathcal{H}$ de fonctions de Ω dans $\mathbb{R}$ est dit stable par convergence monotone bornée si la limite de toute suite croissante et bornée de $\mathcal{H}$ est aussi dans $\mathcal{H}$.

(ii) Un ensemble $\mathcal{H}$ est dit monotone s'il contient les constantes et est stable par convergence monotone bornée.

Observons que l'intersection d'un nombre arbitraire d'ensembles monotones de fonctions réelles définies sur Ω est un ensemble monotone.

Rappelons que si $\mathcal{F}$ est une famille de fonctions sur Ω à valeurs réelles, $\sigma(\mathcal{F})$ désigne la plus petite tribu sur Ω qui rend mesurables les fonctions de $\mathcal{F}$ lorsque $\mathbb{R}$ est muni de sa tribu borélienne. Rappelons enfin qu'un ensemble $\mathcal{C}$ est stable par multiplication si pour tous f, g dans $\mathcal{C}$, le produit fg est dans $\mathcal{C}$.

Le théorème suivant peut être considéré comme une version mesurable du théorème de Stone-Weierstrass.

Théorème I.3.5 **(des classes monotones fonctionnelles).** *Soit $\mathcal{C}$ un ensemble de fonctions réelles bornées sur Ω stable par multiplication et contenant les constantes. Tout espace vectoriel monotone contenant $\mathcal{C}$ contient les fonctions bornées mesurables par rapport à $\sigma(\mathcal{C})$.*

Démonstration. L'ensemble des fonctions réelles bornées sur Ω est un espace vectoriel. On peut ainsi considérer $\mathcal{H}_0$, le plus petit sous-espace vectoriel monotone contenant $\mathcal{C}$. Puisque $\mathcal{C}$ contient les constantes, $\mathcal{H}_0$ les contient aussi. Il suffit de montrer que $\mathcal{H}_0$ contient les fonctions bornées mesurables par rapport à $\sigma(C)$.

Lemme I.3.6. *$\mathcal{H}_0$ est stable par multiplication.*

Démonstration. Elle est semblable à la démonstration du théorème I.3.3. Si A et B sont deux ensembles de fonctions réelles, on note AB l'ensemble des fonctions de la forme fg où f est dans A et g dans B. Dire que A est stable par multiplication revient à dire que AA est inclus dans A. On montrera d'abord que $\mathcal{C}\mathcal{H}_0 \subset \mathcal{H}_0$, puis que $\mathcal{H}_0\mathcal{H}_0 \subset \mathcal{H}_0$.

Pour toute fonction f de $\mathcal{H}_0$, définissons l'ensemble

$$\mathcal{H}_f = \{\, g \in \mathcal{H}_0 \;:\; fg \in \mathcal{H}_0 \,\}\,.$$

C'est un espace vectoriel. Il contient les constantes puisque d'une part toute fonction constante est dans $\mathcal{C}$ et donc dans $\mathcal{H}_0$, et d'autre part parce que f est prise dans $\mathcal{H}_0$. Si f est bornée, cet espace vectoriel est aussi stable par convergence monotone bornée ; en effet, si (g_n) est une suite de fonctions bornées dans $\mathcal{H}_f$, convergeant en croissant vers g dans $\mathcal{H}_0$, l'égalité $fg_n = (f + \|f\|_\infty)g_n - \|f\|_\infty g_n$ et l'appartenance de fg_n et $\|f\|_\infty g_n$ à $\mathcal{H}_0$ montrent que $(f + \|f\|_\infty)g$ et $\|f\|_\infty g$ sont aussi dans $\mathcal{H}_0$, et donc que fg est dans $\mathcal{H}_0$.

Si f est dans $\mathcal{C}$, l'espace $\mathcal{H}_f$ contient $\mathcal{C}$. Donc, puisque nous venons de voir que $\mathcal{H}_f$ est un espace vectoriel stable par convergence monotone bornée, il contient aussi $\mathcal{H}_0$. On a donc $\mathcal{H}_0 = \bigcap_{f\in\mathcal{C}} \mathcal{H}_f$, ce qui montre $\mathcal{C}\mathcal{H}_0 \subset \mathcal{H}_0$.

Si maintenant f est dans $\mathcal{H}_0$, nous déduisons de ce qui précède que $\mathcal{C}$ est inclus dans $\mathcal{H}_f$. Donc, par minimalité de $\mathcal{H}_0$, on a $\mathcal{H}_0 \subset \mathcal{H}_f$. Ainsi, $\mathcal{H}_0 = \bigcap_{f\in\mathcal{H}_0} \mathcal{H}_f$, ce qui montre que $\mathcal{H}_0\mathcal{H}_0 \subset \mathcal{H}_0$. □

Si $\mathcal{B}$ est une tribu, on note $b(\mathcal{B})$ l'ensemble des fonctions $\mathcal{B}$-mesurables bornées.

Lemme I.3.7. *Si $\mathcal{H}_0$ est un espace vectoriel monotone de fonctions bornées, stable par multiplication, alors il coïncide avec $b\big(\sigma(\mathcal{H}_0)\big)$.*

Démonstration. On a l'inclusion évidente $\mathcal{H}_0 \subset b\big(\sigma(\mathcal{H}_0)\big)$.
i) Montrons que $\mathcal{H}_0$ est stable par l'application valeur absolue. Soit donc f une fonction non nulle dans $\mathcal{H}_0$. Quitte à remplacer f par $f/\|f\|_\infty$, on peut supposer $|f| \leq 1$. Observons que

$$|f| = \sqrt{1-(1-f^2)} = 1 - \sum_{i\geq 1} \alpha_i (1-f^2)^i$$

où les α_i sont positifs. Il s'ensuit que $1-|f|$ est la limite croissante des fonctions bornées $\sum_{1\leq i\leq n} \alpha_i(1-f^2)^i$. Puisque chacune de ces fonctions est dans $\mathcal{H}_0$ ainsi que les constantes, la fonction $|f|$ est aussi dans $\mathcal{H}_0$.
ii) Montrons que $\mathcal{H}_0$ est stable par les opérations maximum $\vee$ et minimum $\wedge$. Pour cela, soient f et g deux fonctions de $\mathcal{H}_0$. Puisque $f^+ = (f+|f|)/2$ et $f^- = -(-f)^+$ sont dans $\mathcal{H}_0$, les représentations $f \vee g = g + (f-g)^+$ et $f \wedge g = -\big((-f)\vee(-g)\big)$ montrent la stabilité de $\mathcal{H}_0$ par maximum et minimum.
iii) Montrons que l'ensemble $\mathcal{A} = \{\, A \subset \Omega : \mathbb{1}_A \in \mathcal{H}_0 \,\}$ est une tribu. Puisque $\mathcal{H}_0$ contient les constantes, $\mathcal{A}$ contient Ω. D'autre part, $\mathcal{A}$ est stable par complémentation, puisque si $\mathbb{1}_A$ est dans $\mathcal{H}_0$, alors $1-\mathbb{1}_A$ est aussi dans $\mathcal{H}_0$. Enfin, si A_n, $n\in\mathbb{N}$, est une suite d'éléments de $\mathcal{A}$, $\mathbb{1}_{\cup_{n\in\mathbb{N}} A_n}$ est la limite monotone bornée des fonctions $\max_{i\leq n} \mathbb{1}_{A_i}$ de $\mathcal{H}_0$, et donc appartient à $\mathcal{H}_0$.

Le reste de la démonstration consiste à montrer que $\mathcal{A}$ et $\sigma(\mathcal{H}_0)$ coïncident, puis que $\mathcal{H}_0$ et $b(\mathcal{A})$ coïncident aussi.
iv) Montrons que $\mathcal{A} \subset \sigma(\mathcal{H}_0)$. Si A est dans $\mathcal{A}$, alors $\mathbb{1}_A$ est dans $\mathcal{H}_0$. En écrivant A comme l'image réciproque de $\{\,1\,\}$ par $\mathbb{1}_A$, on voit que A est dans $\sigma(\mathcal{H}_0)$.
v) Montrons que $\sigma(\mathcal{H}_0) \subset \mathcal{A}$. D'après la définition I.1.11, il convient de montrer que toute fonction de $\mathcal{H}_0$ est $\mathcal{A}$-mesurable. Soit donc f dans $\mathcal{H}_0$. Quitte à remplacer f par $f + \|f\|_\infty$, on peut supposer que f est positive. Il suffit de montrer que pour tout t positif, $\{\, f \geq t \,\}$, ou autrement dit $\{\, f/t \geq 1 \,\}$ est dans $\mathcal{A}$. Donc, en remplaçant f par f/t, il suffit de montrer que $\{\, f \geq 1 \,\}$ est dans $\mathcal{A}$. C'est immédiat puisque $\mathbb{1}_{\{f\geq 1\}}$ est limite monotone de la suite $(f\wedge 1)^n$ d'éléments de $\mathcal{H}_0$.
vi) Montrons que $b(\mathcal{A}) \subset \mathcal{H}_0$. La proposition I.2.7 montre que toute fonction positive $\mathcal{A}$-mesurable est limite croissante de fonctions étagées $\mathcal{A}$-mesurables. Donc les fonctions positives bornées et $\mathcal{A}$-mesurables sont dans $\mathcal{H}_0$. En écrivant toute fonction comme la différence de sa partie positive et sa partie négative, l'inclusion $b(\mathcal{A}) \subset \mathcal{H}_0$ s'ensuit.
vii) Montrons enfin que $\mathcal{H}_0 \subset b(\mathcal{A})$. Cela découle du point v) et de l'inclusion $\mathcal{H}_0 \subset b\big(\sigma(\mathcal{H}_0)\big)$. $\square$

Nous pouvons à présent conclure la démonstration du théorème. L'inclusion $\mathcal{C} \subset \mathcal{H}_0$ donne $\sigma(\mathcal{C}) \subset \sigma(\mathcal{H}_0)$, et donc $b\big(\sigma(\mathcal{C})\big) \subset b\big(\sigma(\mathcal{H}_0)\big)$. Enfin, le lemme I.3.7 montre que $b\big(\sigma(\mathcal{H}_0)\big) = \mathcal{H}_0$. □

I.4. Mesures

Définition I.4.1. Soit $(\Omega, \mathcal{A})$ un espace mesurable.

Une application μ de $\mathcal{A}$ dans $\mathbb{R} \cup \{\infty\}$ est σ-additive si pour toute famille A_i d'éléments de $\mathcal{A}$, indexée par un ensemble I fini ou dénombrable, $\mu(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i)$.

On appelle mesure (positive) toute application μ de $\mathcal{A}$ dans $\mathbb{R}^+ \cup \{\infty\}$, σ-additive, et telle que $\mu(\emptyset) = 0$.

On dit qu'une mesure μ est σ-finie s'il existe une famille dénombrable $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, telle que $\Omega = \bigcup_{n\in\mathbb{N}} A_n$ et $\mu(A_n) < \infty$. Une telle famille est appelée une suite d'exhaustion de Ω.

Un espace mesurable muni d'une mesure est appelé espace mesuré. De plus, si $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité ou simplement une probabilité. Une mesure de probabilité est σ-finie.

On appelle mesure signée la différence de deux mesures (positives).

Exemples I.4.2. (i) Jeu de dé. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ les 6 faces possibles d'un dé, muni de la tribu des parties, $\mathcal{P}(\Omega)$. Vérifier que $\mu(A) = \mathrm{card}(A)/6$ est une probabilité. Remarquer que $\mu(A)$ représente bien la probabilité que A survienne : c'est le nombre de faces qui provoquent A, divisé par le nombre total de faces du dé.

(ii) Soit $(\Omega, \mathcal{A})$ un espace mesurable et $\omega \in \Omega$. L'application

$$\delta_\omega \,:\, A \in \mathcal{A} \mapsto \delta_\omega(A) = \mathbb{1}_A(\omega)$$

est une mesure de probabilité, appelée masse de Dirac en ω.

(iii) Sur un ensemble dénombrable Ω muni de la tribu de ses parties, la mesure $\sum_{\omega\in\Omega} \delta_\omega$ est appelée mesure de comptage. On remarquera que si $A \subset \Omega$, $\mu(A)$ est le cardinal de A, donc « compte » le nombre d'éléments de A.

La proposition suivante regroupe quelques propriétés importantes des mesures.

Proposition I.4.3. *Soit $(\Omega, \mathcal{A})$ un espace mesurable et $(A_i)_{i\in I}$, $I \subset \mathbb{N}$, une famille finie ou dénombrable d'ensembles mesurables.*

(i) Si $A_1 \subset A_2$, alors $\mu(A_1) \leq \mu(A_2)$ (i.e. μ est croissante).
(ii) $\mu(\bigcup_{i\in I} A_i) \leq \sum_{i\in I} \mu(A_i)$ (i.e. μ est sous-additive).
(iii) Si $A_i \subset A_{i+1}$ pour tout i, alors $\mu(\bigcup_i A_i) = \lim_{i\to\infty} \mu(A_i)$.
(iv) Si $A_i \supset A_{i+1}$ pour tout i et $\mu(A_{i_0}) < \infty$ pour un certain i_0, alors $\mu(\bigcap_i A_i) = \lim_{i\to\infty} \mu(A_i)$.

Démonstration. (i) A_2 est la réunion disjointe des ensembles mesurables A_1 et $A_2 \setminus A_1$, et l'axiome principal de la définition I.4.1 d'une mesure fournit $\mu(A_2) = \mu(A_1) + \mu(A_2 \setminus A_1) \geq \mu(A_1)$.
(ii) Si I est fini, par exemple $I = \{1, 2, 3, \ldots, n\}$, on procède par récurrence en remarquant que

$$\mu(A_1 \cup A_2) = \mu\big(A_1 \cup (A_2 \setminus A_1)\big) = \mu(A_1) + \mu(A_2 \setminus A_1) \leq \mu(A_1) + \mu(A_2)\,,$$

puisque A_1 et $A_2 \setminus A_1$ sont disjoints et $A_2 \setminus A_1 \subset A_2$. Si I est infini, on peut supposer $I = \mathbb{N}$. On a alors, pour tout $k \geq 0$,

$$\mu\Bigg(\bigcup_{0\leq i\leq k} A_i\Bigg) \leq \sum_{i\in\mathbb{N}} \mu(A_i)\,.$$

En considérant les ensembles croissants $B_k = \bigcup_{0\leq i\leq k} A_i$, le résultat se déduit de (iii) que nous montrons maintenant.
(iii) Soit $B_k = A_{k+1} \setminus A_k$, $k \in \mathbb{N}$. Les ensembles B_k sont disjoints, et comme $A_i = A_0 \cup \bigcup_{0\leq k\leq i-1} B_k$, $i \in \mathbb{N}$, on a

$$\bigcup_i A_i = A_0 \cup \bigcup_k B_k\,.$$

On utilise alors la σ-additivité pour obtenir

$$\begin{aligned}
\mu\Bigg(\bigcup_i A_i\Bigg) &= \mu(A_0) + \sum_{k\geq 0} \mu(B_k) \\
&= \mu(A_0) + \lim_{i\to\infty} \sum_{0\leq k\leq i-1} \mu(B_k) \\
&= \lim_{i\to\infty} \Bigg(\mu(A_0) + \sum_{0\leq k\leq i-1} \mu(B_k)\Bigg) \\
&= \lim_{i\to\infty} \mu(A_i)\,.
\end{aligned}$$

(iv) Soit i_0 tel que $\mu(A_{i_0}) < \infty$. Les $B_i = A_{i_0} \setminus A_i$, $i \geq i_0$ forment une suite croissante et $\mu(A_{i_0}) = \mu(B_i) + \mu(A_i) \geq \mu(B_i)$. Ainsi, la suite $\mu(B_i)$, $i \geq i_0$, est une suite croissante bornée. Donc la limite $\lim_{i\to\infty}\mu(B_i)$ existe et, la première égalité ci-dessous venant de (iii),

$$\lim_{i\to\infty}\mu(B_i)=\mu\Big(\bigcup_{i\geq i_0}B_i\Big)=\mu\Big(A_{i_0}\setminus\bigcap_{i\geq i_0}A_i\Big)=\mu(A_{i_0})-\mu\Big(\bigcap_{i\geq i_0}A_i\Big).$$

Donc

$$\mu\Big(\bigcap_{i\geq i_0}A_i\Big)=\mu(A_{i_0})-\lim_{i\to\infty}\mu(B_i)=\lim_{i\to\infty}\big(\mu(A_{i_0})-\mu(B_i)\big)=\lim_{i\to\infty}\mu(A_i),$$

ce qui démontre l'assertion. □

***Exemples* I.4.4.** (i) Soit $(\Omega,\mathcal{A},\mu)$ un espace mesuré et $A\in\mathcal{A}$. Alors μ_A définie par $\mu_A(B)=\mu(A\cap B)$ est une mesure sur $(\Omega,\mathcal{A})$. On peut bien sûr remplacer ici la tribu $\mathcal{A}$ par la tribu trace de $\mathcal{A}$ sur A composée des ensembles mesurables $A\cap B$, $B\in\mathcal{A}$.
(ii) Si μ est une mesure sur $(\Omega,\mathcal{A})$ et λ un réel positif, alors $\lambda\mu$ définie par $(\lambda\mu)(A)=\lambda\mu(A)$, $A\in\mathcal{A}$, est une mesure. En particulier si $0<\mu(A)<\infty$, alors $\mu_A(\cdot)/\mu(A)$ est une probabilité. C'est la probabilité conditionnelle sachant A (*cf.* chapitre VI).

Comme dans la section précédente où nous avons étudié le transport de structures par des applications, nous souhaitons pouvoir transporter une mesure d'un espace vers un autre.

***Définition* I.4.5.** Soit f une application mesurable d'un espace mesuré $(\Omega,\mathcal{A},\mu)$ dans un espace mesurable $(E,\mathcal{B})$. L'application μ^f de $\mathcal{B}$ dans $\mathbb{R}^+\cup\{\infty\}$ définie par $\mu^f(A)=\mu(f^{-1}(A))$ définit une mesure sur $(E,\mathcal{B})$, appelée mesure image de μ par f.

Nous laissons au lecteur le soin de vérifier que μ^f est bien une mesure. Elle est parfois notée $\mu\circ f^{-1}$.

***Exemple* I.4.6.** Considérons le jeu de dé avec $\Omega=\{1,2,3,4,5,6\}$ et μ la probabilité définie par $\mu(A)=\mathrm{card}(A)/6$. Soit $f:\Omega\to\{0,1\}$ définie par $f(\omega)=1$ si ω est pair, et 0 si ω est impair. On vérifie que

$$\mu^f(\{0\})=\mu^f(\{1\})=1/2,$$

i.e. on a une chance sur deux d'obtenir un chiffre pair en jouant au dé. Cet exemple montre simplement que le formalisme utilisé n'est pas absurde et coïncide avec l'intuition que l'on peut avoir du hasard.

La construction de mesures est un point délicat. En pratique, la tribu sur Ω peut ne pas être donnée de façon très explicite (penser par exemple à la tribu borélienne sur $\mathbb{R}$) et on ne peut pas vraiment définir une mesure en spécifiant sa valeur explicitement pour tout ensemble mesurable. Il est donc souhaitable d'avoir un procédé permettant de définir une mesure sur une partie de la tribu, puis de l'étendre de façon canonique à l'ensemble de la tribu. En particulier, cette extension doit être unique, ce qui ne peut être vrai que si la classe d'ensembles de départ est suffisamment riche. Un premier pas dans cette direction est donné par la proposition suivante, conséquence du théorème des classes monotones. Elle montre que deux mesures coïncident si elles coïncident sur une algèbre qui engendre la tribu.

Proposition I.4.7. *Soient μ, ν deux mesures sur $(\Omega, \mathcal{A})$. Soit $\mathcal{C}$ une algèbre qui engendre $\mathcal{A}$. Si μ et ν coïncident sur $\mathcal{C}$, alors elles sont égales.*

Démonstration. L'ensemble $\mathcal{M} = \{\, A \in \mathcal{A} : \mu(A) = \nu(A) \,\}$ est une classe monotone qui contient $\mathcal{C}$. Puisque $\mathcal{C}$ est stable par intersection finie, par le théorème I.3.3 des classes monotones, $\mathcal{A} \supset \mathcal{M} \supset \mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A}$ et ainsi $\mathcal{M} = \mathcal{A}$. Donc μ et ν coïncident bien partout. $\square$

Exemple I.4.8. Sur un espace produit, une mesure est déterminée par sa valeur sur les pavés (voir définition I.1.9 et I.1.10.i.)

Un deuxième pas nous est fourni par le théorème de prolongement de Carathéodory. Une application μ définie sur une algèbre de Boole $\mathcal{C}$ à valeurs dans $\mathbb{R}^+ \cup \{\, \infty \,\}$ est dite additive si $\mu(\emptyset) = 0$ et si $\mu\big(\bigcup_{1 \leq i \leq n} A_i\big) = \sum_{1 \leq i \leq n} \mu(A_i)$ pour tous $A_1, \ldots, A_n \in \mathcal{C}$ disjoints.

***Théorème I.4.9* (de prolongement).** *Si μ est une fonction additive d'ensembles, positive, définie sur une algèbre de Boole $\mathcal{C}$ de parties de Ω avec $\mu(\Omega) < \infty$, elle se prolonge de façon unique en une mesure sur $(\Omega, \sigma(\mathcal{C}))$.*

La démonstration de ce théorème est admise. On peut se référer à Neveu (1964, §I.5).

Exemples I.4.10. (i) Sur $\mathbb{R}$, les réunions finies d'intervalles forment une algèbre de Boole $\mathcal{C}$. Définissons $\mu(]\, a, b\,]) = b - a$ et prolongeons μ par additivé à $\mathcal{C}$. Précisément, si $A = \bigcup_{1 \leq i \leq n}]\, a_i, b_i\,]$ avec $]\, a_i, b_i\,]$ disjoints,

$$\mu(A) = \sum_{1 \leq i \leq n} (b_i - a_i)\,.$$

Alors μ est une fonction additive d'ensembles et s'étend à la tribu borélienne sur $\mathbb{R}$. On appelle mesure de Lebesgue cette extension. La mesure de Lebesgue d'un intervalle $[\,a, b\,]$, $]\,a, b\,[$ ou $[\,a, b\,[$ est sa longueur $b - a$. On pourrait faire une construction analogue à l'aide de ces différents intervalles. Plus généralement, pour une fonction croissante $F : I \to \mathbb{R}$ sur un intervalle I de $\mathbb{R}$, on peut poser $\mu_F(]\,a, b\,]) = F(b) - F(a)$; ceci définit la famille des mesures dites de Stieltjes.

(ii) Soient $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$ deux espaces mesurés. Sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ on définit la mesure produit par $\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1) \times \mu_2(A_2)$ pour $A_i \in \mathcal{A}_i$, $i = 1, 2$. Par linéarité, on étend la définition aux ensembles élémentaires au sens de la définition I.1.9. La mesure produit s'étend de manière unique à la tribu produit par le théorème I.4.9.

(iii) Si λ est la mesure de Lebesgue sur $\mathbb{R}$, on définit $\lambda \otimes \cdots \otimes \lambda$ sur $\mathbb{R}^d$ (mesure de Lebesgue sur $\mathbb{R}^d$). Alors $\lambda \otimes \cdots \otimes \lambda$ coincide avec l'extension de la fonction additive d'ensembles μ définie d'abord sur les pavés par

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

et étendue par additivité aux réunions finies de pavés.

La mesure de Lebesgue sur $\mathbb{R}$ ou $\mathbb{R}^d$ a la propriété importante d'être invariante par translation. En effet, si x est un vecteur de $\mathbb{R}^d$, A un borélien de $\mathbb{R}^d$, notons

$$\tau_x(A) = \{\, a + x \;:\; a \in A \,\}$$

le translaté de A par x. Montrons que pour tout $x \in \mathbb{R}^d$, et tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda \otimes \cdots \otimes \lambda\big(\tau_x(A)\big) = \lambda \otimes \cdots \otimes \lambda(A)\,.$$

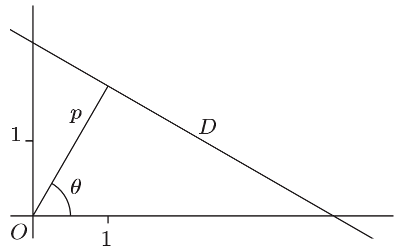
La classe

$$\mathcal{E} = \Big\{\, A =]\,a_1, b_1\,] \times \cdots \times]\,a_d, b_d\,] \subset \mathbb{R}^d \;:\; \forall x \in \mathbb{R}^d\,, \qquad \lambda \otimes \cdots \otimes \lambda\big(\tau_x(A)\big) = \lambda \otimes \cdots \otimes \lambda(A) \,\Big\}$$

est stable par intersection finie et contient tous les pavés. Donc $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}^d)$ d'après le théorème des classes monotones I.3.3 et la définition de $\mathcal{B}(\mathbb{R}^d)$ (exemple I.1.12.iii). On peut de plus démontrer que la mesure de Lebesgue est, à une constante de proportionnalité près, l'unique mesure invariante par translation sur $\mathbb{R}^d$.

(iv) Soit $\mathcal{D}$ l'ensemble des droites du plan. Nous allons construire sur $\mathcal{D}$ une mesure analogue à la mesure de Lebesgue sur $\mathbb{R}^d$, laquelle nous permettra de mesurer des ensembles de droites du plan.

Si D est une droite du plan, on la repèrera par ses coordonnées polaires. Ainsi, on notera $D = D(\theta, p)$ où p est la distance entre la droite D et l'origine O, θ est l'angle entre l'axe Ox et la droite perpendiculaire à D passant par O. On prendra θ dans $[\,0, 2\pi\,[$ et p positif ou nul.



Soit sur $[\,0, 2\pi\,[\times[\,0, \infty\,[$ la mesure de Lebesgue μ définie à partir des pavés par

$$\mu\big([\,\theta_1, \theta_2\,] \times [\,p_1, p_2\,]\big) = (\theta_1 - \theta_2)(p_1 - p_2)\,, \quad \begin{array}{l} 0 \leq \theta_1 \leq \theta_2 \leq 2\pi\,, \\ 0 \leq p_1 \leq p_2 < \infty\,. \end{array}$$

Cette mesure induit une mesure ν sur l'ensemble des droites par

$$\nu(A) = \mu\{\,(\theta, p)\ :\ D(\theta, p) \in A\,\}\,.$$

La mesure ν est donc la mesure image de μ par l'application $(\theta, p) \mapsto D(\theta, p)$. Considérons sur $\mathcal{D}$ l'ensemble des mouvements euclidiens, c'est-à-dire les compositions de translations τ_x et de rotations R_ϕ d'angle ϕ autour de l'origine. Notons $e(\theta)$ le vecteur de coordonnées $(\cos\theta, \sin\theta)$ dans $\mathbb{R}^2$, et notons $\langle x, y\rangle$ le produit scalaire de deux vecteurs $x, y \in \mathbb{R}^2$. Observons que

$$\tau_x\big(D(\theta, p)\big) = D(\theta', p')$$

avec

$$\theta' = \begin{cases} \theta & \text{si } p + \langle x, e(\theta)\rangle > 0 \\ \theta + \pi \quad \bmod 2\pi & \text{sinon} \end{cases}$$

et

$$p' = \begin{cases} p + \langle x, e(\theta)\rangle & \text{si } p + \langle x, e(\theta)\rangle > 0 \\ |p + \langle x, e(\theta)\rangle| & \text{sinon} \end{cases}\,.$$

On voit donc qu'une translation τ_x se traduit par un translation sur (θ, p). Puisque la mesure μ est invariante par translation modulo 2π, ν est invariante par τ_x. De même ν est invariante par toute rotation R_ϕ d'angle ϕ, puisque

$$R_\phi D(\theta, p) = D(\theta + \phi \bmod 2\pi, p)$$

induit encore une translation sur (θ, p).

On démontre que la mesure ν ainsi définie est, à une constante de proportionnalité près, l'unique mesure sur les ensembles de droites qui est invariante par les mouvements euclidiens.

Signalons enfin la définition suivante qui sera très utile pour la suite.

Définition I.4.11. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On dit qu'un ensemble A est négligeable s'il existe un ensemble $B \in \mathcal{A}$ tel que $A \subset B$ et $\mu(B) = 0$.

On dit qu'une fonction mesurable f sur $(\Omega, \mathcal{A}, \mu)$ vérifie une propriété $\mathcal{P}$ μ-presque partout (μ-p.p.) si l'ensemble $\{\, \omega : f(\omega) \text{ ne vérifie pas } \mathcal{P} \,\}$ est négligeable.

Exemples I.4.12. (i) Soit $\Omega = \{\, 1, 2, 3 \,\}$ muni de la tribu de ses parties et μ définie par $\mu(\{\, 1 \,\}) = \mu(\{\, 2 \,\}) = 1$ et $\mu(\{\, 3 \,\}) = 0$. Soit f la fonction mesurable $f(1) = f(2) = 1$ et $f(3) = 0$. Alors f est constante et égale à 1 μ-p.p.

(ii) Soit $\mathbb{R}$ muni de la tribu borélienne et de la mesure de Lebesgue λ. Soit f la fonction $f(\omega) = 1$ si ω est rationnel, et $f(\omega) = 0$ sinon (*i.e.* $f = \mathbb{1}_{\mathbb{Q}}$). Alors f est égal à 0 λ-p.p. En effet, $\lambda(\{\, x \,\}) = 0$ pour tout $x \in \mathbb{R}$ et puisque $\mathbb{Q}$ est dénombrable, $\lambda(\mathbb{Q}) = 0$. De même, la fonction signe, égale à 1 sur $]\, 0, \infty \,[$, valant 0 à l'origine et -1 sur $]\, -\infty, 0 \,[$, est continue λ-p.p. puisque son seul point de discontinuité est 0 et $\lambda(\{\, 0 \,\}) = 0$.

Exercices

Exercice I.1. Soit E une partie (fixée) d'un ensemble Ω, et soit

$$\mathcal{E} = \{\, A \in \mathcal{P}(\Omega) : A \subset E \,\} .$$

Déterminer l'algèbre de Boole engendrée par $\mathcal{E}$.

Exercice I.2. Si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont des tribus sur Ω, on pose

$$\begin{aligned} \mathcal{J} &= \{\, A_1 \cap A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \,\} , \\ \mathcal{U} &= \{\, A_1 \cup A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \,\} . \end{aligned}$$

Démontrer que $\sigma(\mathcal{J}) = \sigma(\mathcal{A}_1 \cup \mathcal{A}_2) = \sigma(\mathcal{U})$.

Exercice I.3. Soit $(\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2)$ un espace mesuré produit. Si $A \in \mathcal{A}$, montrer que pour tout $\omega_1 \in \Omega_1$, la section $A_{\omega_1} = \{\, \omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A \,\}$ est mesurable (élément de $\mathcal{A}_2$).
Indication : considérer $\mathcal{M} = \{\, A \in \mathcal{A} : A_{\omega_1} \in \mathcal{A}_2 \,\}$ et $\mathcal{E}$ la classe des unions finies de pavés. Montrer que $\mathcal{E} \subset \mathcal{M}$, que $\mathcal{M}$ est une classe monotone, puis conclure à l'aide du théorème des classes monotones.

Exercice I.4. Vérifier l'égalité $f^{-1}(U) = \bigcup_{r,m} \bigcap_{n \geq m} f_n^{-1}(U_r)$ de la démonstration du théorème I.2.5.

Exercice I.5. Si $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on note $\phi(x)$ le vecteur x ordonné par ordre croissant, i.e. $\phi(x) = (x_{1,n}, \ldots, x_{n,n})$, où $x_{1,n} = \min_{1 \leq i \leq n} x_i$ et

$$x_{i,n} = \min\bigl(\{\, x_i \, : \, 1 \leq i \leq n \,\} \setminus \{\, x_{j,n} \, : \, 1 \leq j \leq i-1 \,\}\bigr)\,, \quad 2 \leq i \leq n\,.$$

Montrer que ϕ est mesurable.
Indication : on pourra commencer par montrer que $x \mapsto x_{i,n}$ est mesurable pour tout $1 \leq i \leq n$ en considérant les ensembles $\{\, x_{i,n} \leq a \,\}$, $a \in \mathbb{R}$.

Exercice I.6. Un exemple d'ensemble non mesurable. Sur $\mathbb{R}$ on définit la relation d'équivalence $x \sim y$ si $x - y \in \mathbb{Q}$. En utilisant l'axiome du choix (si A est une fonction sur un ensemble I telle que $A(x) \neq \emptyset$ pour tout x de I, il existe une fonction f telle que $f(x) \in A(x)$ pour tout $x \in I$), construire un ensemble $A \in [\,0,1\,[$ qui contient exactement un point de chaque classe d'équivalence. Supposons A mesurable, et soit $\alpha = \lambda(A)$ sa mesure de Lebesgue. Montrer que si $r, s \in \mathbb{Q}$ et $r \neq s$, alors $(A+s) \cap (A+r) = \emptyset$, où $A + x = \{\, y + x \, : \, y \in A \,\}$, et que $\lambda(A+s) = \lambda(A)$. Remarquer que

$$1 = \lambda\bigl([\,0,1\,]\bigr) \leq \lambda\Biggl(\bigcup_{r \in \mathbb{Q} \cap]\,-1,1\,[} (A+r)\Biggr) \leq \lambda\bigl([\,-1,2\,]\bigr) = 3\,.$$

En utilisant la σ-additivité de λ, montrer que cette inégalité conduit d'une part à $\alpha = 0$, d'autre part à $\alpha > 0$. Conclure.

Exercice I.7. Théorème d'Egorov. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré tel que $\mu(\Omega)$ soit fini ; on considère des applications f, f_n, $n \in \mathbb{N}$, de Ω dans $\mathbb{R}$, telles que $f_n \to f$ μ-p.p., c'est-à-dire, telles que

$$\mu\bigl(\{\, \omega \, : \, f_n(\omega) \not\to f(\omega) \,\}\bigr) = 0\,.$$

a) Pour $n \in \mathbb{N}$ et $\varepsilon > 0$, soit $G_{n,\varepsilon} = \{\omega \in \Omega \, : \, |f_n(\omega) - f(\omega)| \geq \varepsilon\}$ et $E_{n,\varepsilon} = \bigcup_{m \geq n} G_{m,\varepsilon}$. Démontrer que pour tout $\varepsilon > 0$,

$$\mu\Bigl(\bigcap_n \bigcup_{m \geq n} G_{m,\varepsilon}\Bigr) = 0$$

et en déduire que $\lim_{n\to\infty} \mu(E_{n,\varepsilon}) = 0$.
c) Déduire de la question précédente que pour tous $\varepsilon, \delta > 0$, il existe $n_0 \in \mathbb{N}$ et $B_{\varepsilon,\delta} \in \mathcal{A}$ tels que $\mu(B_{\varepsilon,\delta}) \leq \delta$ et pour tout $\omega \in \Omega \setminus B_{\varepsilon,\delta}$ et tout $n \geq n_0$, $|f_n(\omega) - f(\omega)| \leq \varepsilon$.
d) Soit $\alpha > 0$; pour tout entier $p \geq 1$, on pose $\varepsilon_p = 1/p$, $\delta_p = \alpha/2^p$, $A_p = B_{\varepsilon_p,\delta_p}$ et $A = \bigcup_{p \geq 1} A_p$. Démontrer que $\mu(A) \leq \alpha$ et que $f_n \to f$ uniformément sur $\Omega \setminus A$.

Exercice I.8. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Une partie $N \subset \Omega$ est dite μ-négligeable si elle est contenue dans un ensemble mesurable A tel que $\mu(A) = 0$. La tribu $\mathcal{B}$ est dite complète pour μ si elle contient tous les ensembles négligeables.

Si $\mathcal{N}$ désigne l'ensemble des parties μ-négligeables, soit

$$\mathcal{A}_\mu = \{\, A \cup N\,;\, A \in \mathcal{A}\,,\, N \in \mathcal{N}\,\}\,.$$

Montrer que $\mathcal{A}_\mu$ est une tribu, appelée la tribu μ-complétée de $\mathcal{A}$.

Exercice I.9. Soient X et Y deux espaces topologiques munis respectivement des tribus boréliennes $\mathcal{B}_X$ et $\mathcal{B}_Y$, μ une mesure sur $\mathcal{B}_X$, et $f : X \to Y$ une fonction continue μ-p.p., c'est-à-dire telle que l'ensemble $N = \{x \in X \,:\, f \text{ discontinue en } x\}$ soit μ-négligeable. Démontrer que f est mesurable de $(X, \overline{\mathcal{B}}_X)$ dans $(Y, \mathcal{B}_Y)$ où $\overline{\mathcal{B}}_X$ est la tribu complétée de $\mathcal{B}_X$ par rapport à μ.

Indication : Pour tout ouvert O de Y, on construira un ouvert V de X tel que

$$f^{-1}(O) \cap (X \setminus N) \subset V \subset f^{-1}(O)\,,$$

et on montrera qu'alors $f^{-1}(O) = V \cup (f^{-1}(O) \cap N)$.

II

INTÉGRATION

Dans tout ce chapitre, nous considérons des fonctions d'un espace mesuré $(\Omega, \mathcal{A}, \mu)$ à valeurs dans $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$. Ces fonctions seront appelées boréliennes sur $(\Omega, \mathcal{A}, \mu)$.

Les théorèmes importants de la théorie de l'intégration sont le théorème de convergence monotone et le théorème de convergence dominée de Lebesgue. De façon grossière, on veut que si une suite croissante de fonctions positives f_n converge simplement, alors la limite des intégrales de f_n est l'intégrale de la limite des f_n.

Cette nécessité conduit naturellement à prendre une définition de l'intégrale utilisant l'approximation des fonctions par des limites croissantes. Mais l'on veut aussi que l'intégrale coïncide avec ce que l'intuition attend lorsque l'on intègre des fonctions étagées.

II.1. Intégrale de fonctions positives

Définition II.1.1. Si $A \in \mathcal{A}$, la fonction indicatrice de A, $f(\omega) = \mathbb{1}_A(\omega)$, est mesurable. Son intégrale par rapport à μ, notée $\int f \, \mathrm{d}\mu$ ou $\int_\Omega f(\omega) \, \mathrm{d}\mu(\omega)$, est définie comme étant $\mu(A)$.

Plus généralement, si $B \in \mathcal{A}$, l'intégrale de $f = \mathbb{1}_A$ sur B par rapport à μ, notée $\int_B f \, \mathrm{d}\mu$ ou $\int_B f(\omega) \, \mathrm{d}\mu(\omega)$, est définie par $\mu(A \cap B)$, ou, de façon équivalente, par $\int \mathbb{1}_B f \, \mathrm{d}\mu$.

En particulier, si $\mathcal{A} = \mathcal{B}(\mathbb{R})$, en prenant μ la mesure de Lebesgue et $A =]\, a, b\,]$, nous obtenons $\int \mathbb{1}_A \, \mathrm{d}\mu = b - a$. C'est donc la longueur de l'intervalle. Cette définition de l'intégrale coïncide, sur les intervalles, avec l'intégrale de Riemann, puisque, au sens de Riemann, $\int_{\mathbb{R}} \mathbb{1}_A(x)\,\mathrm{d}x = \int_a^b 1\,\mathrm{d}x = b-a$.

Maintenant, si $\mathcal{A} = \mathcal{P}(\mathbb{N})$, si μ est la mesure de comptage sur $\mathbb{N}$, et $A \subset \mathbb{N}$, alors $\int \mathbb{1}_A\,\mathrm{d}\mu = \mathrm{card}(A)$.

L'un des avantages de la définition de l'intégrale de Lebesgue est de traiter de la même façon des exemples aussi différents que la mesure de Lebesgue et la mesure de comptage.

Nous savons maintenant intégrer les fonctions indicatrices d'ensembles mesurables, ce qui est bien peu ! Nous souhaitons que l'intégrale soit linéaire (*i.e.* $\int(\alpha f + \beta g)\,\mathrm{d}\mu = \alpha\int f\,\mathrm{d}\mu + \beta\int g\,\mathrm{d}\mu$), ce qui conduit à étendre la définition par linéarité.

***Définition* II.1.2.** Si f est étagée positive, $f(\omega) = \sum_{1\leq i\leq n} a_i \mathbb{1}_{A_i}(\omega)$ avec les A_i mesurables disjoints, on pose

$$\int_B f\,\mathrm{d}\mu = \sum_{1\leq i\leq n} a_i\mu(A_i\cap B) = \sum_{1\leq i\leq n} a_i \int_B \mathbb{1}_{A_i}\,\mathrm{d}\mu\,.$$

Le lecteur peut alors vérifier que la valeur de l'intégrale $\int_B f\,\mathrm{d}\mu$ ne dépend pas de la décomposition de f en somme d'indicatrices et en déduire la linéarité de l'intégrale sur les fonctions étagées positives.

Nous pouvons étendre la définition de l'intégrale aux fonctions positives.

***Définition* II.1.3.** Soit f une fonction mesurable positive définie sur $(\Omega, \mathcal{A}, \mu)$. On définit et note son intégrale par rapport à μ sur l'ensemble mesurable B par

$$\int_B f\,\mathrm{d}\mu = \int_B f(\omega)\,\mathrm{d}\mu(\omega) = \sup\left\{\int_B g\,\mathrm{d}\mu\ :\ g \text{ étagée positive, } g\leq f\right\}.$$

L'intégrale sur Ω est notée $\int_\Omega f\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$.

Observons que l'intégrale d'une fonction positive peut être infinie.

Montrons que cette construction de l'intégrale d'une fonction positive réalise ce que l'on en attend.

***Proposition* II.1.4.** *(i) Si* $0\leq f\leq g$ *alors* $0\leq \int_B f\,\mathrm{d}\mu \leq \int_B g\,\mathrm{d}\mu$.

(ii) Si $A\subset B$ *et* $f\geq 0$, *alors* $\int_A f\,\mathrm{d}\mu \leq \int_B f\,\mathrm{d}\mu$.

(iii) Si $f \geq 0$ *et* $c \geq 0$*, alors* $\int_B cf \, \mathrm{d}\mu = c \int_B f \, \mathrm{d}\mu$.

(iv) $\int_B (f+g) \, \mathrm{d}\mu = \int_B f \, \mathrm{d}\mu + \int_B g \, \mathrm{d}\mu$.

(v) Si $f = 0$ *alors* $\int f \, \mathrm{d}\mu = 0$.

(vi) Si $\mu(B) = 0$*, alors* $\int_B f \, \mathrm{d}\mu = 0$.

(vii) Si $f \geq 0$*, alors* $\int_B f \, \mathrm{d}\mu = \int \mathbb{1}_B f \, \mathrm{d}\mu$.

(viii) Si $f \geq 0$ *et* $\int_B f \, \mathrm{d}\mu = 0$*, alors* $\mathbb{1}_B f = 0$ μ*-p.p.*

Ces propriétés sont encore vraies si les hypothèses sur f (et g) ont seulement lieu μ-presque partout.

Démonstration. Commencer par établir les assertions (i)–(vii) sur les fonctions étagées, puis passer au supremum pour les fonctions positives. Démontrons par exemple, suivant ce schéma, (iii). Remarquons que si $f = \sum_{1 \leq i \leq n} a_i \mathbb{1}_{A_i}$ alors

$$\int_B cf \, \mathrm{d}\mu = \sum_{1 \leq i \leq n} c a_i \mu(B \cap A_i) = c \sum_{1 \leq i \leq n} a_i \mu(B \cap A_i) = c \int_B f \, \mathrm{d}\mu \, .$$

Montrons maintenant (viii). Quitte à remplacer f par $\mathbb{1}_B f$, il suffit de montrer le résultat pour $B = \Omega$. Considérons la suite croissante d'ensembles

$$A_n = \{ \, \omega \in \Omega \, : \, f(\omega) > 1/n \, \} \, , \qquad n \geq 1 \, .$$

On vérifie que $\mathbb{1}_{A_n} \leq n^{-1} f$. Donc, d'après (i) et (iii),

$$\mu(A_n) = \int \mathbb{1}_{A_n} \, \mathrm{d}\mu \leq n \int f \, \mathrm{d}\mu = 0 \, .$$

L'égalité $\{ \, f > 0 \, \} = \bigcup_{n \geq 1} A_n$ et la proposition I.4.3.iii montrent alors que l'ensemble $\{\omega : f(\omega) > 0\}$ est de μ-mesure nulle. Puisque $f \geq 0$, on en déduit que $f = 0$ μ-p.p. □

II.2. Intégrale de fonctions quelconques et théorèmes de convergence

Le théorème suivant est d'un intérêt considérable. Ses nombreuses applications justifient pleinement l'intérêt de la théorie de l'intégrale de Lebesgue. Il permet en particulier d'étendre la définition de l'intégrale des fonctions positives aux fonctions de signe quelconque.

Théorème II.2.1 (de convergence monotone). *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite croissante de fonctions mesurables positives sur* $(\Omega, \mathcal{A}, \mu)$*, convergeant ponctuellement vers* f*. Alors* f *est mesurable et*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu \, .$$

Démonstration. La fonction f est mesurable en vertu du théorème I.2.5. Puisque f_n est croissante et positive, $\int f_n \, \mathrm{d}\mu$ est croissante et positive d'après II.1.4.i, donc admet une limite $\alpha \geq 0$ (éventuellement $\alpha = +\infty$). Puisque $f_n \leq f$, II.1.4.i montre aussi que $\alpha \leq \int f \, \mathrm{d}\mu$.

Soit une fonction étagée $g = \sum_{1\leq j\leq m} b_j \mathbb{1}_{B_j}$ telle que $0 \leq g \leq f$. Soit $0 \leq c < 1$. Notons $\{ f_n \geq cg\} = \{\omega \in \Omega : f_n(\omega) \geq cg(\omega) \}$. Alors,

$$\int f_n \, \mathrm{d}\mu \geq \int \mathbb{1}_{\{f_n \geq cg\}} f_n \, \mathrm{d}\mu \geq c \int g \, \mathbb{1}_{\{f_n\geq cg\}} \, \mathrm{d}\mu = c \sum_{1\leq j\leq m} b_j \mu\big(B_j \cap \{ f_n \geq cg \}\big)$$

d'après II.1.4.i–iii et la définition de l'intégrale sur les fonctions étagées. Donc, quand $n \to \infty$,

$$\alpha \geq c \sum_{1\leq i\leq m} b_j \lim_{n\to\infty} \mu\big(B_j \cap \{ f_n \geq cg \}\big) = c \sum_{1\leq i\leq m} b_j \mu(B_j) = c \int g \, \mathrm{d}\mu \, ,$$

la seconde égalité résultant de I.4.3.iii et du fait que $\bigcup_n \{ f_n \geq cg \} = \Omega$. Or c étant arbitraire dans $[\,0, 1\,[$, on obtient $\alpha \geq \int g \, \mathrm{d}\mu$, ceci pour toute fonction étagée $0 \leq g \leq f$. Par définition de l'intégrale $\int f \, \mathrm{d}\mu$, on en déduit $\alpha \geq \int f \, \mathrm{d}\mu$. Ainsi $\alpha = \int f \, \mathrm{d}\mu$ et le théorème est démontré. □

Le résultat suivant est un exemple d'application du théorème de convergence monotone.

Corollaire II.2.2. *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables positives et soit* $f = \sum_{n\in\mathbb{N}} f_n$*. Alors* $\int f \, \mathrm{d}\mu = \sum_{n\in\mathbb{N}} \int f_n \, \mathrm{d}\mu$.

Démonstration. La suite $g_n = \sum_{0\leq m\leq n} f_m$ est croissante et converge simplement vers f. Le théorème II.2.1 implique $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int \lim_{n\to\infty} g_n \, \mathrm{d}\mu$, ce qui est le résultat. □

Corollaire II.2.3 (Lemme de Fatou). *Soit* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables positives. Alors*

$$\int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu \, .$$

Démonstration. Soit $g_n = \inf_{m\geq n} f_m$. La suite g_n est croissante, converge simplement vers $\liminf_{n\to\infty} f_n$, et de plus $g_n \leq f_n$. Il suffit alors d'appliquer le théorème II.2.1 et d'utiliser la proposition II.1.4.i. $\square$

Nous étendons maintenant l'intégrale des fonctions positives aux fonctions de signe quelconque. Pour cela, si f est une fonction, on note $f^+ = f \vee 0$ sa partie positive et $f^- = -(f \wedge 0)$ sa partie négative, ce qui assure que $f = f^+ - f^-$ et $|f| = f^+ + f^-$. Les fonctions f^+ et f^- sont boréliennes si f l'est.

***Définition* II.2.4.** Soit $f = f^+ - f^-$ une fonction mesurable. On dit que f est μ-intégrable sur B si $\int_B |f|\,\mathrm{d}\mu < \infty$. Si f est μ-intégrable sur B, on définit et écrit son intégrale par rapport à μ sur B par $\int_B f\,\mathrm{d}\mu = \int_B f^+\,\mathrm{d}\mu - \int_B f^-\,\mathrm{d}\mu$.

Si $B = \Omega$, on dit que f est μ-intégrable, et note $\int f\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu$. Si le contexte est clair on dit aussi que f est intégrable.

Alors que l'intégrale d'une fonction positive est toujours définie, éventuellement infinie, nous convenons ici qu'une fonction de signe quelconque est intégrable si et seulement $\int |f|\,\mathrm{d}\mu < \infty$. Il est aisé de vérifier que si f est intégrable et $B \in \mathcal{A}$, alors $\int_B f\,\mathrm{d}\mu = \int \mathbb{1}_B f\,\mathrm{d}\mu$.

***Exemple* II.2.5.** On vérifie sans peine que si $(\Omega, \mathcal{A})$ est un espace mesurable et f est une fonction mesurable à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors pour tout $\omega \in \Omega$, f est intégrable par rapport à la masse de Dirac δ_ω (*cf.* I.4.2.ii) et $\int f\,\mathrm{d}\delta_\omega = f(\omega)$. Plus généralement, si $\mu = \sum_{1\leq i\leq n} a_i\delta_{\omega_i}$, $a_i \geq 0$, $\omega_i \in \Omega$, alors $\int f\,\mathrm{d}\mu = \sum_{1\leq i\leq n} a_i f(\omega_i)$.

Par convergence monotone, les propriétés essentielles de l'intégrale sur les fonctions positives s'étendent à l'intégrale des fonctions de signe quelconque.

***Proposition* II.2.6.** *Si f et g sont intégrables et si $\alpha, \beta \in \mathbb{R}$, alors*

$$\int (\alpha f + \beta g)\,\mathrm{d}\mu = \alpha \int f\,\mathrm{d}\mu + \beta \int g\,\mathrm{d}\mu\,.$$

De plus, si $f \leq g$ alors $\int f\,\mathrm{d}\mu \leq \int g\,\mathrm{d}\mu$.

Démonstration. Supposons d'abord $f, g \geq 0$ ainsi que $\alpha, \beta \geq 0$. D'après la proposition I.2.7, il existe des suites f_n et g_n, $n \in \mathbb{N}$, de fonctions étagées positives qui convergent en croissant vers f et g respectivement. Alors la suite $\alpha f_n + \beta g_n$ converge en croissant vers $\alpha f + \beta g$, et le résultat, dans ce cas, se déduit du théorème de convergence monotone. En général, on sépare parties positive et négative et on distingue selon les signes de α et β.

Si $f \geq g$, alors $f - g \geq 0$. Donc $\int (f-g)\,\mathrm{d}\mu \geq 0$ d'après la proposition II.1.4.i, et la conclusion s'ensuit par linéarité. □

L'énoncé suivant est une forme généralisée du lemme de Fatou.

Corollaire II.2.7. *Soit g une fonction intégrable et soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions intégrables.*
(i) Si $g \leq f_n$, alors $\int \liminf_{n\to\infty} f_n \,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu$.
(ii) Si $f_n \leq g$, alors $\limsup_{n\to\infty} \int f_n \,\mathrm{d}\mu \ \leq \int \limsup_{n\to\infty} f_n \,\mathrm{d}\mu$.

Démonstration. (i) D'après le lemme de Fatou (corollaire II.2.3), on a

$$\int \liminf_{n\to\infty} (f_n - g)\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int (f_n - g)\,\mathrm{d}\mu$$

ce qui démontre (i) par linéarité de l'intégrale.
(ii) De même, le corollaire II.2.3 donne

$$\int \liminf_{n\to\infty} (g - f_n)\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int (g - f_n)\,\mathrm{d}\mu\,.$$ □

Le résultat suivant est encore un corollaire du théorème de convergence monotone, mais nous lui attribuons la valeur d'un théorème compte tenu de son importance.

Théorème II.2.8 (de convergence dominée de Lebesgue). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions telles que $|f_n| \leq g$ où g est intégrable et f_n converge simplement vers f. Alors f est intégrable et*

$$\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu\,.$$

Démonstration. L'intégrabilité de f vient de ce que nécessairement $|f| \leq g$ et $|g| = g$ est intégrable. Puisque $\lim_{n\to\infty} f_n = f$, et $-g \leq f_n \leq g$, le corollaire II.2.7 fournit

$$\begin{aligned}
\limsup_{n\to\infty} \int f_n \,\mathrm{d}\mu &\leq \int \limsup_{n\to\infty} f_n \,\mathrm{d}\mu \\
&= \int f \,\mathrm{d}\mu \\
&= \int \liminf_{n\to\infty} f_n \,\mathrm{d}\mu \\
&\leq \liminf_{n\to\infty} \int f_n \,\mathrm{d}\mu\,.
\end{aligned}$$ □

On peut vérifier simplement que, dans les théorèmes de convergence monotone et de convergence dominée, la convergence simple de f_n vers f peut être remplacée par la convergence μ-presque partout (*i.e.* $f_n(\omega) \to f(\omega)$ pour tout ω en dehors d'un ensemble de mesure nulle pour μ).

Exemple II.2.9. Soit A_i, $i \geq 1$, des ensembles mesurables disjoints, et soit g une fonction intégrable. Alors $\int_{\cup_{i\geq 1} A_i} g \,\mathrm{d}\mu = \sum_{i\geq 1} \int_{A_i} g \,\mathrm{d}\mu$. En effet, la suite de fonctions $f_n = \mathbb{1}_{\cup_{1\leq i\leq n} A_i} g$ converge simplement vers $f = \mathbb{1}_{\cup_{i\geq 1} A_i} g$ et chaque fonction $|f_n|$ est majorée par la fonction intégrable $|g|$. Le théorème de convergence dominée II.2.8 montre que

$$\int_{\cup_{i\geq 1} A_i} g \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \lim_{n\to\infty} \sum_{1\leq i\leq n} \int_{A_i} g \,\mathrm{d}\mu \,.$$

L'inégalité suivante est très importante pour les applications. Elle ne concerne que les mesures de probabilité.

***Théorème II.2.10* (Inégalité de Jensen).** *Si ϕ est une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$, et si f est une fonction borélienne telle que f et $\phi(f)$ sont intégrables par rapport à une mesure de probabilité μ, alors*

$$\phi\Big(\int f \,\mathrm{d}\mu\Big) \leq \int \phi(f) \,\mathrm{d}\mu \,.$$

Démonstration. La convexité de ϕ assure qu'en tout point le graphe de ϕ est au-dessus de sa tangente. Donc pour tout $t \in \mathbb{R}$, il existe β tel que $\phi(x) \geq \phi(t) + \beta(x - t)$ pour tout x (on peut prendre pour β la dérivée à gauche ou à droite de ϕ en t). Appliquons cette inégalité à $t = \int f \,\mathrm{d}\mu$ et $x = f(\omega)$ pour tout ω, et intégrons les deux membres. La conclusion s'ensuit puisque l'intégrale conserve le sens des inégalités (*cf.* proposition II.2.6). □

La démonstration de l'inégalité de Jensen montre que si ϕ est strictement convexe, l'égalité $\phi(\int f \,\mathrm{d}\mu) = \int \phi(f) \,\mathrm{d}\mu$ n'a lieu que si f est μ-presque partout constante. De plus, si l'égalité a lieu pour toute fonction f, alors la fonction ϕ est linéaire.

Dans le cas de la mesure de Lebesgue, mentionnons que l'intégrale construite étend celle de Riemann et donc qu'en pratique le calcul d'une intégrale s'effectue à l'aide des techniques usuelles (calcul de primitives, changement de variables, intégration par parties). En effet, rappelons qu'une fonction f est Riemann intégrable sur un intervalle fermé borné $[\,a, b\,]$ si pour tout $\varepsilon > 0$ il existe des fonctions en escalier g et h avec $g \leq f \leq h$ et $\int_a^b (h-g)(x)\,\mathrm{d}x \leq \varepsilon$. Soit λ la mesure de Lebesgue (*cf.* I.4.10.i). Pour les fonctions en escalier g et h, par définition des intégrales de Riemann et de Lebesgue, $\int_a^b g(x)\,\mathrm{d}x = \int_{[a,b]} g\,\mathrm{d}\lambda$ et $\int_a^b h(x)\,\mathrm{d}x = \int_{[a,b]} h\,\mathrm{d}\lambda$. Donc

$$\int_a^b g(x)\,\mathrm{d}x = \int_{[a,b]} g\,\mathrm{d}\lambda \leq \int_{[a,b]} f\,\mathrm{d}\lambda \leq \int_{[a,b]} h\,\mathrm{d}\lambda = \int_a^b h(x)\,\mathrm{d}x$$

et ainsi $\int_{[a,b]} f\,\mathrm{d}\lambda = \int_a^b f(x)\,\mathrm{d}x$. L'abus usuel de notation confond alors les notations $\mathrm{d}\lambda$ et $\mathrm{d}x$.

La construction de l'intégrale donnée ici est plus générale, d'une part parce que l'on peut intégrer par rapport à d'autres mesures que celle de Lebesgue et sur d'autres espaces que $\mathbb{R}$ ou $\mathbb{R}^d$, d'autre part, parce que même sur $\mathbb{R}$, il existe des fonctions Lebesgue-intégrables (*i.e.* intégrables au sens de ce chapitre) qui ne sont pas Riemann intégrables (*cf.* exercice II.1).

II.3. Théorème de Radon-Nikodym

Étant donnée une mesure sur un espace, nous avons vu que l'on peut éventuellement construire de nouvelles mesures en considérant les mesures images (*cf.* I.4.5). Nous considérons ici une nouvelle façon d'engendrer des mesures. Intuitivement, une barre d'un matériau homogène a un poids proportionnel à sa longueur, et la masse des intervalles de cette barre définit une mesure proportionnelle à la mesure de Lebesgue. La proposition ci-dessous revient à peu près à considérer la masse d'une barre non homogène.

Dans ce paragraphe, toutes les mesures sont supposées σ-finies.

Proposition II.3.1. *Soit f une fonction mesurable, positive et intégrable sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$. La fonction d'ensembles ν, définie sur $\mathcal{A}$ par $\nu(A) = \int_A f\,\mathrm{d}\mu$, $A \in \mathcal{A}$, est une mesure. De plus, si $\mu(A) = 0$, alors $\nu(A) = 0$. Enfin, si g est positive, $\int g\,\mathrm{d}\nu = \int gf\,\mathrm{d}\mu$. De plus, g est ν-intégrable si et seulement si gf est μ-intégrable et dans ce cas, $\int g\,\mathrm{d}\nu = \int gf\,\mathrm{d}\mu$.*

Démonstration. L'exemple II.2.9 montre que la fonction d'ensembles ν est σ-additive : si les A_i, $i \geq 1$, sont mesurables et disjoints,

$$\sum_{i\geq 1} \nu(A_i) = \sum_{i\geq 1} \int_{A_i} f\,\mathrm{d}\mu = \int_{\bigcup_{i\geq 1} A_i} f\,\mathrm{d}\mu = \nu\Big(\bigcup_{i\geq 1} A_i\Big).$$

Si $\mu(A) = 0$, la proposition II.1.4.v montre que $\nu(A) = \int_A f\,\mathrm{d}\mu = 0$.

Le dernier point résulte du schéma général de construction de l'intégrale. Si g est étagée positive, il est clair par construction que $\int g \, d\nu = \int gf \, d\mu$. Si g est positive, on conclut en l'approximant par des fonctions étagées. Enfin, pour g intégrable, on conclut en séparant parties positive et négative, et en utilisant la construction de l'intégrale. $\square$

La proposition précédente décrit en fait un phénomène général.

Définition II.3.2. Soient deux mesures μ et ν définies sur un espace mesurable $(\Omega, \mathcal{A})$.

(i) On dit que ν est absolument continue par rapport à μ si tout ensemble de mesure nulle pour μ est de mesure nulle pour ν (*i.e.* $\mu(A) = 0 \Rightarrow \nu(A) = 0$). On note alors $\nu \ll \mu$. De plus, si $\nu \ll \mu$ et $\mu \ll \nu$, on dit que ν et μ sont équivalentes.

(ii) On dit que μ et ν sont étrangères s'il existe $A \in \mathcal{A}$ tel que $\mu(A) = 0$ et $\nu(A^c) = 0$.

La proposition II.3.1 founit une classe de mesures absolument continues par rapport à μ. Le théorème suivant, appelé théorème de Radon-Nikodym, montre que c'est la situation générale.

***Théorème II.3.3* (de Radon-Nikodym).** *Si μ et ν sont deux mesures σ-finies, telles que $\nu \ll \mu$, alors il existe une fonction mesurable positive f telle que $\nu(A) = \int_A f \, d\mu$ pour tout ensemble mesurable A. La fonction f est appelée la densité de ν par rapport à μ, notée $f = \frac{d\nu}{d\mu}$.*

Démonstration. Voir par exemple Neveu (1964), Rudin (1975) ou Malliavin (1982). $\square$

Théorème II.3.4. *Soient μ, ν deux mesures σ-finies sur $(\Omega, \mathcal{A})$. Alors il existe une unique décomposition $\nu = \nu_{ac} + \nu_\perp$ avec $\nu_{ac} \ll \mu$ et $\nu_\perp$ étrangère par rapport à μ. La mesure ν_{ac} est appelée la partie absolument continue de ν par rapport à μ et $\nu_\perp$ la partie étrangère de ν par rapport à μ.*

Démonstration. On se reportera par exemple à Neveu (1964), Rudin (1975) ou Malliavin (1982). $\square$

II.4. Intégration par rapport à une mesure image

Nous avons défini les mesures images (*cf.* I.4.5). Rappelons que si f est une fonction mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(E, \mathcal{B})$, on note μ^f la mesure sur $\mathcal{B}$ définie par $\mu^f(B) = \mu(f^{-1}(B))$. En pratique, l'intégrale par rapport à une mesure image s'effectue à l'aide du théorème suivant, qui est une formulation abstraite de la formule classique du changement de variable.

***Théorème II.4.1* (de transport).** *Soit f une application mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(E, \mathcal{B})$, et soit ϕ une fonction borélienne de E dans $\mathbb{R}$. Alors si ϕ est à valeurs positives,*

$$\int_E \phi \, \mathrm{d}\mu^f = \int_\Omega \phi \circ f \, \mathrm{d}\mu \, .$$

Si ϕ est à valeurs quelconques, ϕ est μ^f-intégrable si et seulement si $\phi \circ f$ est μ-intégrable et, dans ce cas, l'identité précédente est encore satisfaite.

Démonstration. Elle répète le schéma général de construction de l'intégrale. Si $\phi = \mathbb{1}_B$ pour un $B \in \mathcal{B}$,

$$\begin{aligned}\int_E \mathbb{1}_B \, \mathrm{d}\mu^f = \mu^f(B) = \mu\big(f^{-1}(B)\big) &= \mu\big(\{\, \omega \in \Omega : f(\omega) \in B \,\}\big) \\ &= \int_\Omega \mathbb{1}_B \circ f(\omega) \, \mathrm{d}\mu(\omega)\end{aligned}$$

et la formule est vraie dans ce cas. Si ϕ est étagée, la formule est valide par linéarité (par rapport à ϕ). Si ϕ est positive, soit $(\phi_n)_{n \in \mathbb{N}}$ une suite de fonctions étagées positives convergeant en croissant vers ϕ (I.2.7). Alors $\phi \circ f_n$ est étagée et converge simplement en croissant vers $\phi \circ f$. En utilisant le théorème de convergence monotone (II.2.1), à la fois pour la mesure μ et la mesure μ^f,

$$\int_E \phi \, \mathrm{d}\mu^f = \lim_{n \to \infty} \int_E \phi_n \, \mathrm{d}\mu^f = \lim_{n \to \infty} \int_\Omega \phi_n \circ f \, \mathrm{d}\mu = \int_\Omega \phi \circ f \, \mathrm{d}\mu \, .$$

Dans le cas général, remarquons que

$$\int_E |\phi| \, \mathrm{d}\mu^f = \int_\Omega |\phi \circ f| \, \mathrm{d}\mu$$

et donc $\phi \circ f$ est μ-intégrable si et seulement si ϕ est μ^f intégrable. En posant $\phi = \phi^+ - \phi^-$, on conclut que

$$\begin{aligned}\int_E \phi \, \mathrm{d}\mu^f &= \int_E \phi^+ \, \mathrm{d}\mu^f - \int_E \phi^- \, \mathrm{d}\mu^f \\ &= \int_\Omega \phi^+ \circ f \, \mathrm{d}\mu - \int_\Omega \phi^- \circ f \, \mathrm{d}\mu = \int_\Omega \phi \circ f \, \mathrm{d}\mu \, .\end{aligned}$$

Le théorème est établi. □

Exemple II.4.2. Continuons l'exemple I.4.10.iv en évaluant la mesure de l'ensemble des droites qui coupent un segment S de longueur l donnée, *i.e.*

$$\nu\big(\{\, D \,:\, D \cap S \neq \emptyset \,\}\big)\,.$$

Puisque ν est invariante par les mouvements euclidiens, ou peut supposer que S est sur l'axe Ox, l'une de ses extrémités étant l'origine. Alors

$$\begin{aligned}
\nu\big(\{\, D \,:\, D \cap S \neq \emptyset \,\}\big) &= \int \mathbb{1}_{\{D\cap S\neq\emptyset\}} \,\mathrm{d}\nu(D) \quad \text{(définition II.1.1 de l'intégrale)}\\
&= \int_0^{2\pi}\int_0^{\infty} \mathbb{1}_{\{D(\theta,p)\cap S\neq\emptyset\}} \,\mathrm{d}p\,\mathrm{d}\theta \qquad \text{(par transport)}\\
&= 2\int_0^{\pi/2}\int_0^{\infty} \mathbb{1}_{\{0\leq p\leq l\cos\theta\}} \,\mathrm{d}p\,\mathrm{d}\theta\\
&= 2l\,.
\end{aligned}$$

Supposons maintenant que nous disposons de k segments $S_1, \ldots, S_k$, et posons $S = \bigcup_{1\leq i\leq k} S_i$. Soit $\mathrm{card}(D \cap S)$ le nombre de points d'intersection de D avec S. Alors

$$\begin{aligned}
\frac{1}{2}\int \mathrm{card}(D\cap S)\,\mathrm{d}\nu(D) &= \frac{1}{2}\int \sum_{1\leq i\leq k} \mathbb{1}_{D\cap S_i}\,\mathrm{d}\nu(D)\\
&= \frac{1}{2}\sum_{1\leq i\leq k} \nu\big(\{\, D \,:\, D\cap S_i \neq \emptyset \,\}\big)
\end{aligned}$$

est la somme des longueurs des segments.

Enfin, si nous avons une courbe $\mathcal{C}$ régulière (*i.e.* une courbe paramétrée $\alpha(t) = (x(t), y(t))$, $0 \leq t \leq 1$, avec x, y dérivables), celle-ci peut s'approximer par des segments. En utilisant le théorème de convergence dominée, on peut montrer que

$$\frac{1}{2}\int \mathrm{card}(D\cap \mathcal{C})\,\mathrm{d}\nu(D) = \text{longueur de } \mathcal{C}\,.$$

Ce résultat, loin d'être une abstraction, est tout à fait concret et utile. Il permet par exemple de mesurer la longueur d'une chaîne plane de molécules. Pour cela, il suffit de remarquer que

$$\frac{1}{2}\int \mathrm{card}(D\cap \mathcal{C})\,\mathrm{d}\nu(D) = \frac{1}{2}\int_0^{2\pi}\int_0^{c} \mathrm{card}\big(D(\theta,p)\cap\mathcal{C}\big)\,\mathrm{d}p\,\mathrm{d}\theta\,,$$

où c est un majorant du diamètre de $\mathcal{C}$ (*i.e.* de $\sup_{0\leq s,t\leq 1} |\alpha(s) - \alpha(t)|$) et en supposant que $\mathcal{C}$ est translaté de sorte que $\alpha(0) = 0$ par exemple. En approximant l'intégrale par une somme de Riemann, on voit que pour une subdivision

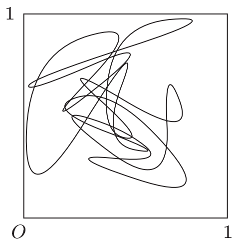
$0 = p_0 < p_1 < \cdots < p_n$ de $[0, c]$ et $0 = \theta_0 < \theta_1 < \cdots < \theta_m = 2\pi$

$$\frac{1}{2}\int \operatorname{card}(D \cap \mathcal{C})\,\mathrm{d}\nu(D) \simeq \frac{1}{2}\sum_{1\le i\le m}(\theta_i - \theta_{i-1})\sum_{1\le j\le n}(p_j - p_{j-1})\operatorname{card}\bigl(D(\theta_i, p_j)\cap\mathcal{C}\bigr)\,.$$

Lorsque $p_j - p_{j-1} = \varepsilon$ est constant, observons que

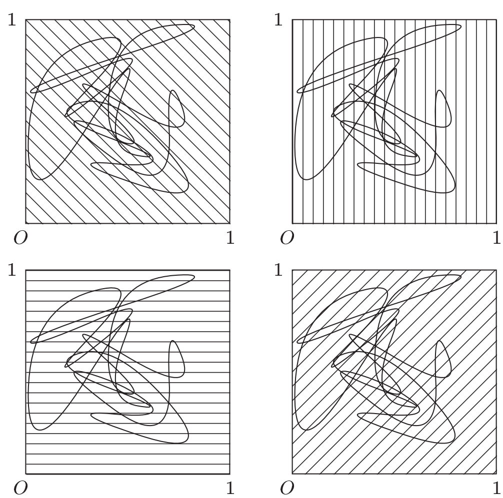
$$\sum_{1\le j\le n}(p_j - p_{j-1})\operatorname{card}(D(\theta_i, p_j)\cap\mathcal{C})$$

est ε fois le nombre d'intersections entre $\mathcal{C}$ et un réseau de droites parallèles équidistantes de ε. On se rend compte assez facilement sur un dessin que pour une courbe assez tortueuse, $m = 8$ (et $\theta_i - \theta_{i-1} = \pi/4$) donnera une très bonne approximation pourvu que c soit assez petit. Implémentons ceci pour mesurer la longueur de la courbe fermée ci-dessous (dans le carré unité).



Prenons $p_j - p_{j-1} = 1/20$ ce qui conduit à considérer un réseau de droites espacées de $0,05$. Prenons aussi $\theta_i - \theta_{i-1} = \pi/4$. Alors la longueur de la courbe est à peu près

$$l \simeq \frac{1}{2}\frac{\pi}{4}\frac{132 + 74 + 146 + 114}{20} \simeq 9{,}1\,.$$



II.5. Théorèmes de Fubini-Tonelli

Soient $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ deux espaces mesurés pour des mesures μ_1 et μ_2 σ-finies. Considérons l'espace produit $\Omega = \Omega_1 \times \Omega_2$ muni de la tribu produit $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ et de la mesure produit $\mu = \mu_1 \otimes \mu_2$ (*cf.* I.4.10.ii). Si $A \in \mathcal{A}$, on peut montrer (exercice I.3) que les sections

$$A_{\omega_1} = \big\{\, \omega_2 \in \Omega_2 \,:\, (\omega_1, \omega_2) \in A \,\big\} \quad \text{et} \quad A_{\omega_2} = \big\{\, \omega_1 \in \Omega_1 \,:\, (\omega_1, \omega_2) \in A \,\big\}$$

sont mesurables ($i.e.$ $A_{\omega_1} \in \mathcal{A}_2$ et $A_{\omega_2} \in \mathcal{A}_1$). De plus

$$\mu(A) = \int_{\Omega_1} \mu_2(A_{\omega_1})\, \mathrm{d}\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(A_{\omega_2})\, \mathrm{d}\mu_2(\omega_2)\,, \tag{1}$$

ce qui peut être réécrit

$$\begin{aligned} \mu(A) &= \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_A\big((\omega_1, \omega_2)\big)\, \mathrm{d}\mu_2(\omega_2)\, \mathrm{d}\mu_1(\omega_1) \\ &= \int_{\Omega_2} \int_{\Omega_1} \mathbb{1}_A\big((\omega_1, \omega_2)\big)\, \mathrm{d}\mu_1(\omega_1)\, \mathrm{d}\mu_2(\omega_2)\,. \end{aligned}$$

En effet, soit

$$\mathcal{M} = \Big\{\, A \in \mathcal{A} \,:\, \mu(A) = \int_{\Omega_1} \mu_2(A_{\omega_1})\, \mathrm{d}\mu(\omega_1) = \int_{\Omega_2} \mu_1(A_{\omega_2})\, \mathrm{d}\mu(\omega_2) \,\Big\}$$

et soit $\mathcal{E}$ la classe des pavés de la forme $A_1 \times A_2$, $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$. On vérifie facilement que $\mathcal{M}$ est une classe monotone qui contient $\mathcal{E}$. Donc, en utilisant le théorème des classes monotones I.3.3 et la définition I.1.9 de la tribu produit, $\mathcal{A} \supset \mathcal{M} \supset \mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}$, ce qui démontre (1).

Autrement dit, on peut dans ce cas permuter l'ordre d'intégration. La construction de l'intégrale montre que cette permutation reste licite lorsqu'on intègre des fonctions à valeurs positives. C'est le théorème de Fubini-Tonelli. Le théorème de Fubini étend ce fait aux fonctions μ-intégrables.

Théorème II.5.1 **(de Fubini).** *Soit f une fonction réelle, définie sur Ω, $\mathcal{A}$-mesurable et μ-intégrable. Alors,*

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

Démonstration. D'après la remarque précédant le théorème, les égalités sont satisfaites lorsque f est positive. On étend les égalités aux fonctions intégrables en séparant parties positive et négative et en utilisant la linéarité de l'intégrale. $\square$

Dans la pratique, pour vérifier qu'une fonction f est intégrable par rapport à une mesure produit $\mu = \mu_1 \otimes \mu_2$, on évalue $\int |f| \,\mathrm{d}\mu$ par permutation des intégrales. Si $\int |f| \,\mathrm{d}\mu$ est finie, on est alors en droit d'utiliser le théorème de Fubini pour le calcul de $\int f \,\mathrm{d}\mu$. Des exemples simples montrent en outre que la permutation de l'ordre d'intégration peut être en défaut si f n'est pas intégrable pour la mesure produit μ.

II.6. Espaces L^p

Nous avons défini la classe des fonctions intégrables (à valeurs dans $\mathbb{R}$) sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$, qu'il est d'usage de noter $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. Pour $0 < p < \infty$, on note $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ (ou simplement $\mathcal{L}^p$ si le contexte est clair) l'ensemble des fonctions réelles de puissance p-ième intégrable, c'est-à-dire l'ensemble des fonctions f mesurables de Ω dans $\mathbb{R}$, telles que $\int |f|^p \,\mathrm{d}\mu < \infty$.

$\mathcal{L}^0$ est défini simplement comme étant l'ensemble des fonctions mesurables de $(\Omega, \mathcal{A}, \mu)$ dans $\mathbb{R}$.

On définit $\mathcal{L}^\infty$ comme étant l'ensemble des fonctions mesurables f de $(\Omega, \mathcal{A}, \mu)$ dans $\mathbb{R}$ telles qu'il existe $c > 0$ avec

$$\mu(\{\, \omega : |f(\omega)| > c \,\}) = 0 \,.$$

C'est l'ensemble des fonctions (mesurables) μ-essentiellement bornées.

Si $f \in \mathcal{L}^p$, $0 < p < \infty$, on pose

$$\|f\|_p = \left(\int_\Omega |f|^p \,\mathrm{d}\mu \right)^{1/p} .$$

Pour $f \in \mathcal{L}^\infty$, on pose

$$\|f\|_\infty = \inf \Big\{\, c > 0 : \mu\{\omega : |f(\omega)| > c \,\} = 0 \Big\} \,,$$

qui est appelé le supremum essentiel, ou la borne essentielle, de f.

Définition II.6.1. Deux réels $p, q \geq 1$ sont conjugués si $p^{-1} + q^{-1} = 1$. On convient que 1 et ∞ sont conjugués.

Théorème II.6.2 **(Inégalité de Hölder).** *Soient p et q conjugués, $1 \leq p \leq \infty$, et $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$. Alors $fg \in \mathcal{L}^1$ et*

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

Démonstration. Si $p = 1$ ou $p = \infty$ l'inégalité est évidente. Si $\|f\|_p \|g\|_q = 0$, alors $fg = 0$ μ-p.p. et l'inégalité de Hölder est triviale. Supposons donc $\|f\|_p \|g\|_q \neq 0$. Par homogénéité, on peut supposer que $\|g\|_q = 1$. Il suffit donc de démontrer que

$$\Big(\int |fg| \, \mathrm{d}\mu \Big)^p \leq \int |f|^p \, \mathrm{d}\mu \,.$$

Soit alors la mesure de probabilité ν de densité $|g|^q$ par rapport μ. L'inégalité à établir devient alors

$$\Big(\int |f| |g|^{1-q} \, \mathrm{d}\nu \Big)^p \leq \int |f|^p |g|^{-q} \, \mathrm{d}\nu$$

qui est une conséquence de l'inégalité de Jensen pour la fonction convexe $\phi(x) = x^p$ (théorème II.2.10), car $(1 - q)p = -q$. (En toute rigueur, afin d'assurer l'hypothèse d'intégrabilité, il conviendrait de travailler avec $|f| \, |g|^{1-q} \wedge n$, $n \geq 1$, en lieu et place de $|f| \, |g|^{1-q}$, et de conclure avec le théorème de convergence monotone.) $\square$

Théorème II.6.3 **(Inégalité de Minkowski).** *Soit $p \geq 1$. Si f, g sont dans $\mathcal{L}^p$, alors $f + g$ est aussi dans $\mathcal{L}^p$ et*

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

Démonstration. Remarquons que p et $p/(p-1)$ sont conjugués. En utilisant l'inégalité triangulaire puis l'inégalité de Hölder,

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g|^p \, \mathrm{d}\mu \\
&\leq \int \big(|f| |f + g|^{p-1} + |g| |f + g|^{p-1} \big) \, \mathrm{d}\mu \\
&\leq \big(\|f\|_p + \|g\|_p \big) \big\| \, |f + g|^{p-1} \big\|_{p/(p-1)} \\
&= \big(\|f\|_p + \|g\|_p \big) \, \|f + g\|_p^{p-1} \,.
\end{aligned}$$

C'est le résultat si $\|f + g\|_p \neq 0$. L'inégalité est triviale si $\|f + g\|_p = 0$. $\square$

De l'inégalité de Minkowski, on déduit que $\|\cdot\|_p$ est une semi-norme sur l'espace $\mathcal{L}^p$ (en effet $\|f\|_p = 0$ n'implique pas $f = 0$ mais seulement $f = 0$ μ-p.p. *cf.* II.1.4.viii). Notons $\mathrm{L}^p = \mathrm{L}^p(\Omega, \mathcal{A}, \mu)$ le quotient de l'espace $\mathcal{L}^p$ par la relation d'équivalence $f = g$ μ-p.p. Autrement dit, un élément f de L^p s'identifie à un représentant de la classe de tous les éléments g de $\mathcal{L}^p$ tels que $f = g$ μ-p.p. Alors $(\mathrm{L}^p, \|\cdot\|_p)$ est un espace vectoriel normé.

Pour les énoncés suivants nous supposerons que la mesure μ est σ-finie.

Théorème II.6.4. *Pour tout $p \geq 1$, l'espace L^p est complet.*

Nous démontrerons ce théorème à la fin de la section V.3.

Théorème II.6.5. *Pour p et q conjugués et $1 \leq p < \infty$, le dual de l'espace L^p est L^q. En d'autres termes, les formes linéaires continues sur L^p sont les fonctions de la forme $f \in \mathrm{L}^p \mapsto \int fg\,\mathrm{d}\mu \in \mathbb{R}$ pour $g \in \mathrm{L}^q$. La norme d'une telle forme linéaire est donnée par l'égalité $\|f\|_p = \sup\{\int fg\,\mathrm{d}\mu : \|g\|_q \leq 1\}$.*

Démonstration esquissée. Si $g \in \mathrm{L}^q$, l'application $f \mapsto \int fg\,\mathrm{d}\mu$ définie sur L^p est linéaire et continue d'après l'inégalité de Hölder. Il convient donc de montrer que toute forme linéaire continue sur L^p est nécessairement de cette forme. Soit Λ une telle forme linéaire, et posons $\nu(A) = \Lambda(\mathbb{1}_A)$. On vérifie que ν est additive (*i.e.* $\nu(A \cup B) = \nu(A) + \nu(B)$ si $A \cap B = \emptyset$), et même est une mesure. Si $\mu(A) = 0$, alors $\mathbb{1}_A = 0$ (dans L^p) et $\nu(A) = 0$, ce qui montre que ν est absolument continue par rapport à μ. D'après le théorème de Radon-Nikodym II.3.3, on a donc $\nu(A) = \Lambda(\mathbb{1}_A) = \int g\mathbb{1}_A\,\mathrm{d}\mu$ pour $g = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Par linéarité, $\Lambda(f) = \int fg\,\mathrm{d}\mu$ sur L^∞. Pour montrer que $g \in \mathrm{L}^q$, écrivons $g = h|g|$ où $|h| = 1$ et h est mesurable. Alors

$$\begin{aligned}
\int |g|^q \mathbb{1}_{[0,n]}(|g|)\,\mathrm{d}\mu &= \int |g|^{q-1}\mathbb{1}_{[0,n]}(|g|)hg\,\mathrm{d}\mu \\
&= \Lambda\big(|g|^{q-1}\mathbb{1}_{[0,n]}(|g|)h\big) \\
&\leq \|\Lambda\|\,\big\||g|^{q-1}\mathbb{1}_{[0,n]}(|g|)\big\|_p \\
&\leq \|\Lambda\|\,\big\|g\,\mathbb{1}_{[0,n]}(|g|)\big\|_q^{q/p}
\end{aligned}$$

et donc $\||g|\,\mathbb{1}_{[0,n]}(|g|)\|_q \leq \|\Lambda\|$. En passant à la limite lorsque n tend vers l'infini, $\|g\|_q \leq \|\Lambda\| < \infty$.

Il reste à montrer que si les formes linéaires $f \mapsto \int fg\,\mathrm{d}\mu$ et Λ coïncident sur L^∞, alors elles coïncident sur L^p. Lorsque $\mu(\Omega) < \infty$, on montre que tout espace L^p est dense dans tout espace L^r et donc que deux formes linéaires continues coïncidant sur L^∞ coïncident sur L^p. Si $\mu(\Omega) = \infty$, on utilise la σ-finitude de la mesure et on partitionne l'espace pour se ramener au cas fini.

La dernière affirmation découle du théorème de Hahn-Banach sur les duaux. □

On prendra garde au fait suivant : L^∞ est le dual de L^1, mais le dual de L^1 n'est pas (en général) L^∞ (voir exercice II.7).

De ce qui précède nous déduisons le résultat suivant qui permet d'utiliser des arguments géométriques dans les espaces $\mathrm{L}^2(\Omega, \mathcal{A}, \mu)$.

Théorème II.6.6. *L'espace* $\mathrm{L}^2(\Omega, \mathcal{A}, \mu)$ *est un espace de Hilbert pour le produit scalaire* $\langle f, g\rangle = \int_\Omega f g \,\mathrm{d}\mu$.

Exercices

Exercice II.1. Un exemple de fonction Lebesgue intégrable qui n'est pas Riemann intégrable : $f(x) = \mathbb{1}_{\mathbb{Q}\cap[0,1]}(x)$, $x \in [\,0,1\,]$. Montrer que $\int f \,\mathrm{d}\lambda = 0$ mais que f n'est pas Riemann intégrable sur $[\,0,1\,]$.

Exercice II.2. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, et soient A et B deux éléments de $\mathcal{A}$. Examiner le lemme de Fatou sur l'exemple suivant : $f_{2n} = \mathbb{1}_A$, $f_{2n+1} = \mathbb{1}_B$.

Exercice II.3. Soit μ une mesure de probabilité sur $I = [\,0,1\,]$. On note

$$m = \int_I x \,\mathrm{d}\mu(x)\,, \qquad v = \int_I (x-m)^2 \,\mathrm{d}\mu(x)\,,$$
$$a = \int_I x^2 \,\mathrm{d}\mu(x) - m^2\,, \qquad b = \Big(\frac{1}{2} - m\Big)^2 + \int_I x(1-x)\,\mathrm{d}\mu(x)\,.$$

Exprimer v et b en fonction de a. En déduire que $a \leq 1/4$ et que $a = 1/4$ pour une unique mesure μ que l'on déterminera.

Exercice II.4. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f, f_n, $n \in \mathbb{N}$, des fonctions mesurables positives intégrables. On suppose que

$$f_n \to f \quad \mu\text{-p.p.} \qquad \text{et} \qquad \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu\,.$$

En utilisant l'inégalité $(f - f_n)^+ \leq f$, démontrer que $\lim_{n\to\infty} \int (f-f_n)^+ \,\mathrm{d}\mu = 0$. En déduire que $f_n \to f$ dans $\mathrm{L}^1(\mu)$.

Exercice II.5. Soit $\mathrm{C}_K^\infty(\mathbb{R})$ l'ensemble des fonctions sur $\mathbb{R}$, infiniment différentiables, à support compact. Montrer que si A est intervalle ouvert, alors $\mathbb{1}_A$ est limite simple de fonctions dans $\mathrm{C}_K^\infty(\mathbb{R})$, majorées par 1.
Indication : on pourra d'abord considérer l'intervalle $[\,0,1\,]$ et les fonctions $\exp(-\varepsilon/x(1-x))$ si $x \in]\,0,1\,[$ et 0 si $x \notin]\,0,1\,[$.

En déduire que $\sigma(\mathrm{C}_K^\infty(\mathbb{R})) = \mathcal{B}(\mathbb{R})$ et qu'une mesure μ est caractérisée par la donnée de $\int f \,\mathrm{d}\mu$ pour toute fonction $f \in \mathrm{C}_K^\infty(\mathbb{R})$.

Exercice II.6. Si $\mu_1 \ll \mu_2 \ll \mu_3$, montrer que $\dfrac{\mathrm{d}\mu_1}{\mathrm{d}\mu_3} = \dfrac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\dfrac{\mathrm{d}\mu_2}{\mathrm{d}\mu_3}$, μ_3-p.p. Si de plus $\mu_2 \ll \mu_1$, alors $\dfrac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1} = \left(\dfrac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\right)^{-1}$, μ_1-p.p. et μ_2-p.p.

Exercice II.7. Cet exercice montre que le dual topologique de $\mathrm{L}^\infty([\,0,1\,],\mathcal{B}([\,0,1\,]),\lambda) = \mathrm{L}^\infty$ n'est pas $\mathrm{L}^1([\,0,1\,],\mathcal{B}([\,0,1\,]),\lambda) = \mathrm{L}^1$. En effet, $C[\,0,1\,] \subset \mathrm{L}^\infty \subset (\mathrm{L}^1)^*$ où * désigne le dual. La masse de Dirac δ_0 est dans le dual de $C[\,0,1\,]$ par la dualité $\langle \delta_0, f\rangle = \int f\,\mathrm{d}\delta_0 = f(0)$. De plus la norme de $\delta_0 \in C[\,0,1\,]^*$ est 1. Par le théorème de Hahn-Banach, montrer que l'on peut prolonger δ_0 en une forme linéaire Λ sur L^∞, de norme 1. Prouver que Λ n'est pas dans L^1.

Exercice II.8. Soit $\mathrm{L}^1([\,0,1\,],\lambda)$ l'espace des fonctions réelles intégrables pour la mesure de Lebesgue λ sur $[\,0,1\,]$. On considère la suite de fonctions

$$a_n(t) = 2 + \sin(nt)\,, \qquad t \in \mathbb{R}\,, \quad n \in \mathbb{N}\,.$$

a) Démontrer que pour toute fonction f de $\mathrm{L}^1([\,0,1\,],\lambda)$, on a

$$\lim_{n\to\infty} \int_{[\,0,1\,]} f(t)a_n(t)\,\mathrm{d}\lambda(t) = 2\int_{[\,0,1\,]} f(t)\,\mathrm{d}\lambda(t)\,.$$

Indication : Utiliser la densité des fonctions de classe C^1 dans $\mathrm{L}^1([\,0,1\,],\lambda)$ et intégrer par parties.

b) Démontrer que pour toute fonction f de $\mathrm{L}^1([\,0,1\,],\lambda)$, on a

$$\lim_{n\to\infty} \int_{[0,1]} \frac{f(t)}{a_n(t)}\,\mathrm{d}\lambda(t) = \beta\int_{[0,1]} f(t)\,\mathrm{d}\lambda(t)$$

où $\beta = (2\pi)^{-1}\int_0^{2\pi}(2+\sin u)^{-1}\mathrm{d}u$.

Indication : Utiliser la densité des fonctions en escalier dans $\mathrm{L}^1([\,0,1\,],\lambda)$.

c) Prouver que $\beta \neq 1/2$.

Exercice II.9. Sur un espace mesuré $(\Omega,\mathcal{A},\mu)$, soient f et g deux fonctions intégrables positives ou nulles telles que $\int f\,\mathrm{d}\mu = \int g\,\mathrm{d}\mu = 1$. On définit les mesures (de probabilité) P et Q de densités f et g par rapport à μ. Si $\|P-Q\|$ désigne la distance en variation totale définie par

$$\|P-Q\| = \sup_{A\in\mathcal{A}} \bigl|P(A)-Q(A)\bigr|\,,$$

démontrer que

$$\|P-Q\| = \frac{1}{2}\int |f-g|\,\mathrm{d}\mu\,.$$

III

MESURES DE PROBABILITÉ

Dans ce chapitre, nous définissons les notions de base des probabilités, à savoir, ce que sont une mesure de probabilité et une variable aléatoire. Il ne faut pas perdre de vue que les mathématiques ne proposent au mieux qu'un modèle de certains mécanismes réels. La définition mathématique d'une variable aléatoire est choquante à première vue, puisque nous verrons qu'il n'y a absolument rien d'aléatoire et de variable dans cette définition ! Mais à l'usage, nous verrons que le calcul des probabilités que l'on peut développer à partir de cette définition coïncide avec l'intuition que l'on peut avoir en observant des phénomènes qualifiés d'aléatoires.

L'axiomatique que nous présentons ici est essentiellement due à Kolmogorov (1903–1987). C'est la plus communément utilisée. Ce n'est pas la seule possible. Il en existe de nombreuses autres et l'on pourra utilement consulter l'ouvrage de Fine (1973) à ce propos.

III.1. Définition et exemples

L'objet de cette section est de transcrire une partie des notions introduites dans les chapitres précédents en termes probabilistes, définissant ainsi les notions fondamentales du calcul des probabilités. Nous commençons par définir ce qu'est une probabilité.

Définition III.1.1. Soit $(\Omega, \mathcal{A})$ un espace mesurable. On appelle probabilité (ou mesure de probabilité) sur $(\Omega, \mathcal{A})$ toute mesure positive P sur $\mathcal{A}$ telle que $P(\Omega) = 1$. On dit que $(\Omega, \mathcal{A}, P)$ est un espace probabilisé. On dit aussi que P est une loi de probabilité, ou simplement une loi.

En particulier, si μ est une mesure sur $(\Omega, \mathcal{A})$ avec $0 < \mu(\Omega) < \infty$, on voit que $P = \mu/\mu(\Omega)$ est une probabilité.

Si P est une probabilité, observons que P est à valeurs dans $[\,0, 1\,]$ puisque pour tout ensemble A mesurable, $P(A) \leq P(\Omega) = 1$. De plus, $P(\emptyset) = 0$.

Donnons à présent quelques exemples de mesures de probabilité. L'appendice donne un inventaire des mesures de probabilité usuelles et de leurs caractéristiques principales.

Exemples III.1.2. (i) Soit $(\Omega, \mathcal{A})$ un espace mesurable. La masse de Dirac δ_x en $x \in \Omega$ est la probabilité définie par $\delta_x(A) = 1$ si $x \in A$ et 0 si $x \notin A$. En d'autres termes, $\delta_x(A) = \mathbb{1}_A(x)$.
(ii) Soit $\Omega = \{\,1, 2, 3, 4, 5, 6\,\}$ muni de la tribu de ses parties et de la mesure $P = \frac{1}{6} \sum_{1 \leq i \leq 6} \delta_i$, proportionnelle à la mesure de comptage. Cette mesure est une probabilité. Cette probabilité sert à modéliser le jet d'un dé. Intuitivement, si $A \in \mathcal{P}(\Omega)$, $P(A)$ est la probabilité que le jet du dé donne un chiffre appartenant à l'ensemble A. Comme on l'attend intuitivement, $P(\{\,i\,\}) = 1/6$ pour tout $i \in \Omega$ et, par exemple, la probabilité de tirer un chiffre pair est $P(\{\,2, 4, 6\,\}) = 1/2$.

(iii) Soit $0 \leq p \leq 1$. La mesure de probabilité $P = p\,\delta_1 + (1-p)\,\delta_0$ est appelé probabilité ou loi de Bernoulli de paramètre (de succès) p. Plus généralement, toute mesure de probabilité concentrée en deux points distincts sera appelée probabilité de Bernoulli. Lorsque $p = 1/2$, elle est utilisée par exemple pour modéliser le jet d'une pièce dans un jeu de pile ou face équilibré. En effet, en comptant 0 pour pile et 1 pour face, elle donne $P(\{\,0\,\}) = P(\{\,1\,\}) = 1/2$.

(iv) Soit $(x_k)_{k \in \mathbb{N}}$ une suite de points d'un espace $(\Omega, \mathcal{A})$ et $(p_k)_{k \in \mathbb{N}}$ des réels positifs de somme égale à 1. On peut construire une probabilité P en posant $P = \sum_{k \geq 0} p_k \delta_{x_k}$. Une telle probabilité est absolument continue par rapport à la mesure de comptage sur l'ensemble $E = \{\,x_k : k \in \mathbb{N}\,\}$, qui à toute partie de E associe son cardinal. Toute mesure de probabilité sur $(E, \mathcal{P}(E))$ est de cette forme.

Par exemple, si $\Omega = \mathbb{N}$ est muni de la tribu des parties $\mathcal{A} = \mathcal{P}(\mathbb{N})$, si $x_k = k$ et $p_k = e^{-\lambda}\lambda^k/k!$, $\lambda > 0$, on obtient la probabilité

$$P = e^{-\lambda} \sum_{k \in \mathbb{N}} \frac{\lambda^k}{k!}\, \delta_k\,,$$

appelée loi de Poisson de paramètre λ, notée $\mathcal{P}(\lambda)$.

Si toujours $\Omega = \mathbb{N}$ est muni de la tribu de ses parties, et si $x_k = k$, $p_k = (1-p)p^k$, la probabilité $P = (1-p) \sum_{k \in \mathbb{N}} p^k \delta_k$ est appelée loi géométrique de paramètre $p \in [\,0, 1\,]$.

Soient les coefficients binomiaux $\mathrm{C}_n^k = n!/k!(n-k)!$, $0 \leq k \leq n$. En prenant $\Omega = \{0, 1, \ldots, n\}$, $x_k = k$ et $p_k = \mathrm{C}_n^k p^k(1-p)^{n-k}$ pour $0 \leq k \leq n$, on obtient la loi dite binomiale de paramètre $p \in [0,1]$ et de taille n, notée $\mathcal{B}(n,p)$. La loi $\mathcal{B}(1,p)$ est une loi de Bernoulli sur $\{0,1\}$.

(v) Si $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ et f est une fonction mesurable positive telle que $\int f \, \mathrm{d}\lambda = 1$, on peut considérer la probabilité $P(A) = \int_A f \, \mathrm{d}\lambda$, $A \in \mathcal{A}$ (*cf.* II.3.1). La fonction f est la densité de P (par rapport à λ) (*cf.* II.3.3).

Définition III.1.3. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Un ensemble $A \in \mathcal{A}$ est appelé un événement. Un événement A a lieu P-presque sûrement (P-p.s.) s'il a lieu P-p.p. (*i.e.* si $P(A) = 1$).

Exemple III.1.4. En reprenant l'exemple III.1.2.ii, l'ensemble $\{2, 4, 6\}$ est un événement. Il modélise le tirage d'un chiffre pair lors d'un lancé de dé. Dans cet exemple, le seul événement qui a lieu presque sûrement est $\{1, 2, 3, 4, 5, 6\}$. Nous verrons par la suite, en particulier au chapitre V, des situations beaucoup moins triviales.

Dans tout ce qui suit, on considère un espace probabilisé $(\Omega, \mathcal{A}, P)$.

Définition III.1.5. On appelle variable aléatoire toute application mesurable définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

Il est d'usage d'utiliser $X, Y, \ldots$ pour noter des variables aléatoires. Pour l'essentiel, on se contentera ici de variables à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Elles seront appelées variables aléatoires réelles ou vectorielles.

Si X et Y sont deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$ avec $X = Y$ P-p.p., on écrit aussi bien $X = Y$ P-p.s. ou $X = Y$ p.s. s'il n'y a pas d'ambiguïté sur P.

Exemples III.1.6. (i) Soit $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$ la tribu borélienne de $[0,1]$ et soit $P(A) = \lambda(A)$, $A \in \mathcal{B}(\mathbb{R})$, la mesure de Lebesgue sur $[0,1]$. L'application identité de $(\Omega, \mathcal{A})$ dans $(\Omega, \mathcal{A})$ est mesurable. C'est donc une variable aléatoire. On appelle aussi P la probabilité uniforme sur $[0,1]$, que l'on notera $\mathcal{U}_{[0,1]}$.

(ii) Sur l'espace probabilisé $(\Omega, \mathcal{A}, \delta_x)$ où $x \in \Omega$, toute variable aléatoire X est δ_x-p.s. constante. En effet, $\delta_x(\{\omega : X(\omega) = c\}) = 1$ si $c = x$ et 0 sinon.

En reprenant les propriétés des mesures, on voit que si P est une probabilité sur $(\Omega, \mathcal{A})$ et si A, B, A_n, $n \in \mathbb{N}$, sont mesurables, alors

(i) $A \subset B \Rightarrow P(A) \leq P(B)$.

(ii) $P(A^c) = 1 - P(A)$.
(iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
(iv) $P(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} P(A_n)$.
(v) Si les ensembles A_n sont croissants, ou décroissants,

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)\,.$$

(vi) De (iii) et (iv) nous déduisons l'inégalité de Bonferoni : pour tout $k \in \mathbb{N}$,

$$\sum_{0\leq n\leq k} P(A_n) - \sum_{0\leq n\leq m\leq k} P\big(A_n \cap A_m\big) \leq P\Big(\bigcup_{0\leq n\leq k} A_n\Big) \leq \sum_{0\leq n\leq k} P(A_n)\,.$$

La minoration se démontre par récurrence, en notant que

$$\begin{aligned} P\Big(\bigcup_{0\leq n\leq k} A_n\Big) &= P\Big(A_0 \cup \bigcup_{1\leq n\leq k} A_n\Big) \\ &= P\Big(\bigcup_{1\leq n\leq k} A_n\Big) + P(A_0) - P\Big(A_0 \cap \bigcup_{1\leq n\leq k} A_n\Big) \\ &\geq P\Big(\bigcup_{1\leq n\leq k} A_n\Big) + P(A_0) - \sum_{1\leq n\leq k} P\big(A_0 \cap A_n\big)\,. \end{aligned}$$

Les énoncés sur les mesures peuvent être reformulés sur les probabilités. En particulier le théorème de prolongement de Carathéodory (I.4.9) montre qu'une probabilité est définie si elle est donnée sur une algèbre qui engendre la tribu.

De même que nous avons défini les mesures images, on peut définir les lois images. La définition suivante introduit la notion fondamentale de loi d'une variable aléatoire.

Définition III.1.7. Soit X une variable aléatoire de $(\Omega, \mathcal{A}, P)$ à valeurs dans $(E, \mathcal{B})$. On appelle loi de X sous la probabilité P la mesure de probabilité image P^X sur $(E, \mathcal{B})$. On notera parfois $\mathcal{L}(X)$ la loi de X.

Il est usuel et commode d'alléger les notations des lois images en posant pour tout $B \in \mathcal{B}$,

$$P^X(B) = P\big(\{\,\omega \in \Omega \,:\, X(\omega) \in B\,\}\big) = P\big(\{\,X \in B\,\}\big) = P\{\,X \in B\,\}\,.$$

En pratique, l'espace $(\Omega, \mathcal{A}, P)$ est un peu mythique. Si l'on se donne une loi, on peut toujours l'écrire comme une loi image par une application mesurable (prendre l'identité pour la variable aléatoire!). Donc toute mesure de probabilité est la loi d'une variable aléatoire. Pour les applications, en général, seule compte la mesure image, et l'on explicite rarement la variable aléatoire et l'espace $(\Omega, \mathcal{A}, P)$. On écrira par exemple « soit X une variable de Bernoulli de paramètre p, c'est-à-dire telle que $P\{X = 1\} = 1 - P\{X = 0\} = p$ » au lieu de « soit X une variable aléatoire de l'espace $(\Omega, \mathcal{A}, P)$ dans $\{0, 1\}$, de loi de Bernoulli, c'est-à-dire telle que $P^X(\{1\}) = 1 - P^X(\{0\}) = p$, ou plus exactement $P\{X = 1\} = 1 - P\{X = 0\} = p$ ». De même, on écrira souvent « soit X une variable aléatoire de loi P » pour dire « soit X une variable aléatoire définie sur l'espace $(\Omega, \mathcal{A}, \mu)$ où μ est une mesure de probabilité telle que la mesure image μ^X est P ».

Il s'ensuit que l'on peut considérer de façon complémentaire les variables aléatoires et leurs lois. Selon le contexte, l'un ou l'autre de ces points de vue est préférable. Souvent nous utiliserons les variables aléatoires. On prendra garde au fait que le langage aura souvent tendance à confondre les variables aléatoires et leurs lois.

La représentation d'une loi par une variable aléatoire n'est pas unique. Par exemple, pour la loi de Bernoulli $\mathcal{B}(1, p)$, on peut choisir $\Omega = \{0, 1\}$ muni de la tribu des parties et de la probabilité $P = p\,\delta_1 + (1 - p)\,\delta_0$ et X l'application identité de $\{0, 1\}$ dans lui-même. On peut aussi choisir $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, P la mesure uniforme sur $[0, 1]$ (Lebesgue) et $X : \Omega \to \{0, 1\}$ définie par $X(\omega) = \mathbb{1}_{[0,p]}(\omega)$. La mesure image P^X est la loi de Bernoulli de paramètre p.

La définition suivante a pour but de résumer les deux classes fondamentales de lois rencontrées dans les exemples précédents.

Définition III.1.8. On dit qu'une loi est discrète si c'est une combinaison linéaire finie ou dénombrable de masses de Dirac. Une variable aléatoire de loi discrète $P = \sum_{i \in I} p_i \delta_{x_i}$ ne prend (presque sûrement) qu'un nombre fini ou dénombrable de valeurs.

Si une loi P est absolument continue par rapport à une mesure μ et si X est de loi P, on dira par abus de langage que X admet la densité f par rapport à μ si $f = \mathrm{d}P/\,\mathrm{d}\mu$. Si μ est la mesure de Lebesgue, on dit simplement que X est de densité f.

III.2. Fonctions de répartition

Soit X une variable aléatoire réelle (*i.e.* X est à valeurs réelles), définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

Définition III.2.1. On appelle fonction de répartition de X, ou de sa loi P^X, et on note F^X, la fonction sur $\mathbb{R}$ définie par

$$F^X(t) = P^X(]-\infty, t\,]) = P(\{\,\omega : X(\omega) \leq t\,\}) = P\{\,X \leq t\,\}\,, \quad t \in \mathbb{R}\,.$$

Propriété III.2.2. *Une fonction de répartition F vérifie les propriétés suivantes :*
(i) $0 \leq F \leq 1$,
(ii) F est croissante, continue à droite avec une limite à gauche en tout point,
(iii) $\lim_{t\to-\infty} F(t) = 0$ *et* $\lim_{t\to\infty} F(t) = 1$.

Réciproquement, une fonction F vérifiant (i)–(iii) est la fonction de répartition d'une variable aléatoire réelle.

Démonstration. (i) vient de ce que P est à valeurs dans $[\,0,1\,]$. La croissance dans (ii) découle de la croissance des mesures (*i.e.* $A \subset B \Rightarrow P(A) \leq P(B)$). La continuité à droite peut être vue comme une conséquence de la proposition I.4.3.iv en remarquant que

$$\{\,X \leq t\,\} = \bigcap_{n\geq 1} \{\,X \leq t + 1/n\,\}$$

et que la croissance de F implique

$$\lim_{h\downarrow 0} F(t+h) = \lim_{n\to\infty} F\big(t + \tfrac{1}{n}\big) = F(t)\,.$$

La limite à gauche est également une conséquence de la croissance de F.

La propriété (iii) vient encore de la proposition I.4.3.iv en remarquant que l'intersection $\bigcap_{n\geq 1}\{\,X \leq -n\,\}$ et vide, et donc

$$0 = P(\emptyset) = \lim_{n\to\infty} P\{\,X \leq -n\,\} = \lim_{n\to\infty} F(-n)$$

tandis que $1 = P(\Omega) = \lim_{n\to\infty} P\{\,X \leq n\,\}$ d'après I.4.3.iii.

Soit maintenant une fonction F vérifiant (i)–(iii). Définissons pour $a < b$ la fonction d'ensembles $\mu_F(]\,a,b\,]) = F(b) - F(a)$. On étend la définition de μ à l'algèbre des unions finies d'intervalles. Le théorème de prolongement I.4.9 permet ensuite de conclure, comme dans l'exemple I.4.10.i, que μ_F s'étend en une mesure de Stieltjes de probabilité. $\square$

Propriété III.2.3. *La fonction de répartition caractérise la loi, c'est-à-dire $F^X = F^Y$ si et seulement si $P^X = P^Y$.*

Démonstration. En effet, si $F^X = F^Y$, alors P^X et P^Y coïncident sur les intervalles, donc sur l'algèbre et la tribu engendrées par les intervalles. La tribu engendrée par les intervalles est la tribu borélienne et le résultat s'ensuit. □

***Propriété* III.2.4.** *Une fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.*

Démonstration. Soit D_n l'ensemble des points de discontinuité avec un saut d'amplitude plus grande que $1/n$; en notant $F(t-)$ la limite à gauche de F en t,

$$D_n = \{\, t \in \mathbb{R} \,:\, F(t) - F(t-) \geq 1/n \,\}\,.$$

Puisque $0 \leq F \leq 1$, nécessairement $\mathrm{card}(D_n) \leq n$. L'ensemble des points de discontinuité est $\bigcup_{n\geq 1} D_n$, et donc est dénombrable. Notons que le même raisonnement s'applique en fait à toute fonction croissante. □

***Exemple* III.2.5.** Soit F une fonction de répartition. Soit $(x_n)_{n\in I}$, $I \subset \mathbb{N}$, la suite des points de discontinuité de F et $a_n = F(x_n) - F(x_n-)$ le saut correspondant. On peut poser $\mathcal{F}_d = \sum_{n\in I} a_n \mathbb{1}_{[x_n,\infty[}$. Soit $\alpha = \lim_{t\to\infty} \mathcal{F}_d(t)$. Si $\alpha = 0$, la fonction F est continue. Sinon, $F_d = \frac{1}{\alpha}\mathcal{F}_d$ est une fonction de répartition discrète. C'est en fait la fonction de répartition de la mesure de probabilité $\frac{1}{\alpha}\sum_{n\in I} a_n \delta_{x_n}$. Si $\alpha = 1$, alors $F = F_d$ est discrète. Sinon, $F_c = \frac{1}{1-\alpha}(F - \mathcal{F}_d)$ est une fonction de répartition continue. Ainsi F est la moyenne $\alpha F_d + (1-\alpha)F_c$ d'une fonction de répartition continue et d'une fonction de répartition discrète.

Notons P_c la mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de fonction de répartition F_c. En décomposant P_c suivant la mesure de Lebesgue λ par le théorème II.3.4, on pourra écrire $P_c = \beta P_{ac} + (1-\beta)P_\perp$ pour un $\beta \in [\,0,1\,]$, P_{ac} étant une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue, et $P_\perp$ lui étant singulière. Notons F_{ac} et $F_\perp$ leur fonction de répartition. L'absolue continuité de P_{ac} par rapport à λ permet d'exprimer

$$F_{ac}(t) = \int_{]-\infty,t]} f \,\mathrm{d}\lambda = \int_{-\infty}^{t} f \,\mathrm{d}\lambda\,, \quad t \in \mathbb{R}\,,$$

pour une fonction de densité $f \geq 0$ (théorème de Radon-Nikodym, II.3.3). Ainsi,

$$F = (1-\alpha)\beta F_{ac} + (1-\alpha)(1-\beta)F_\perp + \alpha F_d\,.$$

La partie donnée par F_d est discrète, et la partie donnée par $F_\perp$ est continue (*i.e.* ne contient aucune masse de Dirac, donc tout point est de mesure nulle), mais étrangère par rapport à la mesure de Lebesgue (un exemple classique d'une telle mesure étrangère est donné dans l'exercice V.13).

Si P^X admet une densité par rapport à la mesure de Lebesgue, sa fonction de répartition s'écrit

$$F(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}\lambda(x)\,,$$

avec $f = \mathrm{d}P^X/\,\mathrm{d}\lambda$, et la densité f s'obtient pratiquement comme la dérivée λ-p.p. de F.

Exemples III.2.6. (i) Soit $\theta > 0$ et soit $F(t) = 1-e^{-\theta t}$ si $t \geq 0$ et $F(t) = 0$ si $t < 0$. C'est une fonction de répartition. Sa densité est $\theta e^{-\theta t}$ si $t \geq 0$ et 0 si $t < 0$. C'est la fonction de répartition de la loi exponentielle de paramètre θ, notée $\mathcal{E}xp(\theta)$.

(ii) $F = \mathbb{1}_{[x,\infty[}$ est la fonction de répartition de la masse de Dirac δ_x en $x \in \mathbb{R}$. (Faire un dessin.)

(iii) $F(t) = \sum_{0\leq k\leq n} \mathrm{C}_k^n\, p^k(1-p)^{n-k}\mathbb{1}_{[k,\infty[}(t)$ est la fonction de répartition de la loi binomiale $\mathcal{B}(n,p)$ (*cf.* exemple III.1.2.iv).

(iv) Soit $f(x) = (2\pi)^{-1/2}e^{-x^2/2}$, $x \in \mathbb{R}$. On vérifie que $\int_{\mathbb{R}} f\,\mathrm{d}\lambda = 1$ et donc que f est une densité. En effet, par un changement de variables en coordonnées polaires,

$$\begin{aligned}
\left(\int_{\mathbb{R}} e^{-x^2/2}\,\mathrm{d}\lambda(x)\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)^2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^{2\pi}\left(\int_0^{\infty} re^{-r^2/2}dr\right)\mathrm{d}\theta = 2\pi\,.
\end{aligned}$$

Sa fonction de répartition $F(t) = (2\pi)^{-1/2}\int_{-\infty}^{t} e^{-x^2/2}\,\mathrm{d}x$, $t \in \mathbb{R}$, est la fonction de répartition d'une loi appelée loi normale ou loi gaussienne, centrée, réduite, notée $\mathcal{N}(0,1)$.

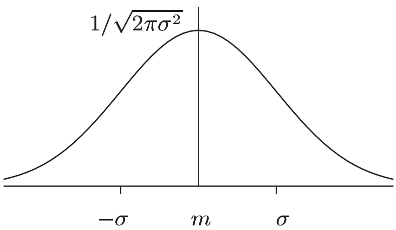
Si X est de fonction de répartition F, alors pour $m \in \mathbb{R}$ et $\sigma > 0$, la variable aléatoire $Y = \sigma X + m$ a pour fonction de répartition $F\big((t-m)/\sigma\big)$ puisque

$$P\{\,\sigma X + m \leq t\,\} = P\Big\{\,X \leq \frac{t-m}{\sigma}\,\Big\}\,.$$

En particulier, si X est de loi normale $\mathcal{N}(0,1)$, on notera $\mathcal{N}(m,\sigma^2)$ la loi de $Y = \sigma X + m$, appelée loi normale de moyenne m et variance σ^2. (La terminologie sera justifiée plus loin.)

Par un changement de variables, la densité de la loi de Y est

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-m)^2}{2\sigma^2}\Big) .$$



La densité représente la “cloche gaussienne”, symétrique par rapport à m, d'autant plus pointue que σ est petit. En particulier, $\mathcal{N}(m,0)$ peut être vue comme la masse de Dirac en m.

(v) $F(t) = t$ si $t \in [\,0,1\,]$, 0 si $t < 0$ et 1 si $t > 1$ est une fonction de répartition (faire un dessin). C'est la fonction de répartition de la mesure de Lebesgue sur $[\,0,1\,]$, appelée loi uniforme et notée $\mathcal{U}_{[0,1]}$. Plus généralement, on définit une loi uniforme $\mathcal{U}_{[a,b]}$ sur un intervalle borné $[\,a,b\,]$ quelconque. On pourrait tout aussi bien considérer les intervalles ouverts ou semi-ouverts.

(vi) Soit la fonction de répartition F d'une loi P, donnée par

$$F(t) = \begin{cases} 0 & \text{si } t \leq 0, \\ t/4 & \text{si } 0 \leq t < 1, \\ 1/2 & \text{si } 1 \leq t < 2, \\ \frac{2}{3} + \frac{1}{3}(1 - e^{-(t-2)}) & \text{si } t \geq 2. \end{cases}$$

Le graphe de F comporte deux points de discontinuité en 1 et 2 d'amplitudes respectives $1/4$ et $1/6$. La partie continue est dérivable presque partout par rapport à la mesure de Lebesgue, de densité

$$f(x) = \frac{1}{4}\, \mathbb{1}_{[0,1[}(x) + \frac{1}{3}\, e^{-(x-2)}\, \mathbb{1}_{[2,\infty[}(x) .$$

La mesure de probabilité P se représente donc comme

$$P = \frac{1}{4}\,\delta_1 + \frac{1}{6}\,\delta_2 + \mu_{ac}$$

avec μ_{ac} la mesure de densité f par rapport à la mesure de Lebesgue.

Une application intéressante des fonctions de répartition est donnée par la proposition suivante, qui montre que pour simuler numériquement une variable aléatoire de fonction de répartition F, il suffit de savoir simuler une variable aléatoire uniforme sur $[\,0,1\,]$. Le résultat fournit également une preuve alternative à la réciproque de la propriété III.2.2.

Proposition III.2.7. *Soit F une fonction de répartition. On appelle fonction de quantile la fonction*

$$F^{\leftarrow}(u) = \inf\{\, x : F(x) > u \,\}\,, \quad u \in]\,0,1\,[\,.$$

Si U est de loi uniforme sur $]\,0,1\,[$, alors $F^{\leftarrow}(U)$ a pour fonction de répartition F.

Démonstration. Observons d'abord que pour tout $u \in]\,0,1\,[$, si $F^{\leftarrow}(u) \leq t$, alors $F(t) \geq u$. En effet, si $F^{\leftarrow}(u) \leq t$, pour tout $s > t$ il existe $x < s$ tel que $F(x) > u$; ainsi $F(s) > u$, et par continuité à droite de F, $F(t) \geq u$. Réciproquement, si $F(t) > u$, alors t appartient à $\{\, x : F(x) > u \,\}$ et donc $F^{\leftarrow}(u) \leq t$. Par voie de conséquence, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} F(t) = P\{\, U < F(t) \,\} &\leq P\{\, F^{\leftarrow}(U) \leq t \,\} \\ &\leq P\{\, F(t) \geq U \,\} = F(t)\,, \end{aligned}$$

de sorte que $P\{\, F^{\leftarrow}(U) \leq t \,\} = F(t)$ et donc $F^{\leftarrow}(U)$ a pour fonction de répartition F. □

Remarquons que la fonction de quantile est bien définie (*i.e.* est finie) sur $]\,0,1\,[$. Elle croissante, et donc elle admet au plus un nombre dénombrable de points de discontinuité (par un argument tout à fait analogue à celui que nous avons utilisé dans la démonstration de la propriété III.2.4, en remarquant que $F^{\leftarrow}(1-u) - F^{\leftarrow}(u) < \infty$ pour tout $u < 1/2$). De plus, on vérifie facilement que si F est inversible, alors $F^{\leftarrow}$ est l'inverse de F.

La propriété suivante nous sera utile ultérieurement.

Proposition III.2.8. *Si $F^{\leftarrow}$ est une fonction de quantile, elle est continue à droite et admet une limite à gauche en tout point.*

Démonstration. La limite à gauche en tout point provient de la croissance de $F^{\leftarrow}$. Pour démontrer la continuité à droite en un point $u \in]\,0,1\,[$, montrons (ce qui suffit par croissance) que $\limsup_{n\to\infty} F^{\leftarrow}(u+\frac{1}{n}) \leq F^{\leftarrow}(u) = t$. Sinon, il existe $\eta > 0$ tel que $\limsup_{n\to\infty} F^{\leftarrow}(u+\frac{1}{n}) > t+\eta$. Le long d'une sous-suite (n_k), $F^{\leftarrow}(u+\frac{1}{n_k}) > t+\eta$. Autrement dit, d'après le raisonnement utilisé dans la démonstration de la proposition précédente, $F(t+\eta) \leq u + \frac{1}{n_k}$; et quand k tend vers l'infini, $F(t+\eta) \leq u$. En particulier, $F^{\leftarrow}(u) > t$, ce qui est impossible puisque $t = F^{\leftarrow}(u)$. □

III.3. Vecteurs aléatoires

Dans ce paragraphe, d est un entier supérieur ou égal à 2.

Définition III.3.1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On appelle vecteur aléatoire une variable aléatoire à valeurs dans $\mathbb{R}^d$ muni de sa tribu borélienne.

En utilisant le lemme I.2.2 et l'exemple I.1.12.iii, on voit que $X = (X_1, \ldots, X_d)$ est un vecteur aléatoire si et seulement si ses composantes sont des variables aléatoires réelles.

Définition III.3.2. On appelle fonction de répartition de X, ou de la loi de X, la fonction

$$t = (t_1, \ldots, t_d) \in \mathbb{R}^d \mapsto F^X(t) = P\{X_1 \le t_1, \ldots, X_d \le t_d\}.$$

La loi de la variable aléatoire X_i est appelée la i-ème loi marginale (ou i-ème marge) de $X = (X_1, \ldots, X_d)$. Elle est donnée par

$$F^{X_i}(t_i) = \lim_{t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_d \to \infty} F^X(t).$$

Comme il ressort de cette définition, la loi d'un vecteur $X = (X_1, \ldots, X_d)$ détermine chacune des lois marginales (loi de X_i, $1 \le i \le d$). L'exemple suivant montre que la réciproque est fausse en général.

Exemples III.3.3. (i) Supposons que $X = (X_1, X_2)$ soit de loi discrète dans $\mathbb{R}^2$ concentrée en les points $(-1, 0)$, $(0, 1)$, $(0, -1)$, $(1, 0)$ tous de probabilité $1/4$. Autrement dit,

$$P^X = \frac{1}{4}\,\delta_{(-1,0)} + \frac{1}{4}\,\delta_{(0,1)} + \frac{1}{4}\,\delta_{(0,-1)} + \frac{1}{4}\,\delta_{(1,0)},$$

ce qui se résume dans le tableau ci-contre. Les lois marginales P^{X_1} et P^{X_2} de P^X sont égales, et données par $P^{X_1} = P^{X_2} = \frac{1}{4}\,\delta_{-1} + \frac{1}{2}\,\delta_0 + \frac{1}{4}\,\delta_1$.

On peut produire un autre vecteur, (Y_1, Y_2), ayant les mêmes lois marginales, dont les probabilités sont données par le tableau ci-contre.

On pourra noter que l'on obtient les lois marginales en sommant les probabilités respectivement sur les lignes et les colonnes de la table.

$X_2 \backslash X_1$	-1	0	1
-1	0	$1/4$	0
0	$1/4$	0	$1/4$
1	0	$1/4$	0

$Y_2 \backslash Y_1$	-1	0	1
-1	$1/16$	$1/8$	$1/16$
0	$1/8$	$1/4$	$1/8$
1	$1/16$	$1/8$	$1/16$

(ii) Soit $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$

($\lambda \otimes \cdots \otimes \lambda$ d-fois) est donnée par

$$f(x) = \frac{1}{(2\pi)^{d/2}} \exp(-\|x\|^2/2)$$

avec $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ et la norme euclidienne $\|x\|^2 = x_1^2 + \cdots + x_d^2$. D'après le théorème de Fubini-Tonelli (II.5.1), les lois marginales sont des lois $\mathcal{N}(0,1)$.

(iii) Plus généralement, supposons que $Z = (X, Y)$ admette une densité $f(x, y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, *i.e.*

$$F(t_1, t_2) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} f(x, y) \, \mathrm{d}x \, \mathrm{d}y \,, \quad t_1, t_2 \in \mathbb{R} \,.$$

La loi de X a pour densité $f^X(x) = \int_{\mathbb{R}} f(x, y) \, \mathrm{d}y$, et celle de Y a pour densité $f^Y(y) = \int_{\mathbb{R}} f(x, y) \, \mathrm{d}x$. En effet,

$$P\{X \leq t_1\} = \lim_{t_2 \to \infty} F(t_1, t_2) = \int_{-\infty}^{t_1} \left(\int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}y \right) dx \,,$$

et l'on conclut avec l'exemple III.2.5.

Si de plus $f(x, y) = h(x)g(y)$ avec $\int h(x) \, \mathrm{d}x = 1$, on voit que $f^X = h$ et $f^Y = g$.

La propriété III.2.3 se généralise sans difficulté.

***Propriété* III.3.4.** *Soient deux vecteurs aléatoires X, Y, définis sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{R}^d$. Alors $P^X = P^Y$ si et seulement si $F^X = F^Y$.*

Démonstration. Comme pour la propriété III.2.3, remarquer que les pavés $]-\infty, a_1] \times \cdots \times]-\infty, a_d]$, $(a_1, \ldots, a_d) \in \mathbb{R}^d$, engendrent $\mathcal{B}(\mathbb{R}^d)$. $\square$

III.4. Moyennes et inégalités

Nous avons vu au chapitre II comment intégrer des fonctions mesurables. Nous pouvons donc intégrer les variables aléatoires.

***Définition* III.4.1.** Soit X une variable aléatoire réelle, définie sur $(\Omega, \mathcal{A}, P)$. Si X est intégrable, on appelle espérance ou espérance mathématique de X (sous la probabilité P) le nombre réel

$$E(X) = \int_\Omega X \,\mathrm{d}P\,.$$

On dit que X est centrée si elle est intégrable et $E(X) = 0$.

L'espérance d'une variable aléatoire n'est donc rien d'autre que sa valeur moyenne. Une mesure de probabilité étant de masse totale égale à 1, l'espérance d'une variable aléatoire constante ou presque sûrement constante est égale à cette constante.

Plus généralement si $X \in \mathrm{L}^p$, $p > 0$, on définit le moment absolu d'ordre p de X par $E(|X|^p) = \int |X|^p \,\mathrm{d}P$. Si p est entier, on peut aussi définir le moment d'ordre p, $E(X^p) = \int X^p \,\mathrm{d}P$.

Rappelons quelques résultats du chapitre II sous une autre formulation. Commençons par le théorème de transport II.4.1. Nous le formulons ici, dans le langage probabiliste, pour des vecteurs aléatoires.

***Théorème* III.4.2 (de transport).** *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et soit ϕ une fonction borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$. Si ϕ est à valeurs positives,*

$$E\big(\phi(X)\big) = \int_\Omega \phi \circ X(\omega) \,\mathrm{d}P(\omega) = \int_{\mathbb{R}^d} \phi(x) \,\mathrm{d}P^X(x)\,.$$

Si ϕ est à valeurs quelconques,

$$\phi(X) \in \mathrm{L}^1(\Omega, \mathcal{A}, P) \text{ si et seulement si } \phi \in \mathrm{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P^X)\,.$$

Dans ce cas, l'égalité précédente a lieu.

En particulier, si X est une variable aléatoire réelle, intégrable,

$$E(X) = \int_\Omega X \,\mathrm{d}P = \int_{\mathbb{R}} x \,\mathrm{d}P^X(x)\,.$$

***Remarque* III.4.3.** Notons les deux faits importants suivants.

(i) Si X est un vecteur aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et si $A \in \mathcal{B}(\mathbb{R}^d)$, la fonction $\mathbb{1}_A$ est mesurable. Par définition de l'intégrale et par transport,

$$\begin{aligned} E\big(\mathbb{1}_A(X)\big) = \int_\Omega \mathbb{1}_A\big(X(\omega)\big)\,\mathrm{d}P(\omega) &= \int_{\mathbb{R}^d} \mathbb{1}_A(x)\,\mathrm{d}P^X(x) \\ &= P^X(A) = P\{X \in A\}\,. \end{aligned}$$

(ii) Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, admettant une densité f. Soit h une bijection sur $\mathbb{R}^d$, de classe C^1, de jacobien $J_h(x) \neq 0$ pour tout x. Le vecteur $Y = h(X)$ a pour densité

$$g(y) = |J_{h^{-1}}(y)| f \circ h^{-1}(y) = \big|J_h\big(h^{-1}(y)\big)\big|^{-1} f \circ h^{-1}(y)\,.$$

En effet, si ϕ est une fonction borélienne bornée (par exemple une indicatrice de borélien), d'après le théorème de transport et la formule de changement de variables pour des intégrales de Lebesgue, on a

$$\begin{aligned} E\big(\phi \circ h(X)\big) &= \int_{\mathbb{R}^d} \phi \circ h(x)\,\mathrm{d}P^X(x) \\ &= \int_{\mathbb{R}^d} \phi \circ h(x) f(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \phi(y)\big|J_{h^{-1}}(y)\big| f \circ h^{-1}(y)\,\mathrm{d}y\,. \end{aligned}$$

Si la densité f est nulle hors d'un ouvert U de $\mathbb{R}^d$, la même formule s'applique si h est définie sur U.

Dans la pratique, la loi de X se décompose le plus souvent en une partie absolument continue par rapport à la mesure de Lebesgue, et une partie absolument continue par rapport à une mesure de comptage. Si P^X a une densité f par rapport à la mesure de Lebesgue, sous les conditions d'intégrabilité du théorème III.4.2,

$$E\big(\phi(X)\big) = \int_{\mathbb{R}^d} \phi(x)\,\mathrm{d}P^X(x) = \int_{\mathbb{R}^d} \phi(x) f(x)\,\mathrm{d}x\,.$$

Si $P^X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$,

$$E\big(\phi(X)\big) = \int_{\mathbb{R}^d} \phi(x)\,\mathrm{d}P^X(x) = \sum_{n \in \mathbb{N}} \phi(x_n) p_n = \sum_{n \in \mathbb{N}} \phi(x_n) P\{X = x_n\}\,.$$

On voit donc qu'en pratique, le calcul de $E(\phi(X))$ ne nécessite pas le calcul de la loi de $\phi(X)$.

***Exemples* III.4.4.** (i) Soit X de loi $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Alors $E(X) = 1/2$: dans un jeu de pile ou face équilibré, on tire en moyenne une fois sur deux pile ($X = 1$) et une fois sur deux face ($X = 0$) !

(ii) Soient $x_1, \ldots, x_n$ des réels et $P_n = n^{-1} \sum_{1 \leq i \leq n} \delta_{x_i}$. Si X est de loi P_n, alors

$$E(X) = \frac{1}{n} \sum_{1 \leq i \leq n} x_i$$

est la moyenne des x_i.

(iii) Si X est de loi binomiale $\mathcal{B}(n, p)$,

$$\begin{aligned}
E(X) &= \sum_{0 \leq k \leq n} k \mathrm{C}_n^k p^k (1-p)^{n-k} \\
&= pn \sum_{1 \leq k \leq n} \mathrm{C}_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= pn \sum_{0 \leq k \leq n-1} \mathrm{C}_{n-1}^{k} p^{k} (1-p)^{(n-1)-k} = pn\,.
\end{aligned}$$

(iv) Si X suit une loi de Poisson $\mathcal{P}(\lambda)$, on vérifie comme dans l'exemple (iii) que $E(X) = \lambda$.

(v) Soit X de loi exponentielle de fonction de répartition $1 - F(t) = e^{-\theta t}$, $t \geq 0$. Elle a pour densité $\theta e^{-\theta x} \mathbb{1}_{[0,\infty[}(x)$. Ainsi, en intégrant par parties,

$$E(X) = \int_0^\infty x\theta e^{-\theta x}\, \mathrm{d}x = \int_0^\infty e^{-\theta x}\, \mathrm{d}x = \frac{1}{\theta}\,.$$

(vi) Soit X de densité $1/\pi(1+x^2)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$ (loi de Cauchy). Alors X n'admet pas d'espérance, mais admet tout moment absolu d'ordre $p < 1$.

(vii) Si X est de loi $\mathcal{N}(0, 1)$, alors, par symétrie,

$$E(X) = \int_{\mathbb{R}} x e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = 0\,.$$

Donc si X est de loi $\mathcal{N}(m, \sigma^2)$ (*cf.* III.2.6.iv), $E(X) = m$.

(viii) Plus généralement, si X est une variable aléatoire réelle, intégrable, la linéarité de l'intégrale implique

$$E(\sigma X + m) = \sigma E(X) + m$$

pour tous σ, $m \in \mathbb{R}$.

Nous rappelons à présent les inégalités de Jensen (II.2.10), Hölder (II.6.2) et Minkowski (II.6.3) pour des variables aléatoires.

Théorème III.4.5. *(i) (Inégalité de Jensen) Si ϕ est convexe sur $\mathbb{R}$ et si X est une variable aléatoire réelle telle que X et $\phi(X)$ sont intégrables, alors*

$$\phi\big(E(X)\big) \leq E\big(\phi(X)\big) .$$

(ii) (Inégalité de Hölder) Si $X \in \mathrm{L}^p$, $Y \in \mathrm{L}^q$, $p, q \geq 1$ et $p^{-1} + q^{-1} = 1$, alors $XY \in \mathrm{L}^1$ et

$$E\big(|XY|\big) \leq \Big(E\big(|X|^p\big)\Big)^{1/p} \Big(E\big(|Y|^q\big)\Big)^{1/q} .$$

(iii) L'application $p \mapsto \big(E\big(|X|^p\big)\big)^{1/p}$ est croissante.

(iv) $\|\cdot\|_p = (E|\cdot|^p)^{1/p}$ est une norme sur $\mathrm{L}^p(\Omega, \mathcal{A}, P)$, $p \geq 1$.

(v) On définit $\|X\|_\infty = \lim_{p\to\infty} \|X\|_p$. C'est une norme, appelée norme supremum essentiel, sur $\mathrm{L}^\infty(\Omega, \mathcal{A}, P) = \{\, X \,:\, \|X\|_\infty < \infty \,\}$.

Noter que l'inégalité triangulaire dans le point (iv) de ce théorème est équivalente à l'inégalité de Minkowski. Le point (iii) de ce théorème découle aussi bien de l'inégalité de Jensen ou de Hölder. Dans la pratique, l'inégalité de Jensen est le plus souvent utilisée pour les fonctions $\phi(x) = |x|$, x^2 et $1/x$ lorsque $x > 0$. En particulier, pour une variable aléatoire X intégrable, $|E(X)| \leq E(|X|)$; pour une variable aléatoire X dont le carré est intégrable, $\big(E(X)\big)^2 \leq E(X^2)$; pour une variable aléatoire X à valeurs strictement positives, $E(1/X) \geq 1/E(X)$.

La définition suivante décrit une mesure de la dispersion d'une variable aléatoire ou de sa loi.

Définition III.4.6. Soit X une variable aléatoire réelle dont le carré est intégrable. On appelle variance de X, ou de sa loi P^X, et on note $\mathrm{Var}(X)$, la quantité

$$\mathrm{Var}(X) = E\Big(\big(X - E(X)\big)^2\Big) .$$

La racine $\sqrt{\mathrm{Var}(X)}$ est appelée l'écart type, parfois noté $\sigma(X)$. Une variable aléatoire d'écart type 1 est dite réduite.

Une expression équivalente de la variance est

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 .$$

En effet, le développement du carré et la linéarité de l'espérance montrent que

$$\mathrm{Var}(X) = E\big(X^2 - 2XE(X) + E(X)^2\big) = E(X^2) - 2E(X)^2 + E(X)^2$$

puisque l'espérance d'une constante est cette constante.

Une autre écriture de la variance, de contenu plus géométrique, est en terme de norme dans l'espace de Hilbert $\mathrm{L}^2(\Omega, \mathcal{A}, P)$, mesurant la distance de X à son espérance :

$$\sqrt{\mathrm{Var}(X)} = \|X - E(X)\|_2 \,.$$

Les variances s'évaluent comme les espérances à partir du théorème de transport.

Exemples III.4.7. (i) Si $\mathrm{Var}(X) = 0$, alors X est p.s. constante, égale à sa moyenne $E(X)$.

(ii) Si X est de loi de Bernoulli $\mathcal{B}(n,p)$, sa variance est $np(1-p)$.

(iii) Si X suit une loi $\mathcal{N}(0,1)$, $E(X) = 0$ et donc

$$\mathrm{Var}(X) = E(X^2) = \int_{\mathbb{R}} x^2 e^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = 1$$

(intégration par parties). Ceci justifie la terminologie de loi normale centrée réduite pour $\mathcal{N}(0,1)$.

(iv) Si α est un nombre réel, $\mathrm{Var}(X+\alpha) = \mathrm{Var}(X)$ et $\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X)$. En particulier, si X est de loi $\mathcal{N}(m, \sigma^2)$, on a $\mathrm{Var}(X) = \sigma^2$ et $\sigma(X) = \sigma$.

La définition et les exemples (iii)–(iv) montrent que plus la variance est grande, plus la variable aléatoire est “dispersée”, c'est-à-dire prend avec forte probabilité des valeurs éloignées de sa moyenne.

Il est parfois plus commode de calculer une espérance à partir de la fonction de répartition.

Proposition III.4.8. *Soit X une variable aléatoire réelle positive, de fonction de répartition $F = F^X$. Alors, pour tout $0 < p < \infty$,*

$$E(X^p) = p \int_0^\infty t^{p-1} P\{X > t\}\,\mathrm{d}t = p \int_0^\infty t^{p-1}\big(1 - F(t)\big)\,\mathrm{d}t \,.$$

De plus, $E(X) < \infty$ si et seulement si pour un ou tout $\varepsilon > 0$,

$$\sum_{n\in\mathbb{N}} P\{X > \varepsilon n\} < \infty \qquad ou \qquad \sum_{n\in\mathbb{N}} 2^n P\{X > \varepsilon 2^n\} < \infty \,.$$

Démonstration. D'après le théorème de Fubini-Tonelli (II.5.1),

$$\begin{aligned} p\int_0^\infty t^{p-1}P\{X>t\}\,\mathrm{d}t &= p\int_0^\infty t^{p-1}E\big(\mathbb{1}_{]t,\infty[}(X)\big)\,\mathrm{d}t \\ &= E\Big(p\int_0^X t^{p-1}\,\mathrm{d}t\Big) \\ &= E(X^p)\,. \end{aligned}$$

Pour la seconde partie, prenons $p=1$ et notons que

$$\sum_{n\in\mathbb{N}} P\{X>n+1\} \le \int_0^\infty P\{X>t\}\,\mathrm{d}t \le \sum_{n\in\mathbb{N}} P\{X>n\}$$

en découpant l'intégrale sur $[\,0,\infty\,[$ suivant les intervalles $[\,n,n+1\,[$. De la même façon, en découpant cette intégrale suivant les intervalles $[\,2^n,2^{n+1}\,[$,

$$\sum_{n\in\mathbb{N}} 2^n P\{X>2^{n+1}\} \le \int_0^\infty P\{X>t\}\,\mathrm{d}t \le 1+\sum_{n\in\mathbb{N}} 2^n P\{X>2^n\}\,.$$

On conclut en remplaçant X par X/ε. □

Les inégalités suivantes sont essentielles dans l'analyse des variables aléatoires réelles.

***Inégalité de Markov* III.4.9.** *Si X est intégrable et $t>0$, alors*

$$P\{X\ge t\} \le \frac{E(X^+)}{t} \le \frac{E(|X|)}{t}\,.$$

Démonstration. Observer que

$$\mathbb{1}_{[t,\infty[}(X) \le \frac{X}{t}\mathbb{1}_{[t,\infty[}(X) \le \frac{X^+}{t} \le \frac{|X|}{t}$$

et intégrer cette inégalité par rapport à P. □

Cette inégalité est utilisée généralement soit pour X positive, soit pour $|X|$. Elle n'est intéressante que si le second membre est plus petit que 1.

***Exemples* III.4.10.** (i) Si $X\in \mathrm{L}^p$, $p>0$, alors

$$P\{X\ge t\} \le \frac{E(|X|^p)}{t^p}$$

pour tout $t>0$ puisque $\{X\ge t\}\subset\{|X|^p\ge t^p\}$.

(ii) Si $X \in \mathrm{L}^2$, l'inégalité de Markov implique l'inégalité de Tchebitchev

$$P\big\{|X - E(X)| \geq t\big\} \leq \frac{\mathrm{Var}(X)}{t^2}\,, \qquad t > 0\,,$$

puisque $E(|X - E(X)|^2) = \mathrm{Var}(X)$.
(iii) Si maintenant $E(e^{\lambda X}) < \infty$ pour $\lambda > 0$, ou seulement $\lambda \in \,]\,0, \lambda_0\,[\,$, $\lambda_0 > 0$, alors, pour tout $t \in \mathbb{R}$,

$$P\{X \geq t\} \leq \inf_{\lambda} e^{-\lambda t} E(e^{\lambda X})$$

puisque $\{X \geq t\} = \{e^{\lambda X} \geq e^{\lambda t}\}$ pour tout $\lambda > 0$. Autrement dit,

$$P\{X \geq t\} \leq e^{-I(t)}$$

où

$$I(t) = \sup_{\lambda} \big(\lambda t - \ln E(e^{\lambda X})\big)\,, \qquad t \in \mathbb{R}\,.$$

Cette inégalité est l'inégalité de Bernstein, Cramér ou Chernoff. Elle est d'un usage fréquent dans l'étude des sommes de variables aléatoires indépendantes et dans la théorie des grandes déviations.

(iv) Nous présentons un exemple d'application du calcul des probabilités à l'approximation des fonctions. Le théorème de Stone-Weierstrass indique que l'ensemble des polynômes est dense dans l'espace $\mathrm{C}[\,0,1\,]$ des fonctions continues sur $[\,0,1\,]$ à valeurs réelles, muni de la norme uniforme $\|\cdot\|_\infty$. Montrons ce résultat de façon plus constructive.

À une fonction $f \in \mathrm{C}[\,0,1\,]$, nous associons son n-ième polynôme de Bernstein,

$$B_n(f,x) = \sum_{0 \leq k \leq n} f(k/n) \mathrm{C}_n^k \, x^k (1-x)^{n-k}\,, \quad x \in [\,0,1\,]\,.$$

Observons que $B_n(f,x) = E(f(Z/n))$ où Z est de loi binomiale $\mathcal{B}(n,x)$. Notons

$$\omega(f,\delta) = \sup\big\{|f(x) - f(y)|\,,\, x,y \in [\,0,1\,]\,,\, |x-y| \leq \delta\big\}$$

le module de continuité de f sur $[\,0,1]$. Ce module est fini pour tout $\delta > 0$ puisque f est continue sur le compact $[\,0,1\,]$, donc uniformément continue. Pour tout $x \in [\,0,1\,]$,

$$\begin{aligned}
\big|f(x) - B_n(f,x)\big| &= \big|f(x) - E\big(f(Z/n)\big)\big| \\
&\leq E\big(\big|f(x) - f(Z/n)\big|\big) \\
&\leq \omega(f,\delta) P\big\{|Z/n - x| \leq \delta\big\} + 2\|f\|_\infty P\big\{|Z/n - x| \geq \delta\big\}\,.
\end{aligned}$$

En utilisant l'inégalité de Tchebitchev (4.10.ii) pour la variable Z de moyenne $E(Z) = nx$ et de variance $\mathrm{Var}(Z) = nx(1-x)$, il vient

$$\begin{aligned}\big|f(x) - B_n(f,x)\big| &\leq \omega(f,\delta) + 2\|f\|_\infty \frac{1}{n^2\delta^2}\mathrm{Var}(Z)\\ &\leq \omega(f,\delta) + 2\|f\|_\infty \frac{1}{n\delta^2}\, x(1-x)\\ &\leq \omega(f,\delta) + \frac{\|f\|_\infty}{2n\delta^2}\,.\end{aligned}$$

Donc,

$$\sup_{0\leq x\leq 1} \big|f(x) - B_n(f,x)\big| \leq \inf_{\delta>0}\Big(\omega(f,\delta) + \frac{\|f\|_\infty}{2n\delta^2}\Big) = o(1) \qquad (n\to\infty)\,,$$

ce qui démontre le théorème de Weierstrass. En particulier, si f est höldérienne d'indice α, $\omega(f,\delta) = c\delta^\alpha$, $0 < \alpha \leq 1$, on obtient

$$\sup_{0\leq x\leq 1} \big|f(x) - B_n(f,x)\big| \leq (2c)^{2/(\alpha+2)}\|f\|_\infty^{\alpha/(\alpha+2)} n^{-\alpha/(\alpha+2)}\,.$$

Nous concluons ce paragraphe par les définitions d'espérance et de variance pour des vecteurs aléatoires.

Définition III.4.11. Si $X = (X_1,\ldots,X_d) \in \mathbb{R}^d$ est un vecteur aléatoire sur un espace probabilisé $(\Omega,\mathcal{A},P)$, on dit que X est de puissance p-ième intégrable ($p > 0$) si chacune de ses composantes l'est, ce qui équivaut à dire que $E(\|X\|^p) < \infty$, où $\|X\|$ est ici la norme euclidienne $(X_1^2 + \cdots + X_d^2)^{1/2}$ du vecteur X. Son espérance est le vecteur de $\mathbb{R}^d$

$$E(X) = \big(E(X_1),\ldots,E(X_d)\big)\,.$$

Sa (matrice carrée de) covariance est

$$\mathrm{Cov}(X) = \Big(E\big((X_i - E(X_i))(X_j - E(X_j))\big)\Big)_{1\leq i,j\leq d}\,.$$

À la variance se substitue à présent une matrice. C'est une matrice symétrique semi-définie positive puisque pour tous réels $\alpha_1,\ldots,\alpha_d$,

$$\sum_{1\leq i,j\leq d} \alpha_i\alpha_j E\big((X_i - E(X_i))(X_j - E(X_j))\big) = E\bigg[\Big(\sum_{1\leq i\leq d}\alpha_i(X_i - E(X_i))\Big)^2\bigg] \geq 0\,.$$

Elle est définie positive si aucune combinaison linéaire des composantes du vecteur aléatoire n'est p.s. constante.

Exemple III.4.12. Soit X la variable aléatoire de loi décrite dans l'exemple III.3.3.ii. Il est aisé de vérifier que le vecteur moyenne de X est le vecteur nul (de $\mathbb{R}^d$), et que sa matrice de covariance est la matrice identité (de $\mathbb{R}^d$). On parlera plus loin de vecteurs gaussiens de moyenne m et de matrice de covariance Γ, où m est un vecteur quelconque de $\mathbb{R}^d$ et Γ est une matrice symétrique semi-définie positive (d, d).

III.5. Fonctions caractéristiques

Nous savons que la fonction de répartition d'une variable aléatoire réelle ou vectorielle X sur $(\Omega, \mathcal{A}, P)$ caractérise sa loi. Autrement dit, sur $\mathbb{R}$ par exemple, la donnée de

$$F^X(t) = E\big(\mathbb{1}_{]-\infty,t]}(X)\big)\,, \quad t \in \mathbb{R}\,,$$

détermine la loi de X. Puisque les indicatrices sont des fonctions boréliennes bornées, la donnée de $E(\phi(X))$ pour toute fonction borélienne bornée ϕ caractérise la loi P^X. La fonction indicatrice $\mathbb{1}_{]-\infty,t]}$ peut être approchée par la suite de fonctions continues bornées

$$\phi_n(x) = \begin{cases} 1 & \text{si } x \leq t, \\ 1 + n(t - x) & \text{si } t \leq x \leq t + \frac{1}{n}, \\ 0 & \text{si } x > t + \frac{1}{n} \end{cases}$$

(faire un dessin). Il s'ensuit, d'après le théorème de convergence dominée, que la donnée de $E(\phi(X))$ pour toute fonction continue bornée sur $\mathbb{R}$ caractérise P^X. Plus généralement, les fonctions indicatrices peuvent être approchées simplement par des fonctions C^∞ bornées ; et donc la donnée de $E(\phi(X))$ pour toute fonction ϕ infiniment dérivable caractérise également P^X. On pourrait même se restreindre aux fonctions C^∞ à support compact ! (*cf.* exercice II.5).

Ces raisonnements et conclusions s'appliquent de la même façon aux vecteurs aléatoires.

Une autre caractérisation intéressante en pratique (voir IV.2, V.4 et V.5) est celle des fonctions caractéristiques, ou transformées de Fourier, qui remplace la classe des fonctions C^∞ bornées par la famille des fonctions sinus et cosinus.

Définition III.5.1. Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$. On appelle fonction caractéristique de X ou de la loi de X, ou transformée de

Fourier, et on note φ^X, la fonction à valeurs complexes

$$\begin{aligned} t \in \mathbb{R}^d \mapsto \varphi^X(t) = E(e^{i\langle t,X\rangle}) &= \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \,\mathrm{d}P^X(x) \\ &= \int_{\mathbb{R}^d} \cos\langle t,x\rangle \,\mathrm{d}P^X(x) + i \int_{\mathbb{R}^d} \sin\langle t,x\rangle \,\mathrm{d}P^X(x)\,. \end{aligned}$$

La fonction caractéristique est à valeurs complexes, de module majoré par 1 (d'après l'inégalité de Jensen), et $\varphi^X(0) = 1$. Si la loi de X a une densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, alors

$$\varphi^X(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} f(x) \,\mathrm{d}x$$

est aussi appelée la transformée de Fourier de la fonction f.

Comme son nom l'indique, la fonction caractéristique caractérise la loi.

Théorème III.5.2. *Si X et Y sont deux vecteurs aléatoires de lois P^X et P^Y telles que $\varphi^X = \varphi^Y$, alors $P^X = P^Y$.*

Démonstration. La démonstration utilise le théorème des classes monotones fonctionnelles (I.3.5). On note $e_1, \ldots, e_d$ la base canonique de $\mathbb{R}^d$. Pour tout $t \in \mathbb{R}^d$, l'égalité des parties réelles (resp. imaginaires) de φ^X et φ^Y donne $E(\cos\langle t,X\rangle) = E(\cos\langle t,Y\rangle)$ (resp. $E(\sin\langle t,X\rangle) = E(\sin\langle t,Y\rangle)$). Notons $\mathscr{C}$ l'ensemble des combinaisons linéaires finies des fonctions $x \mapsto \cos\langle t,x\rangle$ et $x \mapsto \sin\langle t,x\rangle$. En particulier, la fonction $x \mapsto n\sin\langle e_i/n, x\rangle$ appartient à $\mathscr{C}$ et sa limite simple, la projection sur la i-ème coordonnée, est mesurable par rapport à la tribu $\sigma(\mathscr{C})$ engendrée par $\mathscr{C}$. Donc $\sigma(\mathscr{C}) = \mathcal{B}(\mathbb{R}^d)$ (*cf.* I.1.12.iii).

Soit maintenant $\mathcal{H}$ l'espace vectoriel des fonctions boréliennes bornées ϕ telles que $E(\phi(X)) = E(\phi(Y))$. L'espace $\mathcal{H}$ contient les constantes et est stable par convergence monotone bornée (d'après le théorème de convergence monotone II.2.1). De plus $\mathscr{C} \subset \mathcal{H}$ et $\mathscr{C}$ est stable par multiplication (linéariser un produit de sinus et de cosinus). Le théorème des classes monotones fonctionnelles (I.3.5) montre alors que $\mathcal{H}$ contient toute fonction bornée mesurable par rapport à $\sigma(\mathscr{C}) = \mathcal{B}(\mathbb{R}^d)$, donc toute fonction borélienne. Le résultat s'ensuit. ☐

Exemples III.5.3. (i) Si $X = a$ p.s., *i.e.* $P^X = \delta_a$, $a \in \mathbb{R}^d$, alors $\varphi^X(t) = e^{i\langle t,a\rangle}$.
(ii) Si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, Σ une matrice opérant sur $\mathbb{R}^d$ et $m \in \mathbb{R}^d$, alors $Y = \Sigma X + m$ est un vecteur aléatoire de fonction caractéristique

$$\varphi^Y(t) = e^{i\langle t,m\rangle} \varphi^X({}^{\mathrm{t}}\Sigma t)$$

puisque $\langle t, \Sigma X + m\rangle = \langle {}^{\mathrm{t}}\Sigma t, X\rangle + \langle t,m\rangle$.

(iii) Si X suit une loi normale $\mathcal{N}(0,1)$, alors

$$\varphi^X(t) = E\big(e^{i\langle t,X\rangle}\big) = \int_{\mathbb{R}} e^{itx-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}} = e^{-t^2/2} .$$

Une méthode pour calculer cette intégrale est donnée dans l'exercice III.12.

Si Y est de loi $\mathcal{N}(m,\sigma^2)$, Y a la même loi que $\sigma X + m$, et donc

$$\varphi^Y(t) = E\big(e^{it(\sigma X+m)}\big) = e^{itm-\sigma^2t^2/2} .$$

(iv) Si X est de loi exponentielle de densité e^{-x} sur $\mathbb{R}^+$, alors

$$\varphi^X(t) = \int_{\mathbb{R}^+} e^{(it-1)x}\,\mathrm{d}x = \frac{1}{1-it} .$$

(v) Si X est de loi de Poisson de paramètre λ,

$$P\{X=k\} = e^{-\lambda}\lambda^k/k! , \qquad k \in \mathbb{N} ,$$

alors

$$\varphi^X(t) = \sum_{k\in\mathbb{N}} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k\in\mathbb{N}} \frac{(e^{it}\lambda)^k}{k!} = \exp\big(\lambda(e^{it}-1)\big) .$$

(vi) Si X est de loi binomiale de paramètres n et p,

$$P\{\,X=k\,\} = \mathrm{C}_n^k\, p^k(1-p)^{n-k} , \qquad 0 \leq k \leq n ,$$

alors

$$\varphi^X(t) = \sum_{0\leq k\leq n} \mathrm{C}_n^k (e^{itp})^k (1-p)^{n-k} = (1-p+pe^{it})^n .$$

(vii) Si $X=(X_1,\ldots,X_d)$ est un vecteur aléatoire dont la loi est le produit des lois marginales, $P^{X_1}\otimes\cdots\otimes P^{X_d}$, alors $\varphi^X(t) = \varphi^{X_1}(t_1)\cdots\varphi^{X_d}(t_d)$. (Utiliser le théorème de Fubini, II.5.1).

Puisque la transformée de Fourier caractérise la loi, il est souhaitable d'avoir une formule d'inversion permettant d'obtenir effectivement la loi à partir de la fonction caractéristique. Il existe plusieurs formules de ce type permettant de calculer la densité si elle existe, ou la fonction de répartition (voir exercice V.9). En voici une possible.

Théorème III.5.4 (Formule d'inversion de Fourier). *Soit φ^X la fonction caractéristique d'un vecteur aléatoire X, supposée intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la loi de X admet une densité continue bornée f^X par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, donnée, pour tout $x\in\mathbb{R}^d$, par*

$$f^X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle} \varphi^X(t)\,\mathrm{d}t .$$

Démonstration. Voir exercice V.9. □

***Exemples* III.5.5.** (i) Si $f(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$, densité de la loi dite de Laplace, sa transformée de Fourier est $\varphi(t) = 1/(1+t^2)$. Pour le montrer, on se reporte à l'exemple III.5.3.iv, et on remarque que, par symétrie, la transformée de Fourier de f est

$$\frac{1}{2} \cdot \frac{1}{1-it} + \frac{1}{2} \cdot \frac{1}{1+it} = \frac{1}{1+t^2}\,.$$

(ii) Soit $f(x) = 1/\pi(1+x^2)$, $x \in \mathbb{R}$, la densité de la loi dite de Cauchy, sur $\mathbb{R}$. En utilisant l'exemple précédent et le théorème III.5.4, il vient

$$\frac{1}{2}e^{-|x|} = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\frac{1}{1+t^2}\,\mathrm{d}t\,.$$

En changeant t et $-t$ dans l'intégrale, on constate que la tranformée de Fourier de f est $e^{-|t|}$.

Lorsque X est une variable aléatoire réelle, $e^{itX} = \sum_{n\in\mathbb{N}}(it)^n X^n/n!$; en intégrant terme à terme (nous verrons plus loin des conditions permettant de le justifier),

$$\varphi^X(t) = \sum_{n\in\mathbb{N}} \frac{(it)^n}{n!}E(X^n)\,.$$

La formule de Taylor montre alors que les moments de la variable sont proportionnels aux dérivées de la transformée de Fourier. Le résultat rigoureux est le suivant.

***Proposition* III.5.6.** *Soit X une variable aléatoire réelle, de fonction caractéristique $\varphi = \varphi^X$ et de loi P^X.*
(i) Si $E(|X|^n) < \infty$, alors φ est n-fois dérivable, de dérivée k-ième $(k \leq n)$

$$\varphi^{(k)}(t) = i^k \int x^k e^{itx}\,\mathrm{d}P^X(x) = i^k E(X^k e^{itX})\,.$$

En particulier, $\varphi^{(k)}(0) = i^k E(X^k)$.
(ii) Réciproquement, si n est pair et si φ est n-fois dérivable en 0, alors X admet tout moment d'ordre plus petit ou égal à n.

Démonstration. (i) L'inégalité, pour tout $u \in \mathbb{R}$ et tout $n \in \mathbb{N}$,

$$\left| e^{iu} - 1 - \frac{iu}{1!} - \dots - \frac{(iu)^{n-1}}{(n-1)!} \right| \leq \frac{|u|^n}{n!}\,,$$

se démontre en remarquant que $f_1(u) = i\int_0^u e^{ix}\,\mathrm{d}x$ est de module plus petit que $|u|$, et par récurrence $f_n(u) = i\int_0^u f_{n-1}(x)\,\mathrm{d}x$ est de module plus petit que $|u|^n/n!$. Démontrons pour commencer que φ est dérivable en tout point $t \in \mathbb{R}$ lorsque $E(|X|) < \infty$. Pour tout $h \neq 0$,

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \int e^{itx}\,\frac{e^{ihx}-1}{h}\,\mathrm{d}P^X(x)\,.$$

D'après l'inégalité précédente pour $n = 1$,

$$\Big|e^{itx}\,\frac{e^{ihx}-1}{h}\Big| \leq |x|$$

qui est intégrable pour P^X indépendamment de h. D'après le théorème de convergence dominée (II.2.8),

$$\varphi'(t) = \lim_{h\to 0}\int e^{itx}\,\frac{e^{ihx}-1}{h}\,\mathrm{d}P^X(x) = \int ixe^{itx}\,\mathrm{d}P^X(x) = iE(Xe^{itX})\,.$$

Les dérivées d'ordre supérieur se calculent de la même façon.
(ii) Démontrons par récurrence que $E(X^{2k})$ est fini dès que $2k \leq n$, ce qui suffit en vertu du théorème III.4.5.iii. La propriété est vraie pour $k = 0$. Supposons-la montrée pour $k-1$. Par hypothèse, la limite

$$\lim_{h\to 0}\frac{1}{h^2}\big(\varphi^{(2k-2)}(h) + \varphi^{(2k-2)}(-h) - 2\varphi^{(2k-2)}(0)\big)$$

existe et est égale à $\varphi^{(2k)}(0)$. Comme, d'après le point (i), pour tout h réel,

$$\varphi^{(2k-2)}(h) = (-1)^{k-1}\int x^{2k-2}e^{ihx}\,\mathrm{d}P^X(x)\,,$$

on a

$$(-1)^{k-1}\varphi^{(2k)}(0) = \lim_{h\to 0}\frac{2}{h^2}\int x^{2k-2}\big(\cos(hx)-1\big)\,\mathrm{d}P^X(x)\,.$$

Utiliser le lemme de Fatou (II.2.3) et la limite $\lim_{h\to 0}\big(1-\cos(hx)\big)/h^2 = x^2/2$ pour conclure que

$$\int x^{2k}\,\mathrm{d}P^X(x) \leq (-1)^k\varphi^{(2k)}(0) < \infty\,. \qquad \square$$

Application III.5.7. Remarquons qu'en général une loi n'est pas caractérisée par ses moments (exercice III.7). Toutefois, si $\varphi = \varphi^X$ est analytique, la proposition III.5.6 et le théorème III.5.2 montrent que la loi P^X est caractérisée par ses moments. Une condition simple pour que ceci ait lieu est de supposer que $E(e^{\alpha|X|}) < \infty$ pour un $\alpha > 0$. En effet, en intégrant l'inégalité utilisée pour démontrer III.5.6.i et en utilisant III.5.6.i, il vient

$$\left| \varphi(t+h) - \varphi(t) - \frac{h}{1!}\varphi^{(1)}(t) - \cdots - \frac{h^{n-1}}{(n-1)!}\varphi^{(n-1)}(t) \right| \leq E\big(|X|^n\big)\frac{|h|^n}{n!},$$

ce qui fournit l'analyticité de $h \mapsto \varphi(t+h)$ sur $]-\alpha, \alpha[$. Ceci ayant lieu pour chaque réel t, de proche en proche, φ est analytique sur tout $\mathbb{R}$. Un exemple important est le cas particulier des lois concentrées sur un intervalle borné de $\mathbb{R}$. Ceci est résumé dans le théorème dit des moments.

Théorème III.5.8 **(des moments).** *Soient X, Y deux variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$ à valeurs dans un intervalle borné $[a, b]$. Si $E(X^k) = E(Y^k)$ pour tout $k \in \mathbb{N}$, alors X et Y ont même loi.*

La définition suivante est une variante de celle des fonctions caractéristiques. Elle impose cependant des conditions d'intégrabilité sur la loi de la variable aléatoire.

Définition III.5.9. Si X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$, on appelle transformée de Laplace (ou fonction génératrice des moments) la fonction $L^X(s) = E\big(e^{\langle s, X\rangle}\big)$ définie pour les valeurs de s pour lesquelles $e^{\langle s, X\rangle}$ est intégrable.

La transformée de Laplace, si elle est finie dans un voisinage de 0, caractérise la loi, comme la transformée de Fourier (pour les lois sur $\mathbb{R}^+$, la démonstration est analogue à celle du théorème III.5.2, en remplaçant les fonctions $\sin(tx)$ et $\cos(tx)$ par e^{tx} et en remarquant que la fonction identité est limite simple de combinaisons linéaires d'exponentielles de petits paramètres, puisque $x = \lim_{t\to 0} t^{-1}(e^{tx} - 1)$ et $1 = e^{0x}$!). Il existe des formules d'inversion de la transformée de Laplace (voir par exemple l'exercice V.8).

On peut donner un énoncé analogue à la proposition III.5.6 justifiant le nom de fonction génératrice des moments.

Proposition III.5.10. *Soit X une variable aléatoire réelle telle que e^{tX} est intégrable pour t dans un intervalle ouvert contenant 0. Alors la transformée de Laplace L^X est définie sur un intervalle ouvert contenant 0. De plus elle est analytique dans un voisinage de 0 et*

$$L^X(t) = \sum_{n\in\mathbb{N}} \frac{t^n}{n!} E(X^n)$$

pour tout t dans ce voisinage. En particulier, pour tout $n \in \mathbb{N}$,

$$(L^X)^{(n)}(0) = E(X^n)\,.$$

Démonstration. Supposons L^X définie sur $]-\varepsilon, \varepsilon[$ pour un $\varepsilon > 0$. Puisque

$$\sum_{n\in\mathbb{N}} \frac{|tx|^n}{n!} = e^{|tx|} \leq e^{tx} + e^{-tx}\,,$$

le théorème de convergence dominée II.2.8 montre que pour tout $|t| < \varepsilon$,

$$L^X(t) = E(e^{tX}) = \sum_{n\in\mathbb{N}} \frac{1}{n!} E\big((tX)^n\big) = \sum_{n\in\mathbb{N}} \frac{t^n}{n!} E(X^n)\,,$$

ce qui démontre l'analycité de L^X dans un voisinage de 0. $\square$

Exercices

Exercice III.1. Un tiroir contient n paires de chaussures. On choisit au hasard $2r$ chaussures ($2r \leq n$). Quelle est la probabilité qu'il n'y ait parmi ces $2r$ chaussures aucune paire complète ? Quelle est la probabilité qu'il y ait exactement k paire(s) complète(s) ($1 \leq k \leq r$) ?

Exercice III.2. Soit X une variable aléatoire à valeurs dans un ensemble M muni de la tribu de ses parties, telle que $P\{X = x\} > 0$ pour tout $x \in M$. Montrer que M est fini ou dénombrable.
Indication : Pour tout $n \geq 1$, soit $M_n = \{x \in M \,:\, P\{X = x\} > 1/n\}$. Montrer que M_n est fini.

Exercice III.3. (Paradoxe de Bertrand). Soit $\mathcal{C}$ le cercle de centre O et de rayon 1 dans $\mathbb{R}^2$. On cherche à déterminer la probabilité pour que la corde AB de ce cercle, choisie “au hasard”, soit plus grande que le côté du triangle équilatéral inscrit dans le cercle. Faire le calcul dans les différents cas suivants :
a) On fixe un point I du cercle ; on choisit un point M sur le segment OI selon la probabilité uniforme ; on lui associe la corde AB perpendiculaire à OI et passant par M.
b) On fixe A sur le cercle et on choisit B selon la probabilité uniforme sur le cercle.
c) On choisit M dans le disque selon la probabilité uniforme ; AB est alors la corde passant par M et perpendiculaire à OM.

Exercice III.4. La plupart des ordinateurs disposent d'un algorithme permettant de simuler des variables aléatoires uniformes sur $[0,1]$. Supposons donc savoir tirer une variable aléatoire de loi $\mathcal{U}_{[0,1]}$. Utiliser la proposition III.2.7 pour simuler une variable aléatoire de loi

(i) $\mathcal{E}xp(1)$,

(ii) de fonction de répartition $F(x) = 1 - x^{-\alpha}$ si $x \geq 1$, et $F(x) = 0$ si $x \leq 1$ (loi de Paréto),

(iii) de Cauchy de densité $1/\pi(1+x^2)$.

Exercice III.5. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ telle que

$$P\{X = k\} = \frac{e^{-2}2^k}{4k!}(1 + \alpha k)\,, \quad k \in \mathbb{N}\,,$$

où $\alpha > 0$. Déterminer la valeur de α. Calculer l'espérance et la variance de X en remarquant que

$$P\{X = k\} = \frac{1}{4}P\{Y = k\} + \frac{3}{4}P\{T = k\}$$

pour tout k, où $T = Z + 1$ et Y et Z sont deux variables de loi de Poisson de paramètre 2.

Exercice III.6. Soit Ω l'ensemble des $n!$ permutations σ des entiers de 1 à n muni de la probabilité uniforme. Soient $\{c_1, \ldots, c_n\}$ et $\{u_1, \ldots, u_n\}$ des nombres réels. On définit $S(\sigma) = \sum_{1\leq k\leq n} c_k u_{\sigma(k)}$. Posons

$$\begin{aligned} \overline{c} &= \frac{1}{n}\textstyle\sum_{1\leq k\leq n} c_k\,, & \overline{u} &= \frac{1}{n}\textstyle\sum_{1\leq k\leq n} u_k\,,\\ s_c^2 &= \frac{1}{n-1}\textstyle\sum_{1\leq k\leq n}(c_k - \overline{c})^2\,, & s_u^2 &= \frac{1}{n-1}\textstyle\sum_{1\leq k\leq n}(u_k - \overline{u})^2\,. \end{aligned}$$

a) Montrer que l'espérance de S est égale à $n\overline{c}\,\overline{u}$.
b) Calculer la variance de $u_{\sigma(k)}$, puis la covariance de $u_{\sigma(k)}$ et $u_{\sigma(l)}$ $(k \neq l)$.
Indication : Noter que $\sum_{1\leq k\leq n} u_{\sigma(k)} = \sum_{1\leq k\leq n} u_k$.
c) Déterminer la variance de S en fonction de s_c^2 et s_u^2.

Exercice III.7. Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$. Montrer que $Z = e^X$ est de densité $f^Z(z) = (2\pi)^{-1/2}z^{-1}e^{-(\ln z)^2/2}$ si $z > 0$ et $f^Z(z) = 0$ si $z \leq 0$. La loi de Z s'appelle la loi log-normale.

Pour $a \in [-1,1]$, soit $f_a(x) = f^Z(x)(1 + a\sin(2\pi \ln x))$, $x > 0$. Montrer que si Z_a est de densité f_a, alors Z_a et Z ont mêmes moments, et donc que les moments ne caractérisent pas une loi de probabilité (comparer avec III.5.7 et le théorème III.5.8).

Exercice III.8. On dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est échangeable si la loi de X est invariante par permutation des coordonnées, *i.e.* pour toute permutation π de $\{1, 2, \ldots d\}$, X a même loi que $(X_{\pi(1)}, \ldots, X_{\pi(d)})$. Soit donc X un tel vecteur aléatoire, échangeable, de carré intégrable, tel que de plus $X_1 + \cdots + X_d = 1$. Montrer qu'alors $E(X_i) = 1/d$ et

$$\mathrm{Cov}(X_i, X_j) = -\frac{\mathrm{Var} X_1}{d-1}, \qquad i \neq j\,.$$

Indication : étudier $E(X_1 + \cdots + X_d)$ et $E(X_1(X_1 + \cdots + X_d))$.

Exercice III.9. Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, P)$.
a) On suppose que X est de carré intégrable. Démontrer qu'il existe un unique réel x_0 tel que la fonction $g(x) = E((X - x)^2)$ soit minimum en ce point. Déterminer x_0 et $g(x_0)$.
b) On appelle médiane de X un réel m tel que

$$P\{X \geq m\} \geq 1/2 \quad \text{et} \quad P\{X \leq m\} \geq 1/2\,.$$

Démontrer qu'un tel réel existe toujours, mais qu'il n'est pas nécessairement unique. Prouver que si X est intégrable et m est une médiane de X,

$$E\big(|X - m|\big) = \inf\big\{ E\big(|X - \alpha|\big) \,:\, \alpha \in \mathbb{R} \big\}\,.$$

Indication : Établir que si $a < b$,

$$E\big(|X - b|\big) - E\big(|X - a|\big) = \int_a^b \psi(x)\,\mathrm{d}x$$

où $\psi(x) = P\{X \leq x\} - P\{X \geq x\}$ et étudier le signe de la fonction ψ.

Exercice III.10. Soit X une variable aléatoire positive de carré intégrable sur $(\Omega, \mathcal{A}, P)$ et soit $\lambda \in \,]\,0, 1\,[$. Démontrer que

$$(1 - \lambda)E(X) \leq E\big(X \mathbb{1}_{[\lambda E(X), \infty[}(X)\big)\,,$$

et en déduire, par l'inégalité de Cauchy-Schwarz, que

$$P\big\{X \geq \lambda E(X)\big\} \geq (1 - \lambda)^2 \frac{E(X)^2}{E(X^2)}\,.$$

Exercice III.11. Si P est une mesure de probabilité sur $\{1, 2, \ldots, n\}$, on définit l'entropie de P par $H(P) = -\sum_{1 \leq k \leq n} p_k \ln p_k$ où $p_k = P(\{k\})$, avec la convention $0 \ln 0 = 0$. Montrer que H est à valeurs dans $\mathbb{R}^+$ et trouver P telle que $H(P) = 0$. Démontrer que la mesure uniforme sur $\{1, 2, \ldots, n\}$ réalise le maximum de H.

Si P est une mesure de probabilité sur $\mathbb{N}$, on définit de même son entropie par $H(P) = -\sum_{n\in\mathbb{N}} p_n \ln p_n$. Montrer que H est à valeurs dans $\mathbb{R}^+ \cup \{\infty\}$. Quand s'annule-t-elle ? Démontrer que la loi géométrique de paramètre p, $0 < p < 1$, réalise le maximum d'entropie sur l'ensemble des mesures de probabilité sur $\mathbb{N}$ de moyenne inférieure ou égale à $p/(1-p)$.

Si P est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de densité f par rapport à la mesure de Lebesgue, on note $H(P) = \int f(x) \ln f(x)\, \mathrm{d}x$ lorsque cette intégrale a un sens, $H(P) = \infty$ sinon. Calculer l'entropie de la loi normale $\mathcal{N}(0,1)$. Démontrer qu'elle minimise l'entropie de toute mesure de densité f vérifiant $\int_{\mathbb{R}} x f(x)\, \mathrm{d}x = 0$ et $\int_{\mathbb{R}} x^2 f(x)\, \mathrm{d}x = 1$.
Indication : on pourra commencer par montrer que pour toute densité g, $\int \ln(f(x)/g(x)) f(x)\, \mathrm{d}x \geq 0$, puis prendre pour g la densité gaussienne.

Exercice III.12. Montrer que la fonction $\varphi(t) = (2\pi)^{-1/2} \int_{\mathbb{R}} e^{itx - x^2/2}\, \mathrm{d}x$, $t \in \mathbb{R}$, est solution d'une équation différentielle du premier ordre. En déduire la fonction caractéristique de la loi $\mathcal{N}(0,1)$ (*cf.* III.5.3.iii) ainsi que tous les moments de la loi $\mathcal{N}(0,1)$.

Exercice III.13. (Lemme de Riemann-Lebesgue) Soit X une variable aléatoire réelle, de densité f. Montrer que $\lim_{t\to\infty} \varphi^X(t) = 0$.
Indication : on pourra considérer d'abord une densité uniforme, de la forme $\mathbb{1}_{[a,b]}/(b-a)$, puis une densité étagée, et approcher dans L^1 une densité quelconque par une densité étagée.

En déduire que si f admet des dérivées $f^{(1)}, \ldots, f^{(k)}$ intégrables, alors $|\varphi^X(t)| = o(|t|^{-k})$ lorsque $t \to \infty$.

Exercice III.14. Soit P la mesure de probabilité sur $\mathbb{Z}$ définie par

$$P = \sum_{n\geq 2} \frac{c}{n^2 \ln n} (\delta_n + \delta_{-n})$$

où c est la constante de normalisation faisant de P une probabilité. Cette mesure admet-elle un moment d'ordre 1 ? Soit φ la transformée de Fourier de la mesure P. Pour tout entier $N \geq 2$ et tout $t > 0$, on définit

$$f_N(t) = \sum_{2\leq n\leq N} \frac{\sin^2(nt/2)}{tn^2 \ln n}, \qquad g_N(t) = \sum_{n>N} \frac{\sin^2(nt/2)}{tn^2 \ln n}.$$

Démontrer que $f_N(t) \leq tN$ et que $g_N(t) \leq 1/tN \ln N$. Trouver une fonction $t \mapsto N(t)$ de $]\,0,\infty\,[$ dans $\mathbb{N}$ telle que $\lim_{t\to 0} f_{N(t)}(t) = \lim_{t\to 0} g_{N(t)}(t) = 0$. En déduire que φ est dérivable en 0.

Exercice III.15. Soit f une densité sur $\mathbb{R}$, paire (*i.e.* $f(x) = f(-x)$), de fonction caractéristique φ. Pour $x > 0$, soit $g(x) = \int_x^\infty t^{-1} f(t)\,\mathrm{d}t$ et poser $g(-x) = g(x)$. Montrer que g est une densité dont la fonction caractéristique est $t^{-1}\int_0^t \varphi(s)\,\mathrm{d}s$.

IV

INDÉPENDANCE

Dans ce chapitre, nous débutons véritablement les considérations probabilistes.

Si on jette deux fois un dé, le résultat du second jet est intuitivement indépendant du premier. Nous allons formaliser cette intuition, ce qui nous permettra d'évaluer la probabilité de certains événements.

L'indépendance est aux probabilités ce que sont les mesures produit à la théorie de la mesure. En particulier, les sommes de variables aléatoires indépendantes ont pour loi les produits de convolution de mesures. Cette description permet de développer une intuition claire des phénomènes aléatoires modélisés par des répétitions indépendantes d'épreuves (loi des grands nombres, théorème central limite).

IV.1. Indépendance

Définition IV.1.1. Sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, deux événements A, B sont dit indépendants si

$$P(A \cap B) = P(A)P(B)\,.$$

Exemple IV.1.2. On jette deux dés, un bleu et un rouge. Les événements

$$A = \{\ \text{on obtient un nombre inférieur ou égal à 4 avec le dé rouge}\ \}$$

et

$$B = \{\ \text{on obtient un 6 avec le dé bleu}\ \}$$

sont intuitivement indépendants, puisque les deux jets le sont. Nous pouvons modéliser le tirage des deux dés en prenant

$$\Omega = \{\, (i,j) \,:\, 1 \leq i,j \leq 6 \,\}$$

muni de la tribu de ses parties et de la probabilité uniforme P. Clairement, $P(A) = 2/3$ et $P(B) = 1/6$. Observons que

$$A \cap B = \{\, (1,6), (2,6), (3,6), (4,6) \,\}$$

est de probabilité $4/36 = 1/9$, qui est bien le produit de $P(A)$ et $P(B)$.

Remarquons que si deux événements A et B sont indépendants, les tribus $\sigma(\{A\}) = \{\, \emptyset, A, A^{\mathrm{c}}, \Omega \,\}$ et $\sigma(\{B\})$ sont indépendantes au sens où tout élément de $\sigma(\{A\})$ est indépendant de tout élément de $\sigma(\{B\})$. Démontrons par exemple que A et B^{c} sont indépendants. En effet,

$$\begin{aligned} P(A \cap B^{\mathrm{c}}) = P(A) - P(A \cap B) &= P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(B^{\mathrm{c}})\,. \end{aligned}$$

La définition suivante amplifie cette première idée intuitive de l'indépendance dans deux directions, d'une part pour des familles quelconques d'événements, d'autre part pour des tribus.

Définition IV.1.3. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

(i) Une famille quelconque d'événements $A_i \in \mathcal{A}$, $i \in I$, est mutuellement indépendante si pour tout $J \subset I$ fini

$$P\Big(\bigcap_{j \in J} A_j \Big) = \prod_{j \in J} P(A_j)\,.$$

(ii) Une famille quelconque de sous-tribus (ou d'algèbres) $\mathcal{A}_i \subset \mathcal{A}$, $i \in I$, est mutuellement indépendante si toute famille d'événements $A_i \in \mathcal{A}_i$, $i \in I$, est mutuellement indépendante.

Exemples IV.1.4. (i) Prenons $\Omega = [\,0,1\,]$ muni de sa tribu borélienne et P la mesure de Lebesgue sur $[\,0,1\,]$. Soit, pour tout $n \geq 1$,

$$A_n = \bigcup_{1 \leq k \leq 2^{n-1}} \Big] \frac{2(k-1)}{2^n}, \frac{2k-1}{2^n} \Big]\,.$$

La famille $(A_n)_{n \geq 1}$ est mutuellement indépendante (exercice IV.3).

(ii) Reprenons l'exemple du jet de dés. Considérons les événements

$$\begin{aligned} A &= \{\text{ le résultat du dé rouge est impair }\}\,, \\ B &= \{\text{ le résultat du dé bleu est impair }\}\,, \\ C &= \{\text{ la somme des deux dés est impaire }\}\,. \end{aligned}$$

Il est facile de constater que A, B, C sont indépendants deux à deux (c'est-à-dire A et B sont indépendants, B et C sont indépendants, A et C sont indépendants), mais ne sont pas mutuellement indépendants au sens de la définition précédente. En effet,

$$P(A) = P(B) = P(C) = 1/2\,,$$

$$P(A \cap B) = 1/4\ , \qquad P(A \cap C) = 1/4\ , \qquad P(B \cap C) = 1/4\,,$$

alors que $P(A \cap B \cap C) = 0$ car $A \cap B \cap C = \emptyset$ (la somme des dés ne peut être impaire si chacun des deux dés affiche un résultat impair).

(iii) Soient $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, 2$, des espaces probabilisés. En identifiant tout ensemble A_1 de $\mathcal{A}_1$ avec $A_1 \times \Omega_2$ et tout ensemble A_2 de $\mathcal{A}_2$ avec $\Omega_1 \times A_2$, les tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ deviennent des sous-tribus de $\mathcal{A}_1 \otimes \mathcal{A}_2$. Les tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ sont alors indépendantes dans $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$. En effet, observons que $(A_1 \times \Omega_2) \bigcap (\Omega_1 \times A_2) = A_1 \times A_2$ et que par définition de la mesure produit,

$$\begin{aligned} P_1 \otimes P_2\big((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)\big) &= P_1 \otimes P_2(A_1 \times A_2) \\ &= P_1(A_1)P_2(A_2) \\ &= P_1 \otimes P_2(A_1 \times \Omega_2)\, P_1 \otimes P_2(\Omega_1 \times A_2)\,. \end{aligned}$$

Nous convenons pour la suite que l'indépendance d'une famille signifie l'indépendance mutuelle des événements ou des tribus. Toute autre forme d'indépendance (plus faible) sera précisée explicitement.

Les tribus contenant parfois beaucoup d'éléments, il peut être délicat de vérifier leur indépendance. Dans le cas où elle sont engendrées par des algèbres, il suffit de vérifier l'indépendance des algèbres.

Proposition IV.1.5. *Si $\mathcal{C}_1$ et $\mathcal{C}_2$ sont deux algèbres indépendantes dans l'espace probabilisé $(\Omega, \mathcal{A}, P)$, alors les tribus $\sigma(\mathcal{C}_1)$ et $\sigma(\mathcal{C}_2)$ sont indépendantes.*

Démonstration. Soit $A_1 \in \mathcal{C}_1$. La classe monotone

$$\mathcal{M}_2 = \big\{\, A_2 \in \sigma(\mathcal{C}_2)\ :\ P(A_1 \cap A_2) = P(A_1)P(A_2)\,\big\}$$

des événements indépendants de A_1 contient $\mathcal{C}_2$. Elle contient donc la classe monotone engendrée par $\mathcal{C}_2$ qui est égale à $\sigma(\mathcal{C}_2)$ d'après le théorème I.3.3. Soit à présent un élément $A_2 \in \sigma(\mathcal{C}_2)$. La classe monotone

$$\mathcal{M}_1 = \{\, A_1 \in \sigma(\mathcal{C}_1) \,:\, P(A_1 \cap A_2) = P(A_1)P(A_2)\,\}$$

des événements indépendants de A_2 contient $\mathcal{C}_1$ d'après le point précédent, et donc $\sigma(\mathcal{C}_1)$. La conclusion s'ensuit. ☐

Il suffirait de considérer dans la proposition précédente des familles $\mathcal{C}_1$ et $\mathcal{C}_2$ stables par intersection finie.

La définition d'indépendance se formule de façon équivalente en terme de variables aléatoires.

Définition IV.1.6. Une famille quelconque de variables aléatoires X_i, $i \in I$, sur $(\Omega, \mathcal{A}, P)$ et à valeurs dans $(E, \mathcal{B})$ est (mutuellement) indépendante si la famille des tribus engendrées par les X_i est (mutuellement) indépendante, *i.e.* pour tout $J \subset I$ fini et tous les ensembles mesurables $B_j \in \mathcal{B}$, $j \in J$,

$$P\{\, X_j \in B_j \,:\, j \in J\,\} = P\Big(\bigcap_{j\in J}\{\, X_j \in B_j\,\}\Big) = \prod_{j\in J} P\{\, X_j \in B_j\,\}\,.$$

Exemples IV.1.7. (i) Reprenons l'exemple IV.1.2. Soient X_1 et X_2 les projections de Ω sur la première et seconde composante ($X_1(i,j) = i$ et $X_2(i,j) = j$). Ces projections sont des variables aléatoires qui modélisent le tirage de chacun des deux dés. Alors,

$$\begin{aligned} P(A) &= P\big(X_1^{-1}\{\,1,2,3,4\,\} \cap X_2^{-1}\{\,6\,\}\big) \\ &= P\big\{\,(1,6),(2,6),(3,6),(4,6)\,\big\} \\ &= 4/36 = 1/9 = P(A)P(B)\,. \end{aligned}$$

Donc A et B sont bien indépendants. Pour vérifier que X_1 et X_2 sont indépendantes, comme on l'attend intuitivement si notre modèle représente bien la réalité, observons que la tribu $\sigma(X_1)$, engendrée par X_1, est formée des ensembles $A_1 \times \{\,1,2,3,4,5,6\,\}$, $A_1 \subset \{\,1,2,3,4,5,6\,\}$. De façon symétrique,

$$\sigma(X_2) = \{\,\{\,1,2,3,4,5,6\,\} \times B_2 \,:\, B_2 \subset \{\,1,2,3,4,5,6\,\}\,\}\,.$$

Donc si $A \in \sigma(X_1)$ et $B \in \sigma(X_2)$ sont non vides, $A = A_1 \times \{1,2,3,4,5,6\}$, $B = \{1,2,3,4,5,6\} \times B_2$, et

$$\begin{aligned} P(A \cap B) = P(A_1 \times B_2) &= \sum_{(i,j)\in A_1\times B_2} \frac{1}{36} = \sum_{i\in A_1} \frac{1}{6} \sum_{j\in B_2} \frac{1}{6} \\ &= P(A_1)P(B_2) = P(A)P(B)\,. \end{aligned}$$

(ii) Poursuivons l'exemple IV.1.4.i. La famille de variables aléatoires

$$X_n = \mathbb{1}_{A_n} = \sum_{1\le k\le 2^{n-1}} \mathbb{1}_{\left]\frac{2(k-1)}{2^n},\frac{2k-1}{2^n}\right]}\,, \qquad n \ge 1\,,$$

de $[\,0,1\,]$ dans $\{\,0,1\,\}$ est indépendante. Il est aisé de vérifier que la loi de X_n est donnée par $P\{\,X_n = 0\,\} = P\{\,X_n = 1\,\} = 1/2$ (loi de Bernoulli de paramètre $1/2$).

(iii) Il existe une notion de variables aléatoires indépendantes deux à deux, plus faible que l'indépendance mutuelle. Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, chacune de loi de Bernoulli symétrique sur $\{\,-1,+1\,\}$ ($P\{\,X = -1\,\} = P\{\,X = +1\,\} = 1/2$). Soit $Z = XY$. La famille de variables aléatoires (X, Y, Z) est formée de variables indépendantes deux à deux, mais n'est pas mutuellement indépendante. Cet exemple est une variation en terme de variables aléatoires de l'exemple IV.1.4.ii.

Nous reformulons maintenant l'indépendance des variables aléatoires en terme de lois de ces variables. C'est cette formulation qui permet de travailler avec les variables indépendantes. Rappelons que si $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire sur $(\Omega, \mathcal{A}, P)$, la loi de X détermine la loi des marges, c'est-à-dire la loi de chacune des X_i, mais que la réciproque est fausse en général (exemple III.3.3.i). Néanmoins, si les coordonnées sont indépendantes, le résultat suivant montre que la loi du vecteur est déterminée par celles des marges.

On ne considère dans l'énoncé suivant que le cas de variables aléatoires réelles. Le cas de variables aléatoires à valeurs vectorielles est similaire.

Proposition IV.1.8. *Soit $(X_1, \ldots, X_d)$ une famille finie de variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}, P)$. La loi $P^{(X_1,\ldots,X_d)}$ du vecteur aléatoire sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est égale au produit des lois marginales $P^{X_1} \otimes \cdots \otimes P^{X_d}$. Réciproquement, si la loi du vecteur est égale au produit des marges, alors les variables sont indépendantes.*

Démonstration. Si $B = B_1 \times \cdots \times B_d$ est un pavé dans $\mathcal{B}(\mathbb{R}^d)$, par hypothèse d'indépendance,

$$\begin{aligned} P^{(X_1,\ldots,X_d)}(B) &= P\big((X_1,\ldots,X_d)^{-1}(B_1 \times \cdots \times B_d)\big) \\ &= P\big(X_1^{-1}(B_1) \cap \ldots \cap X_d^{-1}(B_d)\big) \\ &= P\big(X_1^{-1}(B_1)\big) \cdots P\big(X_d^{-1}(B_d)\big) \\ &= P^{X_1}(B_1) \cdots P^{X_d}(B_d)\,. \end{aligned}$$

L'identité s'étend à l'algèbre des réunions finies disjointes de pavés, laquelle engendre la tribu borélienne produit $\mathcal{B}(\mathbb{R}^d)$. La réciproque découle des identités précédentes et de la définition d'une loi puisque

$$\begin{aligned} P\big\{\,(X_1,\ldots,X_d) \in B_1 \times \cdots \times B_d\,\big\} &= P^{(X_1,\ldots,X_d)}(B) \\ &= P^{X_1}(B_1) \cdots P^{X_d}(B_d) \\ &= \prod_{1\le i\le d} P\{\,X_i \in B_i\,\}\,. \end{aligned}$$ □

Exemples IV.1.9. (i) Soit $(X,Y) \in \mathbb{R}^2$ un couple de variables aléatoires, de densité $f(x)g(y)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$. Alors X et Y sont indépendantes, et de densité respective f et g si $\int f(x)\,dx = 1$ (et donc $\int g(x)\,dx = 1$).

(ii) Si $X_1,\ldots,X_d$ sont des variables aléatoires réelles indépendantes, si $I_1,\ldots,I_k$ forment une partition de $\{\,1,\ldots,d\,\}$ avec $n_j = \mathrm{card}(I_j)$, et si de plus ϕ_j est mesurable sur $\mathbb{R}^{n_j}$ à valeurs dans $\mathbb{R}$, alors

$$\big\{\,Y_j = \phi_j(X_i \,:\, i \in I_j)\,\big\}_{1\le j\le k}$$

est une famille de k variables aléatoires indépendantes. En effet,

$$\sigma(Y_j) \subset \sigma(X_i \,:\, i \in I_j) = \bigotimes_{i\in I_j} \sigma(X_i)\,,$$

et, par le même argument que dans l'exemple IV.1.4.iii, les tribus $\bigotimes_{i\in I_j}\sigma(X_i)$, $1 \le j \le k$, sont indépendantes.

La proposition IV.1.5 et l'exemple IV.1.9.ii suggèrent la véracité de la proposition suivante.

Proposition IV.1.10. *Sur un espace probabilisé $(\Omega,\mathcal{A},P)$, soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus indépendantes de $\mathcal{A}$. Soit $(J_l)_{l\in L}$ une partition arbitraire de l'ensemble I. La famille de tribus $(\sigma(\mathcal{A}_i \,:\, i \in J_l))_{l\in L}$ est une famille indépendante.*

Démonstration. D'après la définition IV.1.3, il suffit de faire la démonstration lorsque L est fini, $L = \{1, \ldots, n\}$. Il suffit alors de montrer que $\sigma(\mathcal{A}_i : i \in J_1)$ est indépendante de $\sigma(\mathcal{A}_i : i \in J_2 \cup \ldots \cup J_n)$. Autrement dit, nous sommes ramenés au cas d'une partition de I en deux sous-ensembles, J_1 et J_2. Notons $\mathcal{T}_j = \sigma(\mathcal{A}_i : i \in J_j)$, $j = 1, 2$. Comme pour la démonstration de la proposition IV.1.5, nous utilisons un argument de classe monotone.

Pour $j = 1, 2$, soit $\mathcal{E}_j$ la famille des intersections finies d'éléments des tribus $\mathcal{A}_i$, $i \in J_j$. Par définition $\mathcal{E}_j$ est stable par intersection finie, et donc $\mathcal{M}(\mathcal{E}_j) = \mathcal{T}_j$ (*cf.* théorème I.3.3). Fixons $E \in \mathcal{E}_1$ et notons

$$\mathcal{M}(E) = \{ A \in \mathcal{A} : P(E \cap A) = P(E)P(A) \}$$

la classe des événements indépendants de E. Alors $\mathcal{M}(E)$ est une classe monotone qui contient $\mathcal{E}_2$. Donc $\mathcal{M}(E) \supset \mathcal{M}(\mathcal{E}_2) = \mathcal{T}_2$. Si à présent $F \in \mathcal{T}_2$, la classe $\mathcal{M}(F)$ est toujours une classe monotone et contient $\mathcal{E}_1$ par le même argument ; donc elle contient $\mathcal{M}(\mathcal{E}_1) = \mathcal{T}_1$, ce qui démontre le résultat. □

De la proposition IV.1.8 nous déduisons un autre critère d'indépendance.

Corollaire IV.1.11. *Une famille quelconque de variables aléatoires réelles X_i, $i \in I$, sur $(\Omega, \mathcal{A}, P)$ est indépendante si et seulement si pour toute famille finie $J \subset I$ et toute famille finie de fonctions boréliennes ϕ_i, $i \in J$, telles que $\phi_i(X_i)$, $i \in J$, soient intégrables,*

$$E\Big(\prod_{i\in J} \phi_i(X_i)\Big) = \prod_{i\in J} E\big(\phi_i(X_i)\big) .$$

Démonstration. Supposons la famille X_i, $i \in I$, indépendante. Soit J une partie finie de I, que nous pouvons représenter par $J = \{1, \ldots, n\}$. Utilisons alors le théorème de Fubini (II.5.1) pour obtenir

$$\begin{aligned} E\Big(\prod_{1\leq i\leq n} \phi_i(X_i)\Big) &= \int \prod_{1\leq i\leq n} \phi_i(x_i)\, \mathrm{d}P^{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) \\ &= \int \prod_{1\leq i\leq n} \phi_i(x_i)\, \mathrm{d}P^{X_1}(x_1) \otimes \cdots \otimes \mathrm{d}P^{X_n}(x_n) \\ &= \prod_{1\leq i\leq n} \int \phi_i(x_i)\, \mathrm{d}P^{X_i}(x_i) \\ &= \prod_{1\leq i\leq n} E\big(\phi_i(X_i)\big) . \end{aligned}$$

La réciproque s'obtient en considérant pour ϕ_i des indicatrices de boréliens. □

Nous déduisons aussi du corollaire précédent un critère d'indépendance utilisant les fonctions caractéristiques.

Corollaire IV.1.12. *La famille $(X_1, \ldots, X_n)$ de variables aléatoires réelles est indépendante si et seulement si pour tout $(t_1, \ldots, t_n) \in \mathbb{R}^n$,*

$$\varphi^{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = \varphi^{X_1}(t_1) \cdots \varphi^{X_n}(t_n)\,.$$

Démonstration. Le produit $\varphi^{X_1} \cdots \varphi^{X_n}$ est la fonction caractéristique de la loi produit $P^{X_1} \otimes \cdots \otimes P^{X_n}$. C'est le résultat puisque la fonction caractéristique détermine la loi (III.5.2). □

Comme cas particulier du corollaire IV.1.11, nous observons que si $X_1, \ldots, X_n$ sont des variables indépendantes et intégrables,

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)\,.$$

On prendra garde au fait que cette propriété ne caractérise pas en général l'indépendance. Elle décrit en fait une propriété plus faible de non corrélation.

Définition IV.1.13. Deux variables aléatoires réelles $X, Y \in \mathrm{L}^2(\Omega, \mathcal{A}, P)$ sont non corrélées si

$$E(XY) = E(X)E(Y)\,,$$

ou, de façon équivalente, si $E((X - EX)(Y - EY)) = 0$. On dit aussi que les variables centrées $X - E(X)$ et $Y - E(Y)$ sont orthogonales (pour le produit scalaire dans L^2 — *cf.* II.6.6).

Exemples IV.1.14. (i) D'après le corollaire IV.1.11, deux variables aléatoires indépendantes de carré intégrable sont non corrélées.

(ii) Si X est une variable aléatoire réelle de loi normale $\mathcal{N}(0,1)$, alors X et $Y = X^2$ sont non corrélées. En effet, X et Y sont de carré intégrable et

$$E(XY) = E(X^3) = 0 = E(X)E(Y)$$

par application, par exemple, de la proposition III.5.6 pour calculer les moments de la loi normale. Il est clair intuitivement que X et Y ne sont pas indépendantes, ce qui est confirmé par le fait que

$$P\{\, X \geq 1\,,\, Y \geq 1\,\} = P\{\, X \geq 1\,\} \neq P\{\, X \geq 1\,\}P\{\, Y \geq 1\,\}$$

puisque $P\{\, Y \geq 1\,\} < 1$.

Pour les variables non corrélées, on peut facilement évaluer la variance de leur somme.

Proposition IV.1.15. *Si $X_1,\ldots,X_n$ sont deux à deux non corrélées, elles vérifient l'identité de Bienaymé,*

$$\operatorname{Var}\Big(\sum_{1\le i\le n} X_i\Big) = \sum_{1\le i\le n} \operatorname{Var} X_i\,.$$

Nous en déduisons l'inégalité, dite de Bienaymé-Tchebitchev,

$$P\Big\{\,\Big|\sum_{1\le i\le n}(X_i - E(X_i))\Big| \ge t\Big\} \le \frac{1}{t^2}\sum_{1\le i\le n}\operatorname{Var}(X_i)\,,\qquad t>0.$$

Démonstration. Comme $X_i - E(X_i)$ et $X_j - E(X_j)$, $i\neq j$, sont orthogonales dans L^2,

$$\begin{aligned}
\operatorname{Var}\Big(\sum_{1\le i\le n} X_i\Big) &= E\Big(\Big(\sum_{1\le i\le n} X_i - E(X_i)\Big)^2\Big)\\
&= \sum_{1\le i,j\le n} E\Big(\big(X_i - E(X_i)\big)\big(X_j - E(X_j)\big)\Big)\\
&= \sum_{1\le i\le n} E\Big(\big(X_i - E(X_i)\big)^2\Big)\\
&= \sum_{1\le i\le n}\operatorname{Var}(X_i)\,.
\end{aligned}$$

L'inégalité de Bienaymé-Tchebitchev est une conséquence de l'inégalité de Tchebitchev (exemple III.4.10.ii). □

Exemples IV.1.16. (i) Donnons une application du calcul des probabilités à l'étude de la géométrie des espaces vectoriels.

Soient $u_1,\ldots,u_d\in\mathbb{R}^d$, muni de la structure euclidienne, des vecteurs de norme au plus 1. Soient $p_1,\ldots,p_d\in[\,0,1\,]$ et $w=\sum_{1\le i\le d} p_i u_i$. Montrons qu'il existe $\varepsilon_1,\ldots,\varepsilon_d\in\{\,0,1\,\}$ tels que

$$\Big\| w - \sum_{1\le i\le d}\varepsilon_i u_i\Big\| \le \sqrt{d}/2\,.$$

Autrement dit, toute combinaison linéaire des u_i à coefficients dans $[\,0,1\,]$ peut être approximée à $\sqrt{d}/2$ près par une combinaison linéaire à coefficients dans $\{\,0,1\,\}$.

Pour cela, prenons $(X_1,\ldots,X_d)$ un vecteur aléatoire de loi

$$P\{\,(X_1,\ldots,X_d)=(x_1,\ldots,x_d)\,\} = \prod_{1\le i\le d} p_i^{x_i}(1-p_i)^{1-x_i}\,,$$

où $x_i \in \{0,1\}$, et soit

$$X = \sum_{1 \le i \le d} X_i u_i .$$

Les X_i sont mutuellement indépendantes, de loi de Bernoulli de paramètre p_i. En particulier, elles sont non corrélées. Alors,

$$\begin{aligned}
E\big(\|X-w\|^2\big) &= E\Big(\sum_{1 \le i \le d} (X_i - p_i)^2 \|u_i\|^2\Big) \\
&\qquad + 2E\Big(\sum_{1 \le i,j \le d} (X_i - p_i)(X_j - p_j)\langle u_i, u_j\rangle\Big) \\
&= \sum_{1 \le i \le d} \|u_i\|^2 p_i(1-p_i) \\
&\le d/4 .
\end{aligned}$$

Observons alors que si $\|X-w\|^2 > d/4$ pour presque toute valeur de X, alors $E(\|X-w\|^2) > d/4$ (proposition II.2.6). Donc il existe une valeur de $(X_1, \ldots, X_d)$ pour laquelle

$$\|X-w\|^2 \le d/4 ,$$

ce qui est le résultat.

(ii) Voici enfin un exemple d'application en théorie des nombres.

Soit $\nu(n)$ le nombre de diviseurs premiers de n. Nous allons montrer, qu'en un certain sens, pour la plupart des entiers $i \le n$, $\nu(i)$ est de l'ordre de $\ln\ln i$.

Proposition. *Si $a(n)$ est une suite qui tend vers l'infini, alors*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{card}\big\{\, i \le n : |\nu(i) - \ln\ln n| > a(n)\sqrt{\ln\ln n} \,\big\} = 0 .$$

En particulier, pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{card}\bigg\{\, i \le n : \Big|\frac{\nu(i)}{\ln\ln i} - 1\Big| \ge \varepsilon \,\bigg\} = 0 .$$

Pour montrer cette proposition, considérons la suite d'espaces probabilisés $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*), P_n)$ où

$$P_n = \frac{1}{n} \sum_{1 \le i \le n} \delta_i$$

est la loi uniforme sur $\{1, 2, \ldots, n\}$. Pour p premier, soit

$$X_p(i) = \begin{cases} 1 & \text{si } p \text{ divise } i, \\ 0 & \text{sinon }, \qquad i \in \mathbb{N}^* . \end{cases}$$

Les X_p sont des variables aléatoires sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*), P_n)$. Soit $X = \sum_{p \text{ premier}} X_p$. Observons que $X(i) = \nu(i)$ pour tout $i \in \mathbb{N}^*$ et que

$$\begin{aligned} &\frac{1}{n}\mathrm{card}\{ i \leq n : |\nu(i) - \ln\ln n| \geq a(n)\sqrt{\ln\ln n} \} \\ &\qquad\qquad\qquad\qquad = P_n\{ |X - \ln\ln n| \geq a(n)\sqrt{\ln\ln n} \} . \end{aligned}$$

Sous la loi P_n, en notant $\lfloor \cdot \rfloor$ la partie entière, pour tout entier $k \geq 1$

$$E(X_p^k) = P_n\{ X_p = 1 \} = \frac{1}{n} \sum_{1 \leq i \leq n} \mathbb{1}_{\{p \text{ divise } i\}} = \lfloor n/p \rfloor / n = p^{-1} + O(n^{-1}) .$$

En particulier, sous P_n,

$$E(X_p) = \frac{1}{p} + O(n^{-1}) \qquad \text{et} \qquad \mathrm{Var}(X_p) = \frac{1}{p}\Big(1 - \frac{1}{p}\Big) + O(n^{-1}) .$$

Le théorème des nombres premiers indique que le nombre de nombres premiers ne dépassant pas n est $\pi(n) = (n/\ln n)(1 + o(1))$, ce qui permet de montrer que

$$\sum_{\substack{p \text{ premier} \\ p \leq n}} 1/p = \ln\ln n + o(1) , \quad n \to \infty .$$

Donc, sous P_n,

$$E(X) = \sum_{p \leq n} p^{-1} + O(n^{-1}) = \ln\ln n + o(1) .$$

La covariance de X_p et X_q sous la loi P_n est

$$\begin{aligned} \mathrm{Cov}(X_p, X_q) &= E(X_p X_q) - E(X_p)E(X_q) \\ &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \\ &\leq \frac{1}{pq} - \Big(p - \frac{1}{n}\Big)\Big(q - \frac{1}{n}\Big) \\ &\leq \frac{1}{n}\Big(\frac{1}{p} + \frac{1}{q}\Big) . \end{aligned}$$

Donc, sous P_n,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{\substack{p \text{ premier} \\ p\leq n}} \mathrm{Var}(X_p) + 2 \sum_{\substack{p,q \text{ premiers} \\ p<q\leq n}} \mathrm{Cov}(X_p, X_q) \\
&\leq \sum_{p\leq n} \Big(\frac{1}{p}\Big(1-\frac{1}{p}\Big) + O\Big(\frac{1}{n}\Big)\Big) + \sum_{p\neq q\leq n} \frac{1}{n}\Big(\frac{1}{p}+\frac{1}{q}\Big) \\
&\leq \ln\ln n + \frac{1}{n} \sum_{\substack{p,q \text{ premiers} \\ p,q\leq n}} \Big(\frac{1}{p}+\frac{1}{q}\Big) + O\Big(\frac{1}{n}\Big) \sum_{\substack{p \text{ premier} \\ p\leq n}} 1 \\
&= \ln\ln n + \frac{2}{n}\ln\ln n + \frac{\pi(n)}{n} O(1) + O(1) \\
&= \ln\ln n + O(1)\,.
\end{aligned}$$

Appliquons maintenant l'inégalité de Tchebitchev. Pour $t>0$,

$$P_n\big\{\, |X-E(X)| \geq t\sqrt{\mathrm{Var}(X)}\,\big\} \leq 1/t^2\,,$$

ce qui donne

$$\limsup_{n\to\infty} P_n\big\{\, |X-\ln\ln n| \geq t\sqrt{\ln\ln n}\,\big\} \leq 1/t^2$$

et termine la démonstration du premier point de la proposition. Pour obtenir le second, il suffit de remarquer que pour tout $\varepsilon>0$,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\mathrm{card}\big\{\, i\leq n \,:\, |\ln\ln i - \ln\ln n| \geq \varepsilon \,\big\} \\
&\qquad\qquad = \lim_{n\to\infty} \frac{1}{n}\mathrm{card}\big\{\, i\leq n \,:\, i\leq n^{e^{-\varepsilon}} \,\big\} \\
&\qquad\qquad = 0\,.
\end{aligned}$$

IV.2. Sommes de variables aléatoires indépendantes

Remarquons que pour des variables aléatoires indépendantes, de même loi et de carré intégrable, la proposition IV.1.15 montre que si $t>0$,

$$P\bigg\{\,\bigg|\sum_{1\leq i\leq n} \big(X_i - E(X_i)\big)\bigg| \geq t\sqrt{n}\,\bigg\} \leq \frac{\mathrm{Var}(X_1)}{t^2}\,.$$

Ainsi, l'ordre de grandeur de la somme $\sum_{1\leq i\leq n}(X_i - E(X_i))$ est au plus $\sqrt{n}$. Autrement dit, $\sum_{1\leq i\leq n} X_i$ ressemble à un terme déterministe, $\sum_{1\leq i\leq n} E(X_i) = nE(X_1)$ (de l'ordre de n si $E(X_1)\neq 0$), plus un terme aléatoire de l'ordre au plus $\sqrt{n}$. Les résultats de cette section nous servirons, d'une part pour évaluer la loi de $\sum_{1\leq i\leq n} X_i$, d'autre part pour préciser au chapitre V le comportement du terme aléatoire de l'ordre de $\sqrt{n}$ (théorème limite central, V.5.4).

Les sommes de variables aléatoires indépendantes et de même loi jouent un rôle essentiel dans le calcul des probabilités et en statistique. Historiquement, de nombreux travaux leur ont été consacrés. Elles interviennent également dans de nombreux problèmes pratiques. Nous en verrons quelques exemples dans cette partie où nous étudierons comment calculer la loi d'une somme de variables aléatoires indépendantes.

Proposition IV.2.1. *Soient X et Y deux variables aléatoires réelles, indépendantes, sur $(\Omega, \mathcal{A}, P)$. La loi de la somme $X+Y$ est donnée par le produit de convolution $P^X * P^Y$ des lois P^X et P^Y, défini, pour toute fonction borélienne bornée ϕ de $\mathbb{R}$ dans $\mathbb{R}$, par*

$$\begin{aligned}\int_{\mathbb{R}} \phi \,\mathrm{d}(P^X * P^Y) &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(x+y)\,\mathrm{d}P^Y(y)\Big)\,\mathrm{d}P^X(x) \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(x+y)\,\mathrm{d}P^X(x)\Big)\,\mathrm{d}P^Y(y)\,.\end{aligned}$$

Démonstration. On écrit un théorème de transport sur le diagramme suivant :

$$\begin{tikzcd}
(\Omega, \mathcal{A}, P) \arrow[rr, "X+Y"] \arrow[dr, "{(X,Y)}"'] & & (\mathbb{R}, \mathcal{B}(\mathbb{R}), P^{X+Y}) \\
& (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P^{(X,Y)} = P^X \otimes P^Y) \arrow[ur, "U"'] &
\end{tikzcd}$$

où U est la fonction $U(x,y) = x+y$. Ainsi,

$$\begin{aligned}\int_{\mathbb{R}} \phi \,\mathrm{d}P^{X+Y} &= E\big(\phi(X+Y)\big) \\ &= E\big(\phi(U(X,Y))\big) \\ &= \int_{\mathbb{R}^2} \phi\circ U\,\mathrm{d}P^{(X,Y)} \\ &= \int_{\mathbb{R}^2} \phi\circ U\,\mathrm{d}(P^X\otimes P^Y) \\ &= \int_{\mathbb{R}^2} \phi(x+y)\,\mathrm{d}P^X(x) dP^Y(y)\,. \qquad \square\end{aligned}$$

Remarque IV.2.2. Le produit de convolution vérifie un certain nombre de propriétés algébriques issues de la description en terme de variables aléatoires (mais qui ne suffisent cependant pas à le caractériser) :

(i) $P * \delta_0 = P$ (puisque $X + 0 = X$) ;

(ii) (commutativité) $P * Q = Q * P$ (puisque $X + Y = Y + X$) ;

(iii) (associativité) $(P*Q)*R = P*(Q*R)$ (puisque $(X+Y)+Z = X+(Y+Z)$) ;

(iv) (distributivité) $P * (\alpha Q + (1-\alpha)R) = \alpha(P * Q) + (1-\alpha)(P * R)$ pour tout $\alpha \in [\,0,1\,]$; en effet, si ϕ est borélienne et bornée,

$$\begin{aligned}
\int \phi \,\mathrm{d}\big(P * (\alpha Q + (1-\alpha)R)\big) &= \int \phi(x+y)\,\mathrm{d}P(x)\,\mathrm{d}\big(\alpha Q + (1-\alpha)R\big)(y) \\
&= \alpha \int \phi(x+y)\,\mathrm{d}P(x)\,\mathrm{d}Q(y) \\
&\qquad + (1-\alpha)\int \phi(x+y)\,\mathrm{d}P(x)\,\mathrm{d}R(y) \\
&= \alpha \int \phi\,\mathrm{d}(P*Q) + (1-\alpha)\int \phi\,\mathrm{d}(P*R) \\
&= \int \phi\,\mathrm{d}\big(\alpha P * Q + (1-\alpha)P * R\big)\,.
\end{aligned}$$

Les fonctions caractéristiques fournissent un autre moyen de déterminer la loi de la somme de deux variables aléatoires indépendantes.

Proposition IV.2.3. *Si X et Y sont deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}, P)$, la fonction caractéristique de leur somme est donnée par le produit des fonctions caractéristiques*

$$\varphi^{X+Y}(t) = \varphi^X(t)\varphi^Y(t)\,, \quad t \in \mathbb{R}.$$

Démonstration. C'est une conséquence du corollaire IV.1.11, puisque pour tout $t \in \mathbb{R}$,

$$\varphi^{X+Y}(t) = E\big(e^{it(X+Y)}\big) = E\big(e^{itX}e^{itY}\big) = E\big(e^{itX}\big)E\big(e^{itY}\big) = \varphi^X(t)\varphi^Y(t)\,. \quad \square$$

On prendra soin de ne pas confondre la fonction caractéristique d'un couple (X,Y) de variables indépendantes, donnée par $\varphi^{(X,Y)}(s,t) = \varphi^X(s)\varphi^Y(t)$, s, $t \in \mathbb{R}$, avec la fonction caractéristique de la somme $X+Y$ décrite ci-dessus.

Exemples IV.2.4. (i) Si $X = a$ p.s. et $Y = b$ p.s., alors X et Y sont indépendantes et $X + Y = a + b$ p.s. Autrement dit, $\delta_a * \delta_b = \delta_{a+b}$.

(ii) Soient X, Y indépendantes, où X suit la loi de Poisson de paramètre λ, notée $\mathcal{P}(\lambda)$, et Y suit la loi $\mathcal{P}(\mu)$. Alors $X+Y$ est de loi $\mathcal{P}(\lambda+\mu)$. En effet, $\mathcal{P}(\lambda) = \sum_{n\in\mathbb{N}} e^{-\lambda}\frac{\lambda^n}{n!}\delta_n$, et en appliquant IV.2.2.iv,

$$\begin{aligned}
\mathcal{P}(\lambda) * \mathcal{P}(\mu) &= \sum_{n\in\mathbb{N}}\sum_{m\in\mathbb{N}} e^{-\lambda}e^{-\mu}\frac{\lambda^n}{n!}\frac{\mu^m}{m!}\,\delta_n * \delta_m \\
&= e^{-(\lambda+\mu)}\sum_{n\in\mathbb{N}}\sum_{m\in\mathbb{N}} \frac{\lambda^n}{n!}\frac{\mu^m}{m!}\,\delta_{n+m} \\
&= e^{-(\lambda+\mu)}\sum_{k\in\mathbb{N}}\biggl(\sum_{m+n=k} \frac{\lambda^n}{n!}\frac{\mu^m}{m!}\,\delta_k\biggr) \\
&= e^{-(\lambda+\mu)}\sum_{k\in\mathbb{N}} \frac{(\lambda+\mu)^k}{k!}\,\delta_k = \mathcal{P}(\lambda+\mu)\,.
\end{aligned}$$

Une autre démonstration, plus probabiliste, consiste à écrire que pour tout $k\in\mathbb{N}$, $\{\,X+Y=k\,\}$ est la réunion disjointe des événements $\{\,X+n=k\,;\,Y=n\,\}$, $0\le n\le k$. Ainsi, par indépendance de X et Y,

$$\begin{aligned}
P\{\,X+Y=k\,\} &= \sum_{0\le n\le k} P\{\,X+n=k\ ;\ Y=n\,\} \\
&= \sum_{0\le n\le k} P\{\,X=k-n\,\}P\{\,Y=n\,\} \\
&= \sum_{0\le n\le k} e^{-\lambda}\frac{\lambda^{k-n}}{(k-n)!}\,e^{-\mu}\frac{\mu^n}{n!} \\
&= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}\,.
\end{aligned}$$

On peut utiliser enfin les fonctions caractéristiques. Si $t\in\mathbb{R}$,

$$\varphi^{X+Y}(t) = \varphi^X(t)\varphi^Y(t) = e^{\lambda(e^{it}-1)}e^{\mu(e^{it}-1)} = e^{(\lambda+\mu)(e^{it}-1)}$$

est la fonction caractéristique de $\mathcal{P}(\lambda+\mu)$ par le théorème III.5.2 et l'exemple III.5.3.v. Sous réserve qu'elles soient calculables, les transformées de Fourier sont donc un outil très efficace pour l'étude des sommes de variables aléatoires indépendantes.

(iii) Jetons une pièce n fois. Quelle est la loi du nombre de piles ? Modélisons n jets d'une pièce par n variables aléatoires indépendantes $X_1,\dots,X_n$ chacune de loi de Bernoulli $\mathcal{B}(1,p)$, *i.e.*

$$P\{\,X_i=1\,\} = 1 - P\{\,X_i=0\,\} = p$$

(avec $p = 1/2$ si la pièce n'est pas truquée), $X_i = 1$ représentant le tirage de pile au i-ème coup, et $X_i = 0$ le tirage de face. Le nombre de piles est donc $S_n = X_1 + \cdots + X_n$. Montrons que cette somme suit une loi binomiale $\mathcal{B}(n, p)$. À cet effet, observons d'abord que S_n prend les valeurs $0, 1, 2, \ldots, n$. Pour tout $0 \leq k \leq n$, l'événement $\{ S_n = k \}$ est la réunion disjointe des deux événements $\{ S_{n-1} = k\,; X_n = 0 \}$ et $\{ S_{n-1} = k - 1\,; X_n = 1 \}$. Ainsi, par indépendance de S_{n-1} et X_n (exemple IV.1.9.ii),

$$\begin{aligned}
P\{ S_n = k \} &= P\{ S_n = k \ ; \ X_n = 0 \} + P\{ S_n = k \ ; \ X_n = 1 \} \\
&= P\{ S_{n-1} = k \ ; \ X_n = 0 \} + P\{ S_{n-1} = k - 1 \ ; \ X_n = 1 \} \\
&= P\{ S_{n-1} = k \} P\{ X_n = 0 \} + P\{ S_{n-1} = k - 1 \} P\{ X_n = 1 \} \\
&= (1-p) P\{ S_{n-1} = k \} + p P\{ S_{n-1} = k - 1 \} .
\end{aligned}$$

On peut donc démontrer le résultat par récurrence. Rappelons la formule de Pascal, $\mathrm{C}_{n-1}^k + \mathrm{C}_{n-1}^{k-1} = \mathrm{C}_n^k$. Notons que $S_1 = X_1$ est de loi $\mathcal{B}(1, p)$. Si S_{n-1} est de loi $\mathcal{B}(n-1, p)$, alors

$$\begin{aligned}
P\{ S_n = k \} &= (1-p) \mathrm{C}_{n-1}^k \, p^k (1-p)^{n-1-k} + p \mathrm{C}_{n-1}^{k-1} \, p^{k-1} (1-p)^{n-1-(k-1)} \\
&= \mathrm{C}_{n-1}^k \, p^k (1-p)^{n-k} + \mathrm{C}_{n-1}^{k-1} \, p^k (1-p)^{n-k} \\
&= \mathrm{C}_n^k \, p^k (1-p)^{n-k} .
\end{aligned}$$

Une autre démonstration, dans l'esprit de la proposition IV.2.1, consiste à écrire, pour $1 \leq k \leq n$,

$$\begin{aligned}
P\{ S_n = k \} &= E\big(\mathbb{1}_{\{k\}}(S_n) \big) \\
&= \int \mathbb{1}_{\{k\}}(x_1 + \cdots + x_n) \, \mathrm{d}P^{X_1}(x_1) \ldots \, \mathrm{d}P^{X_n}(x_n) \\
&= \sum_{x_1, \ldots, x_n \in \{0,1\}} \mathbb{1}_{\{k\}}(x_1 + \cdots + x_n) \prod_{1 \leq i \leq n} p^{x_i} (1-p)^{1-x_i} \\
&= \sum_{x_1, \ldots, x_n \in \{0,1\}} \mathbb{1}_{\{k\}}(x_1 + \cdots + x_n) p^{\Sigma_{1 \leq i \leq n} x_i} (1-p)^{n - \Sigma_{1 \leq i \leq n} x_i} \\
&= \mathrm{card}\big\{ (x_1, \ldots, x_n) \in \{0,1\}^n \ : \ x_1 + \cdots + x_n = k \big\} p^k (1-p)^{n-k} \\
&= \mathrm{C}_n^k \, p^k (1-p)^{n-k} .
\end{aligned}$$

On peut aussi utiliser le produit de convolution, ou encore les fonctions caractéristiques (voir III.5.3.vi).

La description de la loi binomiale comme loi de la somme de variables de Bernoulli indépendantes permet un calcul rapide de sa moyenne et de sa variance, puisque par linéarité de l'intégrale et par l'identité de Bienaymé,

$$E(S_n) = nE(X_1) = np \quad \text{et} \quad \mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = np(1-p)\,.$$

En particulier, S_n/n est le nombre moyen de piles sur les n jets. Observons que l'inégalité de Tchebitchev dans la version III.4.10.ii montre que pour tout $\varepsilon > 0$,

$$P\Big\{\,\Big|\frac{S_n}{n} - p\Big| \geq \varepsilon\,\Big\} \leq \frac{1}{\varepsilon^2}\mathrm{Var}\Big(\frac{S_n}{n}\Big) = \frac{p(1-p)}{\varepsilon^2 n}\,.$$

Donc la probabilité que S_n/n s'écarte de sa moyenne tend vers 0 lorsque n tend vers l'infini. Ainsi si nous tirons un grand nombre de fois à pile ou face, la proportion de pile sera avec forte probabilité à peu près 1/2 (c'est intuitivement clair, et montre encore que l'axiomatique des probabilités que nous avons utilisée est raisonnable).

Une autre interprétation de la loi binomiale peut être donnée en terme d'urne. Supposons qu'une urne contienne une proportion p de boules lapis-lazuli et $1-p$ de boules vert fluorescent. On tire une boule, note sa couleur, puis la remet dans l'urne. On recommence ainsi n tirages avec remise. Le nombre N_n de boules lapis-lazuli ainsi tirées suit une loi binomiale $\mathcal{B}(n,p)$. En effet, au i-ème tirage, notons $X_i = 1$ si la boule est lapis-lazuli, et 0 sinon. Alors, $N_n = X_1 + \cdots + X_n$ est une somme de variables aléatoires de Bernoulli de paramètre de succès p.

(iv) Soient X, Y deux variables aléatoires réelles, indépendantes, dont les lois admettent des densités respectives f et g par rapport à la mesure de Lebesgue. Il est aisé de constater que la loi de la somme $X+Y$ a une densité h par rapport à la mesure de Lebesgue, donnée par le produit de convolution des fonctions f et g,

$$h(x) = f * g(x) = \int_{\mathbb{R}} f(x-y)g(y)\,\mathrm{d}y = \int_{\mathbb{R}} g(x-y)f(y)\,\mathrm{d}y\,, \quad x \in \mathbb{R}\,.$$

En effet, si ϕ est une fonction borélienne bornée,

$$\begin{aligned}\int \phi \ \mathrm{d}(P^X * P^Y) &= \int\!\!\int \phi(x+y)f(x)g(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int\!\!\int \phi(z)f(z-y)g(y)\,\mathrm{d}y\,\mathrm{d}z \\ &= \int \phi(z)h(z)\,\mathrm{d}z\,.\end{aligned}$$

En particulier, on vérifie ainsi que $\mathcal{N}(0,\sigma_1^2) * \mathcal{N}(0,\sigma_2^2) = \mathcal{N}(0,\sigma_1^2+\sigma_2^2)$. Plus généralement, $\mathcal{N}(m_1,\sigma_1^2) * \mathcal{N}(m_2,\sigma_2^2) = \mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$. Comme précédemment, ces relations peuvent se vérifier rapidement sur les fonctions caractéristiques. Si X_1 et X_2 sont indépendantes de lois respectives $\mathcal{N}(m_1,\sigma_1^2)$ et $\mathcal{N}(m_2,\sigma_2^2)$, pour tout $t \in \mathbb{R}$,

$$\varphi^{X_1+X_2}(t) = \varphi^{X_1}(t)\varphi^{X_2}(t) = e^{im_1t-\sigma_1^2t^2/2}e^{im_2t-\sigma_2^2t^2/2} = e^{i(m_1+m_2)t-(\sigma_1^2+\sigma_2^2)t^2/2}\,.$$

IV.3. Applications de l'indépendance

Nous étudions à présent quelques propriétés théoriques et pratiques de l'indépendance et leurs applications.

La question suivante est motivée par le désir de modéliser des suites infinies de variables aléatoires, des fonctions aléatoires, ou, de façon plus générale, des processus stochastiques : étant données des lois P_i sur $\mathbb{R}$ (par exemple), existe-t-il des variables aléatoires X_i sur un « certain » espace probabilisé $(\Omega, \mathcal{A}, P)$ qui sont indépendantes, et telles que $P^{X_i} = P_i$ pour tout i ? Lorsque l'on a qu'un nombre fini $P_1, \ldots, P_n$ de lois, cela ne pose pas de problème : on prend $\Omega = \mathbb{R}^n$ muni de sa tribu borélienne, $P = P_1 \otimes \cdots \otimes P_n$ et on considère les X_i comme les applications coordonnées, $X_i : \mathbb{R}^n \to \mathbb{R}$, qui à un point de $\mathbb{R}^n$ associe sa i-ème coordonnée.

Qu'en est-il cependant avec une infinité, par exemple dénombrable, de P_i ? Dans l'exemple IV.1.7.ii, nous avons construit une suite infinie de variables de Bernoulli $\mathcal{B}(1, 1/2)$ indépendantes, définies sur l'espace $([\,0,1\,], \mathcal{B}([\,0,1\,]), \lambda)$. Cette construction est spécifique à cet exemple. Soit en général une famille $(E_i, \mathcal{B}_i, P_i)_{i \geq 1}$ d'espaces probabilisés. On cherche à construire un espace probabilisé $(\Omega, \mathcal{A}, P)$ et une famille de variables aléatoires mutuellement indépendantes, X_i, $i \geq 1$, sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $(E_i, \mathcal{B}_i)$, telles que $P^{X_i} = P_i$ pour tout $i \geq 1$. Dans la pratique, tous les E_i sont égaux à $\mathbb{R}$ ou $\mathbb{R}^d$; mais cette notation nous permet de distinguer plus facilement chacun des facteurs de l'espace produit. Comme dans le cas fini, soit $\Omega = \prod_{i \geq 1} E_i$ et X_i la projection sur la i-ème coordonnée. Désignons par $\mathcal{A}$ la tribu produit des $\mathcal{B}_i$, $i \geq 1$, ou de façon équivalente engendrée par les X_i. La tribu $\mathcal{A}$ est engendrée par l'algèbre $\mathcal{C}$ dite des cylindres qui sont les ensembles A de la forme

$$A = C_n \times E_{n+1} \times E_{n+2} \times \cdots$$

où $C_n \in \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n$ est appelé la base du cylindre.

On peut définir sur $\mathcal{C}$ une fonction d'ensemble Q (à valeurs dans $[\,0,1\,]$) par

$$Q(A) = P_1 \otimes \cdots \otimes P_n(C_n)\,,$$

pour tout $A \in \mathcal{C}$ (de base C_n).

***Théorème IV.3.1* (de Kolmogorov).** *La fonction d'ensemble Q se prolonge en une unique probabilité P sur $(\Omega, \mathcal{A})$. Sous P, les X_i sont indépendantes et de loi P_i.*

Démonstration. La deuxième partie de l'énoncé est évidente par construction. Afin d'établir la première partie, on utilise le théorème de prolongement I.4.9. Observons que Q est une fonction d'ensembles additive sur $\mathcal{A}$. Puisque $Q(\Omega)$ est fini, pour montrer la σ-additivité de Q sur $\mathcal{A}$, il suffit de montrer que si $(A_n)_{n\in\mathbb{N}}$ est une suite décroissante dans $\mathcal{C}$ d'intersection vide, alors

$$\lim_{n\to\infty} Q(A_n) = \inf_{n\in\mathbb{N}} Q(A_n) = 0\,.$$

Supposons le contraire, c'est-à-dire que pour une certaine suite $(A_n)_{n\in\mathbb{N}}$ décroissante dans $\mathcal{C}$ et un certain $\varepsilon > 0$, $\inf_{n\in\mathbb{N}} Q(A_n) \geq \varepsilon$ et montrons alors que $\bigcap_{n\in\mathbb{N}} A_n \neq \emptyset$.

Pour tout $k \geq 1$, on peut construire une fonction d'ensembles additive Q^k sur $E^k = \prod_{i\geq k} E_i$ à l'image de la définition de $Q = Q^1$ sur $\Omega = E^1$. Par le théorème de Fubini (II.5.1), pour tout n

$$Q(A_n) = \int Q^2(A_n^{\omega_1})\,\mathrm{d}P_1(\omega_1)$$

où $A_n^{\omega_1} = \{\,\omega \in E^2 \,:\, (\omega_1, \omega) \in A_n\,\}$ est la section de A_n suivant $\omega_1 \in E_1$. Soit

$$B_n = \bigl\{\,\omega_1 \in E_1 \,:\, Q^2(A_n^{\omega_1}) \geq \varepsilon/2\,\bigr\},$$

($A_n^{\omega_1}$ appartient à la tribu cylindrique sur E^2 par l'exercice I.3). Par définition de B_n, puisque $Q^2(A_n^{\omega_1}) < \varepsilon/2$ sur le complémentaire de B_n et $Q^2(A_n^{\omega_1}) \leq 1$ partout,

$$\varepsilon \leq Q(A_n) \leq \tfrac{\varepsilon}{2}\bigl(1 - P_1(B_n)\bigr) + P_1(B_n) \leq \tfrac{\varepsilon}{2} + P_1(B_n)\,.$$

Ainsi, $P_1(B_n) \geq \varepsilon/2$. La suite $(B_n)_{n\in\mathbb{N}}$ décroît dans E_1, et donc, nécessairement, puisque P_1 est une probabilité, $\bigcap_n B_n$ n'est pas vide. Soit ζ_1 un élément de $\bigcap_n B_n$. Ce que l'on a fait pour $Q = Q^1$ et la suite $(A_n)_{n\in\mathbb{N}}$, nous le recommençons pour la fonction d'ensembles Q^2 et la suite $(A_n^{\zeta_1})_{n\in\mathbb{N}}$, car $\inf_n Q^2(A_n^{\zeta_1}) \geq \varepsilon/2$. On construit de cette façon un point $\zeta = (\zeta_i)_{i\geq 1}$ dans Ω tel que pour tout $k \geq 1$, $\inf_n Q^{k+1}(A_n^{\zeta_1,\ldots,\zeta_k}) > 0$ où

$$A_n^{\zeta_1,\ldots,\zeta_k} = \bigl\{\,\omega \in E^{k+1} \,:\, (\zeta_1, \ldots, \zeta_k, \omega) \in A_n\,\bigr\}$$

(qui est un cylindre de E^{k+1}). En particulier, à n fixé, pour tout $k \geq 1$, $A_n^{\zeta_1,\ldots,\zeta_k} \neq \emptyset$. Mais A_n est un cylindre de la forme $A_n = C_N \times E^{N+1}$. Or on peut vérifier, en s'appuyant sur un dessin par exemple, que $A_n^{\zeta_1,\ldots,\zeta_N}$ est non vide si et seulement si $(\zeta_1, \ldots, \zeta_N) \in C_N$. Ainsi $\zeta \in A_n$. Comme ceci à lieu pour n'importe quel n, l'intersection $\bigcap_{n\in\mathbb{N}} A_n$ n'est pas vide puisqu'elle contient ζ. Le théorème est établi. $\square$

En conséquence de ce théorème, nous pouvons parler plus librement d'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires réelles indépendantes sur un espace probabilisé

$(\Omega, \mathcal{A}, P)$. Un raisonnement similaire permet de construire des suites de vecteurs aléatoires indépendants.

Dans de nombreux problèmes de probabilité, on est intéressé par le comportement limite d'une suite de variables aléatoires. Un exemple élémentaire est la suite des proportions de piles dans un tirage successif à pile ou face. Dans de telles situations, les événements dans une tribu engendrée par un nombre fini de variables ont peu d'intérêt, et on ne s'interesse en fait qu'aux événements définis ultimement. Dans les bons cas, ceux-ci appartiennent à une tribu appelée tribu terminale que nous introduisons maintenant.

Définition IV.3.2. Soit $(\mathcal{T}_n)_{n\in\mathbb{N}}$ une famille indépendante de tribus sur $(\Omega, \mathcal{A}, P)$ (par exemple $\mathcal{T}_n = \sigma(X_n)$ où les X_n sont indépendantes). On désigne par $\mathcal{A}_n$ la tribu engendrée par $\mathcal{T}_n, \mathcal{T}_{n+1}, \ldots$ et pose $\mathcal{A}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{A}_n$. La tribu $\mathcal{A}_\infty$ est appelée tribu des événements terminaux ou tribu terminale (de la suite $(\mathcal{T}_n)_{n\in\mathbb{N}}$).

La tribu terminale vérifie la loi du tout ou rien suivante, aussi appelée loi du 0–1.

***Théorème IV.3.3* (loi du 0–1).** *Si $\mathcal{A}_\infty$ est une tribu terminale, alors tout $A \in \mathcal{A}_\infty$ vérifie $P(A) = 0$ ou 1.*

Démonstration. Soit $A \in \mathcal{A}_\infty$ fixé. On considère la classe monotone des événements indépendants de A,

$$\mathcal{M} = \{\, B \in \mathcal{A} : P(A \cap B) = P(A)P(B) \,\}.$$

On se propose de montrer que $\mathcal{M} \supset \mathcal{A}_\infty$. Si tel est le cas, $A \in \mathcal{M}$ et $P(A) = P(A)^2$, et donc $P(A) = 0$ ou 1. Considérons les tribus $\mathcal{B}_n = \sigma(\mathcal{T}_0, \ldots, \mathcal{T}_n)$, $n \in \mathbb{N}$, et posons $\mathcal{B}_\infty = \bigcup_n \mathcal{B}_n$. En tant que réunion croissante, $\mathcal{B}_\infty$ est une algèbre. Nous savons, d'après la proposition IV.1.10, que les tribus $\mathcal{B}_n$ et $\mathcal{A}_{n+1}$ sont indépendantes. Il s'ensuit que tout élément de $\mathcal{B}_n$ est indépendant de A. Ainsi, puisque $\mathcal{A}_\infty \subset \mathcal{A}_{n+1}$, il s'ensuit $\mathcal{B}_\infty \subset \mathcal{M}$. Donc, en utilisant le théorème des classes monotones I.3.3, $\sigma(\mathcal{B}_\infty) = \mathcal{M}(\mathcal{B}_\infty) \subset \mathcal{M}$. Il reste à voir que $\sigma(\mathcal{B}_\infty) \supset \mathcal{A}_\infty$, ce qui est intuitivement clair. En effet, pour tout k,

$$\mathcal{T}_k \subset \mathcal{B}_k \subset \mathcal{B}_\infty \subset \sigma(\mathcal{B}_\infty).$$

Donc pour tout n, $\mathcal{A}_n = \sigma(\mathcal{T}_k : k \geq n) \subset \sigma(\mathcal{B}_\infty)$, d'où le résultat. □

Exemples IV.3.4. (i) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants de $(\Omega, \mathcal{A}, P)$; alors

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m = \{\, A_n \text{ a lieu une infinité de fois } \}$$

est un événement terminal pour la suite de tribus $\mathcal{T}_n = \sigma(A_n) = \{\, \emptyset, \Omega, A_n, A_n^{\mathrm{c}} \,\}$; donc $P(A) = 0$ ou 1.

On abrège souvent l'expression « A_n a lieu une infinité de fois » par « A_n infiniment souvent » ou « A_n i.s. ». Remarquer que $P(A_n \text{ i.s.}) = 0$ signifie que presque sûrement seulement un nombre fini d'événements A_n surviennent. C'est-à-dire que pour presque tout $\omega \in \Omega$, il existe un $n(\omega)$ fini tel que si $n \geq n(\omega)$ alors $\omega \notin A_n$, *i.e.* A_n n'a pas lieu. On fera très attention au fait que l'entier $n(\omega)$ dépend de ω.

(ii) Soient X_n, $n \in \mathbb{N}$, des variables aléatoires réelles indépendantes, $\mathcal{T}_n = \sigma(X_n)$, et soit a_n des réels positifs, tels que $\lim_{n\to\infty} a_n = \infty$. Considérons l'événement

$$A = \Big\{\, \omega \in \Omega \,:\, \frac{1}{a_n}\big(X_1(\omega) + \cdots + X_n(\omega)\big) \text{ converge lorsque } n \to \infty \,\Big\}.$$

Alors $A \in \mathcal{A}_\infty$ car, pour tout k,

$$A = \Big\{\, \omega \in \Omega \,:\, \frac{1}{a_n}\big(X_k(\omega) + \cdots + X_n(\omega)\big) \text{ converge lorsque } n \to \infty \,\Big\}.$$

Donc A est soit de probabilité pleine, soit de probabilité nulle.

Le résultat suivant est connu sous le nom de lemme de Borel-Cantelli, mais le statut de théorème est justifié par son importance pratique.

Théorème IV.3.5 **(Lemme de Borel-Cantelli).** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.*
(i) Si $\sum_{n\in\mathbb{N}} P(A_n) < \infty$ alors $P(A_n \text{ i.s.}) = 0$.
(ii) Si la suite $(A_n)_{n\in\mathbb{N}}$ est indépendante alors $\sum_{n\in\mathbb{N}} P(A_n) = \infty$ implique $P(A_n \text{ i.s.}) = 1$.

Démonstration. La partie (i) est évidente : pour tout n,

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m \subset \bigcup_{m\geq n} A_m\,,$$

et donc $P(A_n \text{ i.s.}) = P(A) \leq \sum_{m\geq n} P(A_m)$ qui tend vers 0 avec n si la série converge.

La partie (ii) s'obtient en remarquant d'abord que pour tout n et tout $N \geq n$,

$$
\begin{aligned}
P\Big(\bigcup_{n\leq m\leq N} A_m\Big) &= 1 - P\Big(\bigcap_{n\leq m\leq N} A_m^{\mathrm{c}}\Big) \\
&= 1 - \prod_{n\leq m\leq N} \big(1 - P(A_m)\big) \,.
\end{aligned}
$$

Comme $1 - x \leq e^{-x}$ pour tout $x \geq 0$,

$$
P\Big(\bigcup_{n\leq m\leq N} A_m\Big) \geq 1 - \exp\Big(-\sum_{n\leq m\leq N} P(A_m)\Big) \,.
$$

Lorsque N tend vers l'infini, $\sum_{n\leq m\leq N} P(A_m)$ tend, pour tout n, vers l'infini par hypothèse, et donc

$$
P\Big(\bigcup_{m\geq n} A_m\Big) = 1 \,.
$$

Il ne reste plus qu'à remarquer que

$$
P(A) = \lim_{n\to\infty} P\Big(\bigcup_{m\geq n} A_m\Big) \,. \qquad \square
$$

Il existe de nombreuses versions du lemme de Borel-Cantelli. Il suffit par exemple de supposer que les A_n sont indépendants deux à deux (voir aussi exercice IV.15).

***Exemples* IV.3.6.** (i) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles telle que, pour un $M \in \mathbb{R}$, $\sum_{n\in\mathbb{N}} P\{\, X_n \geq M \,\} < \infty$. Alors, d'après la partie (i) du lemme de Borel-Cantelli, $P\{\, X_n \geq M$ i.s.$\,\} = 0$. Autrement dit,

$$
P\Big\{\bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \{\, X_m < M \,\}\Big\} = 1 \,.
$$

Donc $\limsup_{n\to\infty} X_n \leq M$ p.s. De la même façon, si $\sum_n P\{\, X_n \leq M \,\} < \infty$, alors $\liminf_{n\to\infty} X_n \geq M$ p.s.

(ii) On jette une infinité de fois une pièce équilibrée. Quelle est la probabilité d'obtenir une infinité de fois deux piles consécutifs ? On représente le jeu par une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires sur $(\Omega, \mathcal{A}, P)$, indépendantes, avec $P\{\, X_n = 1 \,\} = P\{\, X_n = 0 \,\} = 1/2$. Posons $A_n = \{\, X_n = X_{n+1} = 1 \,\}$. On s'intéresse à $P(A_n$ i.s.$)$. Il est clair que les A_n ne forment pas une suite indépendante, puisque par exemple la $(n+1)$-ème variable détermine à la fois A_n et A_{n+1}. En revanche, la sous-suite $(A_{2n})_{n\in\mathbb{N}}$ forme une suite indépendante. En outre, $P(A_{2n}) = 1/4$ pour tout n, et donc $\sum_n P(A_{2n}) = \infty$. Ainsi par le lemme de Borel-Cantelli, $P(A_{2n}$ i.s.$) = 1$. Comme $\{\, A_{2n}$ i.s.$\,\} \subset \{\, A_n$ i.s.$\,\}$, on conclut que $P(A_n$ i.s.$\,) = 1$.

(iii) Donnons un nouvel exemple d'application du calcul des probabilités à l'étude des nombres. Considérons l'espace probabilisé $([\,0,1\,], \mathcal{B}([\,0,1\,]),\ P = \lambda)$, et soit U la fonction identité de $[\,0,1\,]$ dans lui-même. C'est une variable aléatoire de loi uniforme sur $[\,0,1\,]$. Pour tout nombre réel x, notons $\lfloor x \rfloor$ sa partie entière. Pour $\omega \in [\,0,1\,]$, écrivons le développement dyadique du nombre réel $U(\omega)$,

$$U(\omega) = \sum_{i\geq 1} 2^{-i} U_i(\omega)$$

avec $U_i = \lfloor 2^{i+1}U \rfloor - \lfloor 2^i U \rfloor \in \{\, 0,1\,\}$. Les U_i sont mesurables, puisque images de U par des applications mesurables. Ce sont donc des variables aléatoires. Si $(u_1,\ldots,u_n) \in \{\,0,1\,\}^n$,

$$\begin{aligned} P\{\, (U_1,\ldots,U_n) &= (u_1,\ldots,u_n)\,\} \\ &= \lambda\Big(\{\, x \in [\,0,1\,] : (x_1,\ldots,x_n) = (u_1,\ldots,u_n)\,\}\Big) \\ &= 2^{-n}\,. \end{aligned}$$

En particulier,

$$P\{\, U_i = 0\,\} = P\{\, U_i = 1\,\} = 1/2$$

et les U_i, $i \geq 1$, forment une suite de variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}(1, 1/2)$ (ici nous utilisons le théorème de prolongement IV.3.1 puisque nous n'avons en fait montré l'indépendance des U_i que pour $1 \leq i \leq n$ et tout $n \geq 1$).

Le développement dyadique d'un nombre dans $[\,0,1\,]$ est constitué de blocs de 0 et de 1. Par exemple le développement

$$0\,,\,0\,,\,1\,,\,1\,,\,1\,,\,1\,,\,0\,,\,0\,,\,1\,,\,1\,,\,1\,,\,0\,,\,1\,,\,0\,,\,0\,,\,0\,,\,1\,,\,\ldots$$

est formé de blocs de longueur $2, 4, 2, 3, 1, 1, 3, \ldots$

Soit N_n le nombre de blocs dans les n premiers chiffres. C'est une variable aléatoire puisque pour tout entier k positif, l'événement $\{ N = k \}$ est égal à

$$\begin{aligned}\bigcup_{1\leq n_1,\ldots,n_k\leq n} \Big(&\{ U_1 = \cdots = U_{n_1} \} \cap \{ U_{n_1} \neq U_{n_1+1} \} \\ &\cap \{ U_{n_1+1} = \cdots = U_{n_1+n_2} \} \cap \{ U_{n_1+n_2} \neq U_{n_1+n_2+1} \} \\ &\ldots \\ &\cap \{ U_{n_1+\cdots+n_{n_{k-1}}+1} = \cdots = U_{n_1+\ldots+n_k} \} \cap \{ U_{n_1+\cdots+n_k} \neq U_{n_1+\cdots+n_k+1} \} \\ &\cap \{ U_{n_1+\cdots+n_k} = \cdots = U_n \} \Big) ,\end{aligned}$$

et chaque ensemble

$$\begin{aligned}\{ U_i = \cdots = U_{i+j} \} = \big(\{ 0 = U_i \} \cap \ldots \cap \{ 0 = U_{i+j} \} \big) \\ \cup \big(\{ 1 = U_i \} \cap \ldots \cap \{ 1 = U_{i+j} \} \big)\end{aligned}$$

est mesurable. On pourrait ainsi calculer la loi de N_n, mais c'est un peu lourd. Pour obtenir des informations sur N_n, définissons la fonction génératrice

$$G_n(s) = \sum_{k\geq 1} s^k P\{ N_n = k \} .$$

Cette fonction est définie au moins sur l'intervalle $]-1,1[$ puisque $P\{ N_n = k \}$ est dans $[0,1]$. De plus, pour tout $k \geq 1$,

$$P\{ N_n = k \} = \frac{1}{k!} \frac{d^k}{ds^k} \big(G_n(s) \big) \Big|_{s=0} .$$

Cette fonction peut être calculée comme suit. Observons que pour $u \in \{ 0,1 \}$,

$$\begin{aligned}&P\{ N_n = k \,;\, U_n = u \} \\ &\qquad = P\{ N_{n-1} = k \,;\, U_{n-1} = u \,;\, U_n = u \} \\ &\qquad\qquad + P\{ N_{n-1} = k-1 \,;\, U_{n-1} = 1-u \,;\, U_n = u \} \\ &\qquad = \frac{1}{2} P\{ N_{n-1} = k \,;\, U_{n-1} = u \} + \frac{1}{2} P\{ N_{n-1} = k-1 \,;\, U_{n-1} = 1-u \}\end{aligned}$$

puisque N_{n-1} est $\sigma(U_1,\ldots,U_{n-1})$-mesurable et que U_n est indépendante de $U_1,\ldots,U_{n-1}$. On a donc

$$\begin{aligned}
G_n(s) &= \sum_{k\geq 1} s^k\Big(P\{\,N_n=k\,;\,U_n=0\,\}+P\{\,N_n=k\,;\,U_n=1\,\}\Big)\\
&= \frac12\sum_{k\geq 1} s^k\Big(P\{\,N_{n-1}=k\,;\,U_{n-1}=1\,\}+P\{\,N_{n-1}=k-1\,;\,U_{n-1}=0\,\}\\
&\qquad\qquad +P\{\,N_{n-1}=k\,;\,U_{n-1}=0\,\}+P\{\,N_{n-1}=k-1\,;\,U_{n-1}=1\,\}\Big)\\
&= \frac12\sum_{k\geq 1} s^k\Big(P\{\,N_{n-1}=k\,\}+P\{\,N_{n-1}=k-1\,\}\Big)\\
&= \frac{1+s}{2}\,G_{n-1}(s)\,.
\end{aligned}$$

Ainsi, par récurrence, et puisque $G_1(s)=s$,

$$G_n(s)=\Big(\frac{1+s}{2}\Big)^{n-1}s\,.$$

Nous pouvons maintenant encadrer N_n de la façon suivante, similaire à l'exemple III.4.10.iii. Observons que pour tout $s\leq 1$ et tout $c_n\geq 0$,

$$P\Big\{\,N_n\leq \frac n2 - c_n\,\Big\}\leq P\{\,s^{N_n}\geq s^{\frac n2 - c_n}\,\}\leq s^{c_n-n/2}G_n(s)\,.$$

Donc, en optimisant en s, et pour n assez grand,

$$P\Big\{\,N_n\leq \frac n2 - c_n\,\Big\}\leq \inf_{0\leq s\leq 1} s^{c_n-n/2}G_n(s)=e^{-2c_n^2/n(1+o(1))}$$

pourvu que $\lim_{n\to\infty} c_n^2/n=\infty$. En prenant $c_n=\sqrt{n\ln n}$, on voit que

$$\sum_{n\geq 1}P\Big\{\,N_n\leq \frac n2 - c_n\,\Big\}<\infty\,.$$

Donc, par le lemme de Borel-Cantelli, pour presque tout ω,

$$N_n(\omega)\geq \frac n2 - \sqrt{n\ln n}$$

pour n assez grand.

De même, pour c_n tel que $\lim_{n\to\infty} c_n^2/n=\infty$,

$$P\Big\{\,N_n\geq \frac n2 + c_n\,\Big\}\leq \inf_{s\geq 1} s^{-c_n-n/2}G_n(s)=e^{-2c_n^2/n(1+o(1))}$$

et presque sûrement pour n assez grand,

$$N_n(\omega) \leq \frac{n}{2} + \sqrt{n \ln n}\,.$$

Ainsi, nous avons montré que pour presque tout nombre de $[\,0,1\,]$, le nombre de blocs dans les n premiers chiffres du développement dyadique est compris entre $\frac{n}{2} - \sqrt{n \ln n}$ et $\frac{n}{2} + \sqrt{n \ln n}$ pour n assez grand. On voit de plus que

$$G_n'(1) = \sum_{k \geq 1} kP\{\, N_n = 1\,\} = E(N_n) = \frac{n+1}{2} \sim \frac{n}{2}\,, \qquad n \to \infty\,.$$

Donc, « en moyenne », un nombre a de l'ordre de $n/2$ blocs dans les n premiers chiffres.

IV.4. Vecteurs aléatoires gaussiens et lois gaussiennes

Rappelons qu'une variable aléatoire réelle $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ suit une loi normale ou gaussienne $\mathcal{N}(m, \sigma^2)$ de moyenne m et de variance σ^2 si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right), \qquad x \in \mathbb{R}\,.$$

De façon équivalente, sa transformée de Fourier est

$$\varphi^X(t) = e^{imt - \sigma^2 t^2/2}\,, \qquad t \in \mathbb{R}\,.$$

Rappelons aussi que si Y suit une loi $\mathcal{N}(0,1)$, alors $X = m + \sigma Y$ suit une loi $\mathcal{N}(m, \sigma^2)$. On notera pour toutes ces définitions que, dans la famille des lois gaussiennes, les paramètres de moyenne m et de variance σ^2 caractérisent une loi donnée. Une variable gaussienne a des moments de tous ordres (III.5.6).

Nous allons nous intéresser à présent à des variables aléatoires gaussiennes à valeurs dans $\mathbb{R}^d$, ou vecteurs aléatoires gaussiens.

Définition IV.4.1. Une variable aléatoire

$$X = (X_1, \ldots, X_d) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\,,$$

à valeurs dans $\mathbb{R}^d$, est dite gaussienne si pour tout $\alpha = (\alpha_1, \ldots, \alpha_d)$ de $\mathbb{R}^d$,

$$\langle \alpha, X \rangle = \sum_{1 \leq i \leq d} \alpha_i X_i$$

est une variable aléatoire gaussienne.

Dans la définition IV.4.1, la variable aléatoire $\langle \alpha, X \rangle$ réelle gaussienne est caractérisée par sa moyenne

$$E\biggl(\sum_{1 \le i \le d} \alpha_i X_i\biggr) = \sum_{1 \le i \le d} \alpha_i E(X_i)\,,$$

et sa variance

$$\mathrm{Var}\biggl(\sum_{1 \le i \le d} \alpha_i X_i\biggr) = \sum_{1 \le i,j \le d} \alpha_i \alpha_j E\bigl((X_i - E(X_i))(X_j - E(X_j))\bigr)\,.$$

Ainsi, le vecteur gaussien $X = (X_1, \ldots, X_d)$ est entièrement caractérisé par son vecteur moyen

$$m = E(X) = \bigl(E(X_1), \ldots, E(X_d)\bigr)$$

et sa matrice de covariance

$$\Gamma = \Bigl(E\bigl((X_i - E(X_i))(X_j - E(X_j))\bigr)\Bigr)_{1 \le i,j \le d}\,.$$

On voit très simplement sur ces formules comment se ramener au cas d'un vecteur gaussien centré ($E(X) = (0, \ldots, 0)$), en soustrayant simplement la moyenne ; dans ce cas, la matrice de covariance s'écrit $(E(X_i X_j))_{1 \le i,j \le d}$; nous le supposerons souvent par la suite.

En terme de transformée de Fourier, si $u = (u_1, \ldots, u_d) \in \mathbb{R}^d$, et si le vecteur aléatoire $X = (X_1, \ldots, X_d)$ est gaussien,

$$\varphi^X(u) = E\bigl(\exp(i\langle u, X \rangle)\bigr) = \exp\left(i\langle u, m \rangle - \frac{1}{2}\,{}^{\mathrm{t}}u\,\Gamma\,u\right)\,.$$

Notons que si $X = (X_1, \ldots, X_d)$ est un vecteur gaussien, ses marges $X_1, \ldots, X_d$ sont gaussiennes (considérer pour α le i-ème vecteur de base dans la définition IV.4.1), mais, sauf si elles sont indépendantes, la réciproque est fausse (par exemple si Z est gaussienne réelle de loi $\mathcal{N}(0,1)$ et ε est de loi de Bernoulli symétrique $P\{\,\varepsilon = +1\,\} = P\{\,\varepsilon = -1\,\} = 1/2$, indépendante de Z, alors $(Z, \varepsilon Z)$ n'est pas gaussien, mais de marges gaussiennes).

Un exemple élémentaire, mais fondamental, est constitué par un vecteur $G = (G_1, \ldots, G_d)$ dont les composantes sont indépendantes, de loi $\mathcal{N}(0,1)$. Le vecteur G est centré et sa matrice de covariance est la matrice identité. La loi de G a pour densité $(2\pi)^{-d/2} \exp(-\|x\|^2/2)$ par rapport à la mesure de Lebesgue $\mathrm{d}x$ sur $\mathbb{R}^d$ (où l'on rappelle que $\|x\|^2 = x_1^2 + \cdots + x_d^2$ pour tout $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$). On note $\mathcal{N}(0, \mathrm{Id})$ la loi de G. Montrons alors l'existence de vecteurs gaussiens de matrice de covariance Γ.

Toute matrice de covariance Γ étant symétrique et semi-définie positive peut être écrite $\Gamma = A\,{}^{\mathrm{t}}A$, où A est une matrice carrée.

Proposition IV.4.2. *Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien centré de matrice de covariance $\Gamma = A\,{}^{\mathrm{t}}A$. Alors X a même loi que AG où G est de loi $\mathcal{N}(0, \mathrm{Id})$. On note $\mathcal{N}(0, \Gamma)$ la loi de X.*

Démonstration. Le vecteur aléatoire AG est bien gaussien, puisque toute combinaison linéaire de ses coordonnées est une combinaison linéaire des coordonnées de G. Notons $A = (a_{i,j})_{1 \leq i,j \leq d}$. Pour tous $1 \leq i, j \leq d$,

$$\begin{aligned} E\big((AG)_i (AG)_j\big) &= E\bigg(\Big(\sum_{1 \leq k \leq d} a_{i,k} G_k\Big)\Big(\sum_{1 \leq l \leq d} a_{j,l} G_l\Big)\bigg) \\ &= \sum_{1 \leq k \leq d} a_{i,k} a_{j,k} = \Gamma_{i,j} = E(X_i X_j)\,. \end{aligned}$$ □

En vertu de cette proposition, on voit immédiatement par changement de variable que, si A est inversible, pour tout borélien B de $\mathbb{R}^d$,

$$P\{\, X \in B \,\} = \frac{1}{(2\pi)^{d/2} |\mathrm{dét}\, A|} \int_B \exp\Big(-\frac{1}{2} \langle A^{-1}x, A^{-1}x \rangle\Big)\, \mathrm{d}x\,.$$

En effet, d'après la proposition précédente,

$$P\{\, X \in B \,\} = P\big\{\, G \in A^{-1}(B) \,\big\}\,,$$

ce qui se calcule en utilisant la densité de G.

Remarquons aussi que la proposition IV.4.2 nous dit que pour simuler numériquement un vecteur X de loi $\mathcal{N}(0, \Gamma)$, il suffit de savoir simuler des variables aléatoires réelles de loi $\mathcal{N}(0, 1)$. Ayant simulé un vecteur aléatoire X de loi $\mathcal{N}(0, \Gamma)$, on peut simuler un vecteur de loi $\mathcal{N}(m, \Gamma)$ en considérant $X + m$.

On peut être plus précis sur la décomposition précédente de la matrice de covariance Γ. Même si celle-ci est dégénérée, on peut toujours l'écrire sous la forme $\Gamma = P \Delta {}^{\mathrm{t}}P$ où P est une matrice orthogonale (*i.e.* $P^{-1} = {}^{\mathrm{t}}P$) et Δ est une matrice diagonale positive, avec éventuellement des zéros sur la diagonale rendant compte des dégénérescences de Γ (ou de X). (Un cas extrême est par exemple

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

qui est la matrice de covariance du vecteur gaussien $X = (X_1, X_2, X_3)$ où X_1 suit la loi $\mathcal{N}(0,1)$ et X_2 et X_3 sont de loi $\mathcal{N}(0,0)$ *i.e.* $X_2 = X_3 = 0$ p.s.) Ainsi,

$$\Gamma = P\sqrt{\Delta}\sqrt{\Delta}\,{}^{\mathrm{t}}P = P\sqrt{\Delta}\;{}^{\mathrm{t}}(P\sqrt{\Delta})\,,$$

et $A = P\sqrt{\Delta}$.

En terme de changement de base, on notera que le vecteur gaussien ${}^{\mathrm{t}}PX$ a pour matrice de covariance la matrice diagonale Δ (puisque $P\sqrt{\Delta}G$ a même loi que X, donc ${}^{\mathrm{t}}PX$ a même loi que $\sqrt{\Delta}G$, qui a pour matrice de covariance Δ).

La diagonalisation de la matrice de covariance Γ d'un vecteur gaussien centré X nous a donc permis de déterminer une nouvelle base dans laquelle les composantes de X sont orthogonales. L'intérêt de cette observation provient du théorème suivant qui est une autre façon de formuler la proposition IV.4.2.

Théorème IV.4.3. *Soit $X = (X_i, \ldots, X_d)$ un vecteur gaussien dans $\mathbb{R}^d$ de matrice de covariance Γ. Si les composantes de X sont deux à deux non corrélées (i.e. Γ est diagonale), alors la famille $(X_1, \ldots, X_d)$ est mutuellement indépendante.*

Démonstration. Nous nous contentons du cas centré. Pour tous $u_1, \ldots, u_d \in \mathbb{R}$,

$$\begin{aligned}
E\Big(\exp\Big(i\sum_{1\le k\le d} u_k X_k\Big)\Big) &= \exp\Big(-\frac{1}{2}E\Big(\Big(\sum_{1\le k\le d} u_k X_k\Big)^2\Big)\Big)\\
&= \exp\Big(-\frac{1}{2}\sum_{1\le k\le d} u_k E(X_k^2)\Big)\\
&= \prod_{1\le k\le d} E\big(\exp(iu_k X_k)\big)\,.
\end{aligned}$$

Ainsi la fonction caractéristique du vecteur $X = (X_1, \ldots, X_d)$ est le produit des fonctions caractéristiques des lois marginales. Le corollaire IV.1.12 montre alors que les variables aléatoires X_i sont indépendantes. On notera que l'orthogonalité $E(X_i X_j) = 0$, $i \neq j$, est évidemment nécessaire. □

Si donc ${}^{\mathrm{t}}PX$ a pour covariance la matrice diagonale Δ, c'est que dans la nouvelle base de $\mathbb{R}^d$, de matrice de passage ${}^{\mathrm{t}}P$, le nouveau vecteur gaussien est à composantes indépendantes. On se ramène preque toujours à cette réduction. Comme Δ peut avoir des zéros sur la diagonale, le nombre de termes diagonaux non nuls est en fait le rang de Γ ou le rang du vecteur gaussien X : il est en effet à valeurs dans un sous-espace de $\mathbb{R}^d$ de dimension égale à ce rang.

Exercices

Exercice IV.1. Une urne contient r boules rouges et b boules blanches. On tire ces boules une à une, sans remise, jusqu'à épuisement. Pour $0 \leq k \leq b$, quelle est la probabilité pour qu'exactement k boules blanches soient tirées avant la première boule rouge ?

Exercice IV.2. Deux joueurs A et B jouent une suite de parties indépendantes. Lors de chacune d'elles, ils ont respectivement les probabilités p pour A et $q = 1 - p$ pour B de gagner. Le vainqueur final est celui des deux joueurs qui le premier obtient 2 victoires de plus que son adversaire. Quelle est la probabilité pour que A soit vainqueur ?

Exercice IV.3. Vérifier l'indépendance des A_n dans l'exemple IV.1.4.i.

Exercice IV.4. Soient X et Y deux variables définies sur $(\Omega, \mathcal{A}, P)$, ne pouvant prendre que deux valeurs distinctes. Montrer que X et Y sont indépendantes si et seulement si $E(XY) = E(X)E(Y)$.
Indication : Si X prend les valeurs x_1, x_2 et Y les valeurs y_1, y_2, déduire de l'hypothèse que

$$E\big((X - x_i)(Y - y_j)\big) = E(X - x_i)E(Y - y_j)\,, \quad i, j = 1,\, 2\,.$$

Exercice IV.5. Soit X une variable aléatoire réelle et soient f et g deux fonctions croissantes de $\mathbb{R}$ dans $\mathbb{R}$. On suppose que $E(f(X)^2) < \infty$ et $E(g(X)^2) < \infty$. Démontrer que

$$E\big(f(X)g(X)\big) \geq E\big(f(X)\big)E\big(g(X)\big)\,.$$

Indication : remarquer que $(f(x) - f(y))(g(x) - g(y)) \geq 0$ pour tous $x, y \in \mathbb{R}$ et utiliser le théorème de Fubini après avoir introduit une variable Y indépendante de X et de même loi que X.

En déduire que si $|X| < 1$ p.s.,

$$E\Big(\frac{1}{1 - X^2}\Big) \leq E\Big(\frac{1}{1 - X}\Big)\Big(\frac{1}{1 + X}\Big)\,.$$

Exercice IV.6. Soient X et Y deux variables aléatoires indépendantes, de même loi exponentielle de densité $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}_{[0,\infty[}(x)$, $\theta > 0$. Déterminer les densités des lois de X^3, $|X - Y|$, $\min(X, Y^3)$. Même question lorsque X et Y suivent la loi uniforme sur $[-1, 1]$.

Exercice IV.7. Soient F et G deux fonctions de répartition et U une variable aléatoire de loi uniforme sur $]0, 1[$. Montrer que $V(x, y) = \min(F(x), G(y))$ est la fonction de répartition du vecteur aléatoire $(F^{\leftarrow}(U), G^{\leftarrow}(U))$. En particulier, V est de marges F et G.

Montrer que si H est une fonction de répartition sur $\mathbb{R}^2$ de marges F et G, alors $H \leq V$.

Exercice IV.8. Soient X_i, $1 \leq i \leq n$, des variables aléatoires indépendantes, X_i étant de fonction de répartition F_i. Soit $m_n = \min_{1\leq i\leq n} X_i$ et $M_n = \max_{1\leq i\leq n} X_i$. Montrer que la fonction de répartition de M_n en x est $\prod_{1\leq i\leq n} F_i(x)$, que celle de m_n est $1 - \prod_{1\leq i\leq n}(1 - F_i(x))$ et que

$$P\{\, x_1 < m_n \leq M_n \leq x_2 \,\} = \prod_{1\leq i\leq n} \bigl(F_i(x_2) - F_i(x_1)\bigr)\,.$$

Indication : $\{\, M_n \leq x \,\} = \bigcap_{1\leq i\leq n}\{\, X_i \leq x \,\}$.

Exercice IV.9. Soient $X_1, \ldots, X_n$ des variables indépendantes de même loi exponentielle de paramètre 1. Montrer que $P\{\, \exists\, i, j \,:\, X_i = X_j \,\} = 0$. On pose

$$Z = \min_{1\leq i\leq n} X_i \qquad \text{et} \qquad N = \min\{\, 1 \leq i \leq n \,:\, X_i = Z \,\}\,.$$

Déterminer la loi de Z. Établir que

$$P\{\, N = k\,,\, Z > t \,\} = e^{-nt}/n\,, \qquad k = 1, \ldots, n\,, \qquad t > 0\,.$$

En déduire que Z et N sont des variables aléatoires indépendantes et préciser la loi de N.

Exercice IV.10. Soit P une loi sur $\mathbb{R}$ dont on suppose qu'elle admet une transformée de Laplace $L(t) = \int e^{tx}\,\mathrm{d}P(x)$ pour $|t|$ petit. Soit P^{*n} la n-ième convoluée de P avec elle-même, définie par $P^{*1} = P$ et $P^{*n} = P^{*(n-1)} * P$ (*i.e.* P^{*n} est la loi d'une somme de n variables aléatoires indépendantes de loi P). Soit t tel que $L(t)$ existe et soit P_t la loi définie par sa densité $\dfrac{\mathrm{d}P_t}{\mathrm{d}P} = \dfrac{e^{tx}}{L(t)}$. Montrer que P_t^{*n} admet une densité par rapport à P^{*n} donnée par $\dfrac{\mathrm{d}P_t^{*n}}{\mathrm{d}P^{*n}} = \dfrac{e^{tx}}{L(t)^n}$.

Montrer que $P^{*n}([\,x, \infty\,[) \leq e^{-tx}L(t)^n P_t^{*n}([\,x, \infty\,[)$ pour $t > 0$ (comparer cette inégalité avec celle de Chernoff, III.4.10.iii).

Exercice IV.11. On appelle loi gamma de paramètre $p > 0$ et on note γ_p la loi de densité $\gamma_p(x) = (\Gamma(p))^{-1}x^{p-1}e^{-x}$ sur $\mathbb{R}^+$, où $\Gamma(p)$ assure que $\int \gamma_p(x)\,\mathrm{d}x = 1$. Montrer que $\Gamma(p) = (p-1)\Gamma(p-1)$ et que pour p entier, $\Gamma(p) = (p-1)!$.

Montrer que $\Gamma_p * \Gamma_q = \Gamma_{p+q}$. En déduire la loi de $\lambda_1 + \cdots + \lambda_n$ où les λ_i sont des variables aléatoires indépendantes et de loi exponentielle de paramètre 1.

Montrer que la fonction caractéristique de la loi Γ_p est $(1 - it)^{-p}$.

Soit maintenant $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et de même loi exponentielle. Soit $S_n = X_1 + \cdots + X_n$ leur somme. Pour $t \geq 0$, soit $N(t) = \mathrm{card}\{\, i \,:\, S_i \leq t \,\}$. En évaluant $P\{\, N(t) \geq k \,\}$, montrer que $N(t)$ suit une loi de Poisson de paramètre t.

Exercice IV.12. Soient $X_1, \ldots, X_n, X_{n+1}$ des variables aléatoires indépendantes de loi exponentielle de paramètre 1. Calculer la loi de la somme $S_k = X_1 + \cdots + X_k$, $1 \leq k \leq n+1$. Démontrer que la loi du vecteur $(U_1, \ldots, U_n)$ défini par $U_i = S_i/S_{n+1}$, $i = 1, \ldots, n$, a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ donnée par $n!\, \mathbb{1}_D$, où

$$D = \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \,;\, 0 \leq x_1 \leq \cdots \leq x_n \leq 1 \right\}.$$

Exercice IV.13. Soient $X_1, \ldots, X_n$ des variables aléatoires réelles, indépendantes, de même loi de fonction de répartition F ayant une densité f. Ces variables, ordonnées par ordre croissant, sont notées $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Clairement les $X_{i,n}$, $1 \leq i \leq n$, ne sont pas indépendantes puisque par construction $X_{i,n} \leq X_{i+1,n}$.
a) Montrer que la probabilité que k des variables $X_1, \ldots, X_n$ soient inférieures à x et $n-k$ soient supérieures à x est $\mathrm{C}_n^k F(x)^k (1 - F(x))^{n-k}$. En déduire que $P\{ X_{i,n} \leq x \} = \sum_{i \leq k \leq n} \mathrm{C}_n^k F(x)^k (1 - F(x))^{n-k}$, et que $X_{i,n}$ admet pour densité

$$f_{i,n}(x) = i\, \mathrm{C}_n^i\, f(x) F(x)^{i-1} \big(1 - F(x)\big)^{n-i}, \quad x \in \mathbb{R}.$$

b) Montrer par un argument analogue que pour $x, y \in \mathbb{R}$

$$P\{ X_{i,n} \leq x ; X_{i+1,n} > y \} = \mathrm{C}_n^i F(x)^i \big(1 - F(y)\big)^{n-i}.$$

c) En déduire la fonction de répartition du couple $(X_{i,n}, X_{i+1,n})$.
d) Montrer que le couple $(X_{i,n}, X_{i+1,n})$ admet pour densité

$$f_{i,i+1,n}(x,y) = i(n-i)\mathrm{C}_n^i f(x) f(y) F(x)^{i-1} \big(1 - F(y)\big)^{n-i-1}, -\infty < x < y < \infty.$$

e) Soit $S_{i+1,n} = X_{i+1,n} - X_{i,n}$. Montrer que le couple $(X_{i,n}, S_{i+1,n})$ admet pour densité

$$g(x,s) = i(n-i)\mathrm{C}_n^i f(x) f(x+s) F(x)^{i-1} \big(1 - F(x+s)\big)^{n-i-1}, \qquad x \in \mathbb{R}, \ s \geq 0.$$

f) Supposons les X_i de loi exponentielle de paramètre 1. Montrer qu'alors $S_{i+1,n}$ est de loi exponentielle de paramètre $n-i$.

Exercice IV.14. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, de même loi de Bernoulli de paramètre p. Pour tout $n \geq 1$, on définit par récurrence, $T_n = \inf\{ k > T_{n-1} \,;\, X_k = 1 \}$ si cet infimum est fini, $T_n = \infty$ sinon, et $T_0 = 0$. Démontrer que les variables aléatoires T_1, $T_2 - T_1, \ldots, T_n - T_{n-1}, \ldots$ sont indépendantes et de même loi. Calculer la loi de T_1 et sa fonction caractéristique. En déduire la loi de T_n.

Exercice IV.15. Versions du lemme de Borel-Cantelli.

(i) Si $\sum_{n\geq 1} P(A_n) = \infty$ et $\liminf_{n\to\infty} \dfrac{\left(\sum_{1\leq i\leq n} P(A_i)\right)^2}{\sum_{1\leq i\leq n}\sum_{1\leq j\leq n} P(A_i\cap A_j)} = 1$ alors $P(A_n \text{ i.s. }) = 1$ (Rényi).

Indication : Appliquer l'inégalité de l'exercice III.10 à $X = \sum_{1\leq i\leq n} \mathbb{1}_{A_i}$ pour tout $n\geq 1$ pour démontrer que $\sum_{i\geq 1} \mathbb{1}_{A_i} = \infty$ p.s.

(ii) Si $\sum_{n\geq 1} P(A_n) = \infty$ et $P(A_i\cap A_j) \leq cP(A_i)P(A_j)$ pour un $c > 0$ et tous $i\neq j$, alors $P(A_n \text{ i.s.}) > 0$ (Kotska).

Exercice IV.16. Inégalité de Kolmogorov. Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes d'espérance 0 et de variance finie. Soit $S_n = X_1+\cdots+X_n$. Montrer l'inégalité de Kolmogorov,

$$P\Big\{ \max_{1\leq k\leq n} |S_k| \geq t \Big\} \leq t^{-2} \sum_{1\leq i\leq n} \mathrm{Var}(X_i)$$

pour tout $t > 0$.

Indication : considérer les événements disjoints

$$A_k = \bigcap_{j<k} \{ |S_j| < t \} \cap \{ |S_k| \geq t \}, \qquad 1\leq k\leq n,$$

et commencer par montrer la minoration

$$E(S_n^2) \geq \sum_{1\leq k\leq n} \int_{A_k} S_k^2 \,\mathrm{d}P.$$

Puis utiliser l'inégalité de Markov,

$$P(A_k) \leq t^{-2} \int_{A_k} S_k^2 \,\mathrm{d}P.$$

Exercice IV.17. Trouver une fonction h de $\mathbb{R}$ dans $\mathbb{R}$ et un réel $c > 0$ tel que la fonction

$$f(x,y) = \frac{c^2}{2\pi} e^{-(x^2+y^2)/2} + h(x)h(y), \qquad (x,y)\in\mathbb{R}^2,$$

soit la densité de la loi d'un vecteur non gaussien de $\mathbb{R}^2$, dont les lois marginales sont gaussiennes.

Exercice IV.18. Soit (X,Y) un vecteur gaussien, centré, à valeurs dans $\mathbb{R}^2$, de matrice de covariance $\Sigma = \begin{pmatrix} 3 & 6 \\ 6 & 12 \end{pmatrix}$. Démontrer que X et Y sont proportionnelles.

Exercice IV.19. Soit X une variable aléatoire suivant une loi $\mathcal{N}(0,1)$, et soit ε une variable de Bernoulli telle que $P\{\,\varepsilon = +1\,\} = P\{\,\varepsilon = -1\,\} = 1/2$, indépendante de X. Démontrer que εX et $\varepsilon|X|$ ont même loi que X. Le couple $(X, \varepsilon X)$ est-il gaussien ?

Exercice IV.20. Soit X un vecteur gaussien centré, à valeurs dans $\mathbb{R}^d$, et soit Y une copie indépendante de X. On pose $X_\theta = X\cos\theta + Y\sin\theta$ et $X'_\theta = -X\sin\theta + Y\cos\theta$, $\theta \in [\,0, 2\pi\,]$. Démontrer que pour tout θ, X_θ et X'_θ sont indépendantes, de même loi que X.

Exercice IV.21. Soient X et Y deux vecteurs aléatoires de $\mathbb{R}^d$, indépendants et de même loi, tels que $X+Y$ et $X-Y$ sont indépendants. On désigne par φ la fonction caractéristique de la loi de X.
a) Montrer que pour tous $s, t \in \mathbb{R}^d$,

$$\varphi(s+t)\varphi(s-t) = \varphi(s)^2|\varphi(t)|^2 \,.$$

En déduire l'existence d'une fonction continue ψ sur $\mathbb{R}^d$ telle que $\varphi = e^\psi$.
b) On pose $\psi_p(t) = \frac{1}{2}(\psi(t) + \psi(-t))$ et $\psi_i(t) = \frac{1}{2}(\psi(t) - \psi(-t))$, $t \in \mathbb{R}^d$. Démontrer qu'il existe $m \in \mathbb{R}^d$ tel que $\psi_i(t) = i\langle m, t\rangle$, $t \in \mathbb{R}^d$.
c) Soit $Q(s,t) = \psi_p(s+t) - \psi_p(s) - \psi_p(t)$, $s, t \in \mathbb{R}^d$. Démontrer que Q est réelle, symétrique négative. Établir que Q est bilinéaire.
d) Déduire de ce qui précède que la loi de X est gaussienne.

Exercice IV.22. (Lois infiniment divisibles) Soit X une variable aléatoire réelle sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de loi μ ; on dit que μ est infiniment divisible si, pour chaque entier $n \geq 1$, il existe des variables aléatoires réelles $X_{1,n}, \ldots, X_{n,n}$ indépendantes et de même loi ν_n telles que la loi de la somme $X_{1,n} + \cdots + X_{n,n}$ soit μ.
a) Démontrer qu'une loi μ est infiniment divisible si et seulement si sa fonction caractéristique φ est, pour tout entier $n \geq 1$, la puissance n-ième d'une fonction caractéristique.
b) μ est-elle infiniment divisible dans les cas suivant : (i) $\mu = \delta_a$, $a \in \mathbb{R}$; (ii) μ est la loi gaussienne de moyenne m et de variance σ^2 ; (iii) μ est la loi de Poisson de paramètre λ ; (iv) μ est la loi de Cauchy (on rappelle que la fonction caractéristique de la loi de Cauchy est donnée par $e^{-|t|}$) ?
c) Soit X de loi μ de Bernoulli sur $\{\,0,1\,\}$ de paramètre $0 < p < 1$; soient également Y et Z des variables aléatoires indépendantes de loi commune ν telles que la somme $Y+Z$ soit de loi μ.
(i) Si B est un intervalle ne contenant pas 0 et $1/2$, démontrer que $\mu(B+B) = 0$ (où $B+B = \{\,x+y : x, y \in B\,\}$). En déduire que $\nu \otimes \nu(B \times B) = 0$.
(ii) Déduire de la question précédente que Y ne peut prendre que les valeurs 0 et $1/2$.

(iii) Conclure que μ n'est pas infiniment divisible.
d) Soit φ une fonction caractéristique, et soit $\lambda > 0$. On définit

$$\Phi(t) = e^{\lambda(\varphi(t)-1)}, \quad t \in \mathbb{R}.$$

Sur $(\Omega, \mathcal{A}, P)$, on considère une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de même loi de fonction caractéristique φ, ainsi qu'un variable aléatoire N suivant une loi de Poisson de paramètre λ, indépendante de la suite $(X_n)_{n\in\mathbb{N}}$. Pour chaque $\omega \in \Omega$, on pose

$$Y(\omega) = \sum_{1\leq k\leq N(\omega)} X_k(\omega)$$

(avec la convention $\sum_{1\leq k\leq 0} = 0$). Démontrer que Y est une variable aléatoire de fonction caractéristique Φ. Montrer que la loi de Y est infiniment divisible.

V

CONVERGENCE DE SUITES DE VARIABLES ALÉATOIRES

Il existe de nombreuses notions de convergence de variables aléatoires. Elle sont essentielles pour les applications. Elles servent surtout à montrer que les phénomènes aléatoires présentent certaines régularités, à partir desquelles on peut identifier certaines de leurs propriétés. Par exemple, nous avons vu à l'exemple IV.2.4.iii que la fréquence observée des piles dans un jeu de pile ou face, après n tirages, est « proche » de la probabilité (déterministe) p d'obtenir pile, pourvu que n soit grand. Donc, si p est inconnue (par exemple nous ne savons pas si la pièce est truquée), nous avons là un moyen de l'approximer.

Dans tout ce chapitre, les suites de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ sont supposées construites sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. Pour plus de simplicité, nous ne considérons que des variables aléatoires à valeurs réelles. Les énoncés et les résultats subsistent sans modifications pour des vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ (pour l'essentiel, remplacer les valeurs absolues par une norme sur $\mathbb{R}^d$).

V.1. Convergence presque sûre

Définition V.1.1. Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$, définie sur $(\Omega, \mathcal{A}, P)$, converge presque sûrement (p.s.) vers la variable aléatoire X, définie sur $(\Omega, \mathcal{A}, P)$, si

$$P\Big\{\, \omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega) \,\Big\} = 1\,.$$

Dans ce cas, on note $\lim_{n\to\infty} X_n = X$ p.s. ou $X_n \to X$ p.s. lorsque $n \to \infty$.

Observons que l'événement $\{\lim_{n\to\infty} X_n = X\}$ est bien mesurable puisque égal à

$$\bigcap_{p\geq 1}\bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{\,|X_n - X| < 1/p\,\}\,.$$

Or si $(A_p)_{p\in\mathbb{N}}$ est une suite d'événements, la proposition I.4.3.i et le fait qu'une mesure de probabilité est bornée par 1 montrent que $P(\bigcap_{p\in\mathbb{N}} A_p) = 1$ si et seulement si $P(A_p) = 1$ pour tout p. Il s'ensuit que X_n converge vers X p.s. si et seulement si

$$\forall\varepsilon > 0\,,\ P\Big(\bigcup_{m\in\mathbb{N}}\bigcap_{n\geq m}\{\,|X_n - X| < \varepsilon\,\}\Big) = 1$$

(prendre $1/(p+1) < \varepsilon \leq p$). Cette condition peut aussi s'écrire, par passage au complémentaire,

$$\forall\varepsilon > 0\,,\ P\Big(\bigcap_{m\in\mathbb{N}}\bigcup_{n\geq m}\{\,|X_n - X| \geq \varepsilon\,\}\Big) = 0\,.$$

Elle est alors équivalente à

$$\forall\varepsilon > 0\,,\ P\big\{\,|X_n - X| \geq \varepsilon \text{ i.s. }\,\big\} = 0\,. \tag{1}$$

Par convergence monotone, c'est encore équivalent à

$$\forall\varepsilon > 0\,,\ \lim_{m\to\infty} P\Big\{\sup_{n\geq m}|X_n - X| \geq \varepsilon\,\Big\} = 0\,. \tag{2}$$

La convergence p.s. peut aussi être décrite à l'aide du critère de Cauchy. Par exemple $X_n \to X$ p.s. si et seulement si

$$\forall\varepsilon > 0\,,\ P\Big(\bigcup_{n\in\mathbb{N}}\bigcap_{m\geq n}\{\,|X_n - X_m| < \varepsilon\,\}\Big) = 1\,.$$

On peut aussi dire, quitte à enlever un ensemble de mesure nulle (celui pour lequel $X_n(\omega)$ ne converge pas vers $X(\omega)$), que $X_n \to X$ p.s. si et seulement si X_n converge ponctuellement vers X, en tant que suite de fonctions de Ω dans $\mathbb{R}$. Il s'ensuit que si ϕ est une fonction continue sur $\mathbb{R}$, alors $\phi(X_n)$ converge vers $\phi(X)$ presque sûrement. En particulier, si X_n et Y_n, $n\in\mathbb{N}$, sont deux suites de variables aléatoires réelles convergeant presque sûrement vers X et Y, alors pour tous a et b réels, $aX_n + bY_n$ converge presque sûrement vers $aX + bY$ et X_nY_n converge presque sûrement vers XY.

Un des outils classiques pour montrer la convergence presque sûre est le lemme de Borel-Cantelli.

Proposition V.1.2 (Lemme de Borel-Cantelli). *Soient X_n, $n \in \mathbb{N}$, et X, des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, P)$.*
(i) Si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} P\{\, |X_n - X| \geq \varepsilon \,\} < \infty$, alors $X_n \to X$ p.s.
(ii) Si les $(X_n)_{n \in \mathbb{N}}$ sont mutuellements indépendantes, alors $X_n \to 0$ p.s. si et seulement si $\sum_{n \in \mathbb{N}} P\{\, |X_n| \geq \varepsilon \,\} < \infty$ pour tout $\varepsilon > 0$.

Démonstration. Pour montrer (i), soit $\varepsilon > 0$ et les événements

$$A_n = \{\, |X_n - X| \geq \varepsilon \,\}, \quad n \in \mathbb{N}.$$

Appliquons le lemme de Borel-Cantelli (IV.3.5) aux A_n. Il vient $P(A_n \text{ i.s.}) = 0$, ce qui fournit le résultat d'après (1). (Voir aussi IV.3.6.i.)

(ii) se démontre de façon analogue à partir de la partie indépendante du lemme de Borel-Cantelli. (Noter qu'il convient de supposer X nulle, ou constante, sans quoi les événements A_n ne sont pas nécessairement indépendants.) □

***Exemples* V.1.3.** (i) Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de Bernoulli $\mathcal{B}(1,p)$, *i.e.* $P\{\, X_i = 1 \,\} = 1 - P\{\, X_i = 0 \,\} = p$. Soit $U_n = \sum_{1 \leq i \leq n} 2^{-i} X_i$. Pour montrer la convergence p.s de U_n, on peut appliquer le critère de Cauchy, en remarquant que $n < m$ implique

$$|U_m - U_n| \leq \sum_{n+1 \leq i \leq m} 2^{-i} \leq 2^{-n}.$$

Ainsi

$$\begin{aligned}
\bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \{\, \omega : |U_n(\omega) - U_m(\omega)| < \varepsilon \,\} &\supset \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \{\, \omega : 2^{-n} < \varepsilon \,\} \\
&= \bigcup_{n \in \mathbb{N}} \{\, \omega : 2^{-n} < \varepsilon \,\} \\
&= \Omega .
\end{aligned}$$

Notons U la limite $\sum_{i \geq 1} 2^{-i} X_i$. C'est une variable aléatoire à valeurs dans $[\,0,1\,]$ puisque $0 \leq U \leq \sum_{i \geq 1} 2^{-i} = 1$.

(ii) Souvent on ne peut appliquer le lemme de Borel-Cantelli que sur des sous-suites, et un argument supplémentaire est nécessaire pour conclure. Un exemple de cette situation est le suivant. Soient X_i, $i \geq 1$, des variables aléatoires indépendantes, de loi exponentielle $P\{\, X_i > t \,\} = e^{-t}$, $t \geq 0$. Soit $M_n = \max_{1 \leq i \leq n} X_i$. Alors

$$P\{\, M_n \leq t \,\} = P\Big(\bigcap_{1 \leq i \leq n} \{\, X_i \leq t \,\} \Big) = (1 - e^{-t})^n .$$

Montrons que $M_n/\ln n \to 1$ p.s., c'est-à-dire que le maximum de n variables aléatoires exponentielles, indépendantes, se comporte p.s. comme $\ln n$ lorsque n est grand. La démonstration consiste à prouver que pour tout $\varepsilon > 0$, on a une minoration $M_n \geq (1-\varepsilon)\ln n$ p.s. pour tout n assez grand, puis que l'on a une majoration $M_n \leq (1+\varepsilon)\ln n$ p.s. pour tout n assez grand.

Commençons par la minoration. Soit $0 < \varepsilon \leq 1$. On a

$$\begin{aligned} P\{\, M_n \leq (1-\varepsilon)\ln n \,\} &= \big(1 - n^{-1+\varepsilon}\big)^n \\ &= \exp\big(n\ln(1-n^{1-\varepsilon})\big) \\ &= \exp\big(-n^{\varepsilon}(1+o(1))\big) \end{aligned}$$

lorsque n tend vers l'infini. Donc $\sum_n P\{\, M_n/\ln n \leq 1-\varepsilon \,\} < \infty$ pour tout $\varepsilon > 0$, et par le lemme de Borel-Cantelli IV.3.5, $M_n/\ln n \geq 1-\varepsilon$ p.s. pour n assez grand. En particulier,

$$\liminf_{n\to\infty} M_n/\ln n \geq 1-\varepsilon \quad \text{p.s.}$$

Établissons maintenant la majoration. On a

$$\begin{aligned} P\{\, M_n \geq (1+\varepsilon)\ln n \,\} &= 1 - P\{\, M_n < (1+\varepsilon)\ln n \,\} \\ &= 1 - \big(1-n^{-1-\varepsilon}\big)^n \\ &= 1 - \exp\big(n\ln(1-n^{-1-\varepsilon})\big) \\ &= 1 - \exp\big(-n^{-\varepsilon}(1+o(1))\big) \\ &= n^{-\varepsilon}(1+o(1)) \end{aligned}$$

lorsque n tend vers l'infini. Soit la sous-suite $n_k = \lfloor (k+1)^{\delta} \rfloor$, $k \in \mathbb{N}$, avec $\delta\varepsilon > 1$ où $\lfloor\cdot\rfloor$ est la fonction partie entière. Alors

$$\sum_{k\in\mathbb{N}} P\{\, M_{n_k} \geq (1+\varepsilon)\ln n_k \,\} < \infty\,.$$

Donc $\limsup_{k\to\infty} M_{n_k}/\ln n_k \leq 1+\varepsilon$ p.s. Pour conclure la majoration, pour tout $n \in \mathbb{N}^*$, il existe $k \in \mathbb{N}$ tel que $n_k \leq n < n_{k+1}$. Et donc, en remarquant que la suite M_n est croissante,

$$\frac{M_n}{\ln n} \leq \frac{M_{n_{k+1}}}{\ln n_{k+1}} \cdot \frac{\ln n_{k+1}}{\ln n_k}\,.$$

Puisque $\lim_{k\to\infty}(\ln n_k)/(\ln n_{k+1}) = 1$, on conclut que

$$\limsup_{n\to\infty} M_n/\ln n \leq 1+\varepsilon \quad \text{p.s.}$$

Ainsi, pour tout $0 < \varepsilon \leq 1$, presque sûrement,

$$1 - \varepsilon \leq \liminf_{n\to\infty} \frac{M_n}{\ln n} \leq \limsup_{n\to\infty} \frac{M_n}{\ln n} \leq 1 + \varepsilon\,.$$

En choisissant $\varepsilon \to 0$ le long d'une sous-suite dénombrable, il s'ensuit que $M_n/\ln n \to 1$ p.s. L'argument que nous avons utilisé dans la majoration, consistant à découper l'ensemble $\mathbb{N}$ en des blocs $[\,n_k, n_{k+1}\,[$, s'appelle un argument de bloc.

V.2. Convergence en probabilité

La convergence en probabilité, appelée aussi convergence en mesure, ou dans $\mathrm{L}^0(\Omega, \mathcal{A}, P)$, est définie comme suit.

Définition V.2.1. Soient X_n, $n \in \mathbb{N}$, X, des variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$. On dit que X_n converge en probabilité vers X, et on note $X_n \xrightarrow{P} X$, ou $\lim_{n\to\infty} X_n = X$ en probabilité, ou $P - \lim_{n\to\infty} X_n = X$, si pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} P\big\{\,|X_n - X| \geq \varepsilon\,\big\} = 0\,.$$

Autrement dit, pour tout $\varepsilon > 0$, il existe n_0 tel que pour tout entier $n \geq n_0$, $P\{\,|X_n - X| \geq \varepsilon\,\} \leq \varepsilon$.

On mesure tout de suite la différence avec la convergence presque sûre qui exige un supremum (voir (2), p. 110). En particulier, la convergence p.s. implique la convergence en probabilité.

Exemples V.2.2. (i) Soit X_i, $i \geq 1$, des variables aléatoires réelles non corrélées, telles que $E(X_i) = 0$ et $\mathrm{Var}(X_i) = \sigma^2$ pour tout $i \geq 1$. Alors leurs moyennes partielles $n^{-1}\sum_{1\leq i\leq n} X_i$ convergent en probabilité vers 0. En effet, pour tout $\varepsilon > 0$,

$$P\bigg\{\,\bigg|\frac{1}{n}\sum_{1\leq i\leq n} X_i\bigg| \geq \varepsilon\,\bigg\} \leq \frac{1}{n^2\varepsilon^2}\mathrm{Var}\Big(\sum_{1\leq i\leq n} X_i\Big) = \frac{\sigma^2}{n\varepsilon^2}$$

d'après l'inégalité de Tchebitchev (III.4.10.ii). Ceci complète l'exemple IV.2.4.iii et la discussion introduisant ce chapitre. Il convient de comprendre que cette convergence a un sens tout à fait concret. Il suffit de tirer une pièce non truquée une cinquantaine de fois pour voir que la proportion de piles se stabilise vers 1/2 si l'on n'est pas trop malchanceux. C'est exactement ce que dit le résultat de convergence.

(ii) Soit $\Omega = [\,0,1\,]$ muni de sa tribu borélienne et de la probabilité uniforme P (mesure de Lebesgue). Définissons pour $\omega \in \Omega$, et $n \in \mathbb{N}$, $1 \leq k \leq 2^n$, $i = 2^n + k - 1$,

$$X_i(\omega) = \mathbb{1}_{](k-1)/2^n, k/2^n]}(\omega)\,.$$

Alors pour tout $\omega \in \,]\,0,1\,]$, $\liminf_{i\to\infty} X_i(\omega) = 0$ et $\limsup_{n\to\infty} X_n(\omega) = 1$, de sorte que la suite $(X_i)_{i\geq 1}$ ne converge pas presque sûrement. Cependant, pour tout $\varepsilon \in \,]\,0,1\,[$, $i = 2^n + k - 1$, $1 \leq k \leq 2^n$, $n \in \mathbb{N}$, on a $P\{\,|X_i| \geq \varepsilon\,\} = 2^{-n}$, de sorte que X_i converge en probabilité vers 0.

(iii) Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, de loi de Bernoulli, avec $P\{\,X_n = 1\,\} = 1 - P\{\,X_n = 0\,\} = p_n$. Alors

$$X_n \xrightarrow{P} 0 \qquad \longleftrightarrow \qquad \lim_{n\to\infty} p_n = 0$$

puisque $P\{\,|X_n| \geq \varepsilon\,\} = p_n$ si $0 < \varepsilon < 1$. Mais le lemme de Borel-Cantelli V.1.2 montre que

$$\lim_{n\to\infty} X_n = 0 \text{ p.s.} \qquad \longleftrightarrow \qquad \sum_n P\{\,|X_n| \geq \varepsilon\,\} < \infty \text{ pour tout } \varepsilon > 0$$

ce qui est équivalent à $\sum_{n\in\mathbb{N}} p_n < \infty$.

Il est possible de définir une distance qui métrise la convergence en probabilité des variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Si X et Y sont deux variables aléatoires, définies sur $(\Omega, \mathcal{A}, P)$, on pose,

$$d(X,Y) = E\big(|X-Y| \wedge 1\big)\,.$$

Puisque $|X-Y| \wedge 1 \geq 0$, la proposition II.1.4.viii montre que $d(X,Y) = 0$ si et seulement si $X = Y$ p.s. On vérifie facilement que $d(\cdot,\cdot)$ est une distance. Dans ce qui suit, on pourrait aussi utiliser la distance

$$d'(X,Y) = E\Big(\frac{|X-Y|}{1+|X-Y|}\Big)\,.$$

Lemme V.2.3. *La suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X si et seulement si*

$$\lim_{n\to\infty} d(X_n, X) = 0\,.$$

Démonstration. En utilisant l'inégalité de Markov (III.4.9), pour tout $\varepsilon \in \,]\,0,1\,]$,

$$P\{\,|X_n - X| \geq \varepsilon\,\} = P\{\,|X_n - X| \wedge 1 \geq \varepsilon\,\} \leq d(X_n, X)/\varepsilon\,.$$

D'autre part, pour tout $\varepsilon > 0$,

$$\begin{aligned} d(X_n, X) &= \int_{\{|X_n - X| \geq \varepsilon\}} \big(|X_n - X| \wedge 1\big)\,\mathrm{d}P \\ &\qquad + \int_{\{|X_n - X| < \varepsilon\}} \big(|X_n - X| \wedge 1\big)\,\mathrm{d}P \\ &\leq P\big\{\,|X_n - X| \geq \varepsilon\,\big\} + \varepsilon\,. \end{aligned}$$

Si $X_n \xrightarrow{P} X$, il existe n_0 tel que si $n \geq n_0$, $P\{\,|X_n - X| \geq \varepsilon\} \leq \varepsilon$. Donc $d(X_n, X) \leq 2\varepsilon$ pour $n \geq n_0$, et $\lim_{n\to\infty} d(X_n, X) = 0$. □

Répétons encore que la convergence en probabilité n'implique pas la convergence presque sûre. Mais on a le résultat important suivant.

Théorème V.2.4. *Soient X_n, $n \in \mathbb{N}$, X, des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, P)$. Alors X_n converge en probabilité vers X si et seulement si de toute suite déterministe croissante d'entiers (n'), on peut extraire une sous-suite (n'_k) telle que $X_{n'_k} \to X$ p.s.*

Démonstration. Suffisance : Soit (n') une suite croissante d'entiers. Extrayons (n'_k) telle que $X_{n'_k} \to X$ p.s. En particulier, $X_{n'_k} \xrightarrow{P} X$. Or, si de toute sous-suite on peut extraire une sous-suite convergente vers une même limite, la suite converge. On en déduit que X_n converge en probabilité vers X.
Nécessité : Il suffit de considérer $(n') = (n)$. Pour tout $k \geq 1$, soit n_k le plus petit entier tel

$$P\{\,|X_{n_k} - X| \geq 1/k\,\} \leq 2^{-k}\,.$$

Alors,

$$\sum_{k\geq 1} P\big\{\,|X_{n_k} - X| \geq 1/k\,\big\} < \infty\,.$$

En particulier, par le lemme de Borel-Cantelli IV.3.5, $|X_{n_k} - X| < 1/k$ p.s. à partir d'un certain rang, et donc, $\lim_{k\to\infty} X_{n_k} = X$ p.s. □

Il résulte de cette démonstration que la convergence presque sûre n'est pas métrisable, car si elle l'était, elle coïnciderait avec la convergence en probabilité.

La convergence en probabilité est stable par les opérations algébriques usuelles.

Proposition V.2.5. *Soient $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, deux suites de variables aléatoires réelles définies sur un espace $(\Omega, \mathcal{A}, P)$. Supposons que X_n (resp. Y_n) converge en probabilité vers une variable aléatoire X (resp. Y) définie sur $(\Omega, \mathcal{A}, P)$.*

(i) Si ϕ est une application continue de $\mathbb{R}$ à valeurs dans $\mathbb{R}$, alors $\phi(X_n) \xrightarrow{P} \phi(X)$.

(ii) Pour tous α, $\beta \in \mathbb{R}$, $\alpha X_n + \beta Y_n \xrightarrow{P} \alpha X + \beta Y$.

(iii) De plus, $\langle X_n, Y_n \rangle \xrightarrow{P} \langle X, Y \rangle$.

Démonstration. Vérifions par exemple (ii). Soit (n') une suite partielle. On peut extraire une sous-suite (n'') telle que $X_{n''} \to X$ p.s. De (n''), on peut extraire une sous-suite (n''') tel que $Y_{n'''} \to Y$ p.s. Alors $\alpha X_{n'''} + \beta Y_{n'''} \to \alpha X + \beta Y$ p.s. On conclut à l'aide du théorème V.2.4. (Il peut être instructif de démontrer cette proposition sans l'aide du théorème V.2.4.) $\square$

Enfin, l'espace $\mathrm{L}^0(\Omega, \mathcal{A}, P)$ est complet pour la distance d métrisant la convergence en probabilité.

Théorème V.2.6. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, P)$. Supposons qu'elle vérifie le critère de Cauchy en probabilité, c'est-à-dire que*

$$\forall \varepsilon > 0 \, , \; \exists n_0 \, , \; \forall n \geq n_0 \quad P\big\{ \, |X_n - X_{n_0}| \geq \varepsilon \, \big\} \leq \varepsilon \, ,$$

ou de façon équivalente, que

$$\forall \varepsilon > 0 \, , \; \exists n_0 \, , \; \forall n \geq n_0 \, , \; d(X_n, X_{n_0}) \leq \varepsilon \, .$$

Alors X_n converge en probabilité .

Démonstration. En considérant $\varepsilon = 1/2^k$ dans la condition de Cauchy en probabilité, on peut construire une suite d'indices n_k telle que

$$P\big\{ \, |X_{n_k} - X_{n_{k+1}}| \geq 2^{-k} \, \big\} \leq 2^{-k} \, .$$

Alors le lemme de Borel-Cantelli (IV.3.5) montre qu'il existe pour presque tout ω un entier $k_0(\omega) < \infty$ tel que si $k \geq k_0(\omega)$, $|X_{n_k}(\omega) - X_{n_{k+1}}(\omega)| \leq 2^{-k}$. Alors, la suite $X_{n_k}(\omega)$ est de Cauchy. En effet, soit $\varepsilon > 0$ et $p > l > k_0(\omega)$; on a

$$\begin{aligned} \big| X_{n_l}(\omega) - X_{n_p}(\omega) \big| &\leq \sum_{l \leq k \leq p-1} \big| X_{n_k}(\omega) - X_{n_{k+1}}(\omega) \big| \\ &\leq \sum_{l \leq k \leq p-1} 2^{-k} \\ &\leq \sum_{k \geq l} 2^{-k} \leq 2^{-l+1} \, . \end{aligned}$$

Ainsi, si $l_0 \geq k_0(\omega)$ et $2^{-l_0+1} < \varepsilon$, pour tous $p > l > l_0$, $|X_{n_l}(\omega) - X_{n_p}(\omega)| \leq \varepsilon$. Donc X_{n_k} converge p.s. vers une limite X. En particulier, cette sous-suite converge en probabilité vers X. Observons que dans un espace métrisable, une suite de Cauchy dont une sous-suite converge est une suite convergente. Puisque la convergence en probabilité dans l'espace $\mathrm{L}^0(\Omega, \mathcal{A}, P)$ est métrisable, la suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers X. □

Comme pour les suites usuelles (non aléatoires), l'intérêt du critère de Cauchy et du théorème V.2.6 est qu'il assure l'existence d'une limite sans que nous ayons besoin de la calculer explicitement.

V.3. Convergence dans L^p

Nous avons introduit les espaces L^p au chapitre II. Rappelons qu'une variable aléatoire réelle X, définie sur $(\Omega, \mathcal{A}, P)$, est dans $\mathrm{L}^p(\Omega, \mathcal{A}, P)$, $p > 0$, si $E(|X|^p)$ est fini. L'espace $\mathrm{L}^p(\Omega, \mathcal{A}, P)$ est muni de la norme,

$$\|X\|_p = \Big(E\big(|X|^p\big)\Big)^{1/p},$$

qui en fait un espace complet (théorème II.6.4). En particulier, on peut définir une notion de convergence.

***Définition* V.3.1.** Soient $(X_n)_{n\in\mathbb{N}}$, X, des variables aléatoires réelles dans $\mathrm{L}^p(\Omega, \mathcal{A}, P)$, $0 < p < \infty$. On dit que X_n converge vers X dans L^p si $\lim_{n\to\infty} \|X_n - X\|_p = 0$, ou de façon équivalente, $\lim_{n\to\infty} E\big(|X_n - X|^p\big) = 0$.

L'inégalité de Markov III.4.10.i montre que pour tout $p > 0$, la convergence dans L^p implique la convergence en probabilité. Le lemme V.2.3 justifie la terminologie de convergence dans L^0 pour cette dernière.

Les exemples suivants montrent qu'en général la convergence en probabilité, ou même presque sûre, n'implique pas la convergence dans L^p.

***Exemples* V.3.2.** (i) Soit $\Omega =]\,0,1\,]$ muni de sa tribu borélienne et de la probabilité uniforme P. Soit $\alpha > 0$ et

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{]0,1/n]}(\omega)\,, \quad n \geq 1\,.$$

Pour tout $\varepsilon \in \,]\,0,1\,[$, nous avons $P\{\,|X_n| \geq \varepsilon\,\} = 1/n$. Par conséquent, X_n converge en probabilité vers 0. Mais $X_n \notin \mathrm{L}^p$ dès que $\alpha p \geq 1$ puisque

$$E\big(X_n^p\big) = \int_0^{1/n} \omega^{-\alpha p}\,\mathrm{d}\omega = +\infty\,.$$

(ii) Soit $\Omega = \mathbb{R}$ muni de sa tribu borélienne. Pour tout $n \geq 1$, soit X_n une variable aléatoire de loi $(1 - n^{-p})\delta_0 + n^{-p}\delta_n$, c'est-à-dire telle que

$$P\{\, X_n = n \,\} = n^{-p} = 1 - P\{\, X_n = 0 \,\}\,, \qquad p > 1.$$

Si $\varepsilon > 0$, pour tout $n \geq 1/\varepsilon$, $P\{\, |X_n| \geq \varepsilon \,\} = n^{-p}$, et comme $p > 1$, $X_n \to 0$ p.s. (lemme de Borel-Cantelli). En revanche, $E(|X_n|^p) = 1$ pour tout n.

Pour lier la convergence dans les espaces L^p à celle en probabilité, il faut utiliser la notion d'équiintégrabilité, aussi appelée intégrabilité uniforme.

***Définition* V.3.3.** Une famille quelconque $(X_i)_{i \in I}$ de variables aléatoires réelles, définies et intégrables sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ est dite équiintégrable ou uniformément intégrable si

$$\lim_{c \to \infty} \sup_{i \in I} \int_{\{|X_i| > c\}} |X_i| \,\mathrm{d}P = 0\,.$$

Les théorèmes de convergence monotone (II.2.1) ou de convergence dominée (II.2.8) montrent qu'une famille finie de variables aléatoires intégrables est uniformément intégrable. De même, si $|X_i| \leq Y$ p.s. pour tout $i \in I$ et Y est intégrable, alors $(X_i)_{i \in I}$ est équiintégrable.

Rappelons que par convergence dominée, si X est intégrable, pour tout $\varepsilon > 0$ il existe $\eta > 0$ tel que si $P(A) \leq \eta$ alors $\int_A |X| \,\mathrm{d}P \leq \varepsilon$. En effet, le théorème de convergence dominée (II.2.8) implique que pour c assez grand,

$$\int_{\{|X| > c\}} |X| \,\mathrm{d}P \leq \varepsilon/2\,.$$

De plus,

$$\begin{aligned} \int_A |X| \,\mathrm{d}P &\leq \int_{A \cap \{|X| > c\}} |X| \,\mathrm{d}P + \int_{A \cap \{|X| \leq c\}} |X| \,\mathrm{d}P \\ &\leq \varepsilon/2 + cP(A) \\ &\leq \varepsilon/2 + c\eta\,. \end{aligned}$$

Il suffit donc de prendre $\eta = \varepsilon/2c$.

L'analogue uniforme est donné par la proposition suivante.

***Proposition* V.3.4.** *La famille de variables aléatoires réelles intégrables $(X_i)_{i \in I}$ définies sur $(\Omega, \mathcal{A}, P)$ est uniformément intégrable si et seulement si* *(i) pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $A \in \mathcal{A}$, $P(A) \leq \eta$ implique*

$$\forall i \in I \ , \ \int_A |X_i| \,\mathrm{d}P \leq \varepsilon \,,$$

et

(ii) $\sup_{i\in I} \int |X_i| \,\mathrm{d}P < \infty$ *(ou* $\sup_{i\in I} E\big(|X_i|\big) < \infty$*).*

Démonstration. Supposons l'intégrabilité uniforme. Pour tout $\varepsilon > 0$, il existe $c > 0$ tel que

$$\sup_{i\in I} \int_{\{|X_i|>c\}} |X_i| \,\mathrm{d}P \leq \varepsilon/2 \,.$$

Si $A \in \mathcal{A}$, pour tout $i \in I$,

$$\begin{aligned} \int_A |X_i| \,\mathrm{d}P &\leq \int_{A\cap\{|X_i|>c\}} |X_i| \,\mathrm{d}P + \int_{A\cap\{|X_i|\leq c\}} |X_i| \,\mathrm{d}P \\ &\leq \varepsilon/2 + cP(A) \,, \end{aligned}$$

ce qui démontre (i) en prenant $\eta = \varepsilon/2c$ et (ii) en prenant $A = \Omega$.

Réciproquement, soit $M = \sup_{i\in I} \int |X_i| \,\mathrm{d}P < \infty$. Soit de plus $\varepsilon > 0$ et $\eta > 0$ fournis par (i). Posons $c_0 = M/\eta$. Pour tout $c \geq c_0$ et tout i, $P\{\,|X_i| > c\,\} \leq \eta$ (inégalité de Markov, III.4.9). Appliquons alors (i) à $A = \{\,|X_i| > c\,\}$ pour chaque i pour obtenir $\sup_{i\in I} \int_{\{|X_i|>c\}} |X_i| \,\mathrm{d}P \leq \varepsilon$. La conclusion s'ensuit. □

L'intérêt de la notion d'équiintégrabilité apparaît dans le théorème suivant, lequel peut être vu comme une amélioration du théorème de convergence dominée.

Théorème V.3.5. *Soient X_n, $n \in \mathbb{N}$, X, des variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, P)$. Supposons chaque X_n intégrable. Alors, il y a équivalence entre les deux points suivants :*

(i) $X_n \xrightarrow{P} X$ et la famille $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable ;

(ii) X est intégrable et $\lim_{n\to\infty} \|X_n - X\|_1 = 0$.

Démonstration. (i)⇒(ii). Si $X_n \xrightarrow{P} X$, le théorème V.2.4 montre qu'on peut extraire une sous-suite (n_k) telle que X_{n_k} converge vers X p.s. Par le lemme de Fatou (II.2.3) et la proposition V.3.4,

$$E\big(|X|\big) \leq \liminf_{k\to\infty} E\big(|X_{n_k}|\big) \leq \sup_{n\in\mathbb{N}} E\big(|X_n|\big) < \infty \,.$$

Donc $X \in \mathrm{L}^1$. Maintenant, pour tout $\varepsilon > 0$,

$$\begin{aligned} E\big(|X_n - X|\big) &\leq \int_{\{|X_n-X|<\varepsilon\}} |X_n - X|\,\mathrm{d}P + \int_{\{|X_n-X|\geq\varepsilon\}} |X_n|\,\mathrm{d}P \\ &\qquad + \int_{\{|X_n-X|\geq\varepsilon\}} |X|\,\mathrm{d}P. \\ &\leq \varepsilon + \int_{\{|X_n-X|\geq\varepsilon\}} |X_n|\,\mathrm{d}P + \int_{\{|X_n-X|\geq\varepsilon\}} |X|\,\mathrm{d}P\,. \end{aligned}$$

Puisque $X \in \mathrm{L}^1$, $(X_n, n \in \mathbb{N}, X)$ est encore uniformément intégrable. Appliquons la proposition V.3.4 à cette famille et désignons par $\eta = \eta(\varepsilon) > 0$ le réel intervenant dans le point (i) de cette proposition. Par hypothèse, $P\{\,|X_n - X| \geq \varepsilon\,\} \leq \eta$ pour n assez grand. La proposition V.3.4, avec $A = \{\,|X_n - X| \geq \varepsilon\,\}$, montre donc que pour tout n assez grand,

$$\int_{\{|X_n-X|\geq\varepsilon\}} |X_n|\,\mathrm{d}P \leq \varepsilon \quad \text{et} \quad \int_{\{|X_n-X|\geq\varepsilon\}} |X|\,\mathrm{d}P \leq \varepsilon\,.$$

Alors $E(|X_n - X|) \leq 3\varepsilon$. Puisque ε est arbitraire, X_n converge vers X dans L^1.
(ii)$\Rightarrow$(i). Soit $\varepsilon > 0$ et n_0 tel que $\|X_n - X\|_1 \leq \varepsilon$ pour $n > n_0$. Puisque $X \in \mathrm{L}^1$, la famille X, X_n, $n \leq n_0$, est uniformément intégrable. D'après la proposition V.3.4, il existe $\eta > 0$ tel que si $P(A) \leq \eta$,

$$\int_A |X|\,\mathrm{d}P \leq \varepsilon/2 \quad \text{et} \quad \int_A |X_n|\,\mathrm{d}P \leq \varepsilon/2$$

pour tout $n \leq n_0$. Lorsque $n > n_0$, par l'inégalité triangulaire,

$$\int_A |X_n|\,\mathrm{d}P \leq \int_A |X|\,\mathrm{d}P + \|X_n - X\|_1 \leq 2\varepsilon\,.$$

Il s'ensuit que la suite $(X_n)_{n\in\mathbb{N}}$ vérifie le point (i) de la proposition V.3.4. Le point (ii) est immédiat puisque par l'inégalité triangulaire $E(|X_n|)$ est majoré par $E(|X_n - X|) + E(|X|)$. Donc la suite $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable. □

Corollaire V.3.6. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles, définies sur un espace $(\Omega, \mathcal{A}, P)$, telle que pour un $p > 1$, $\sup_{n\in\mathbb{N}} E(|X_n|^p) < \infty$. Si $X_n \xrightarrow{P} X$, alors pour tout $q < p$, $\lim_{n\to\infty} \|X_n - X\|_q = 0$.*

Démonstration. Pour tout $c > 0$, et tout entier n,

$$\begin{aligned} \int_{\{|X_n|^q>c\}} |X_n|^q\,\mathrm{d}P &\leq c^{(q-p)/q} \int_{\{|X_n|^q>c\}} |X_n|^p\,\mathrm{d}P \\ &\leq c^{(q-p)/q} \sup_{k\in\mathbb{N}} E\big(|X_k|^p\big)\,. \end{aligned}$$

Comme $q < p$, le terme de droite tend vers 0 lorsque $c \to \infty$, uniformément en n. Donc la suite $(|X_n|^q)_{n\in\mathbb{N}}$ est uniformément intégrable. Comme $|X_n|^p \xrightarrow{P} |X|^p$, par le lemme de Fatou, $E(|X|^p) < \infty$, et donc aussi $E(|X|^q) < \infty$. Par conséquent, en utilisant l'inégalité triangulaire pour la norme $\|\cdot\|_p$, la suite $(|X_n - X|^q)_{n\in\mathbb{N}}$ est aussi uniformément intégrable. Si X_n converge en probabilité vers X, alors $|X_n - X|^q$ converge en probabilité vers 0. Le théorème V.3.5 montre alors que la suite $(|X_n - X|^q)_{n\in\mathbb{N}}$ converge vers 0 dans L^1, et donc que la suite $(X_n)_{n\in\mathbb{N}}$ converge vers X dans L^q. $\square$

Démonstration du théorème II.6.4. Nous montrons le résultat lorsque $\mu(\Omega) < \infty$. Le cas général s'en déduit en considérant une suite d'exhaustion $(A_n)_{n\in\mathbb{N}}$ de $(\Omega, \mathcal{A}, \mu)$ et en utilisant un argument diagonal. Quitte à remplacer μ par $\mu/\mu(\Omega)$, nous pouvons supposer que μ est une probabilité.

On remarque d'abord qu'une suite de Cauchy $(X_n)_{n\in\mathbb{N}}$ dans L^p est équiintégrable. Nous pouvons en extraire une sous-suite qui est p.s. de Cauchy, et donc converge p.s. vers une limite X. La limite est dans L^p puisque les X_n sont équiintégrables. La limite ne dépend pas de la sous-suite considérée puisque la suite est de Cauchy. Donc X_n converge dans L^p. $\square$

V.4. Convergence en loi

C'est le quatrième type de convergence que nous étudierons. C'est le plus faible, mais peut-être aussi le plus important. Il est souvent utilisé dans les applications.

Nous avons vu que deux variables aléatoires, X, Y, sur $(\Omega, \mathcal{A}, P)$ ont même loi, ou que les lois P^X et P^Y sont égales, si et seulement si leurs fonctions de répartition sont égales (III.2.3), *i.e.*

$$F^X = F^Y\,,$$

ou (*cf.* III.5) si pour toute fonction $\phi : \mathbb{R} \mapsto \mathbb{R}$ continue bornée

$$\int \phi(X)\,\mathrm{d}P = \int \phi(Y)\,\mathrm{d}P\,,$$

ou encore si leurs fonctions caractéristiques sont égales (III.5.2), *i.e.*

$$\varphi^X = \varphi^Y\,.$$

Ces diverses égalités donnent lieu à des définitions de convergence.

Définition et théorème V.4.1. *Soient X_n, $n \in \mathbb{N}$ et X, des variables aléatoires réelles, définies sur $(\Omega, \mathcal{A}, P)$. On dit que X_n converge en loi vers X, ou que les lois P^{X_n} convergent étroitement vers la loi P^X, si l'une des quatre conditions équivalentes suivantes est vérifiée :*

(i) $\lim_{n\to\infty} F^{X_n}(t) = F^X(t)$ *en tout point de continuité t de F^X ;*

(ii) $\lim_{n\to\infty} \int \phi(X_n)\,\mathrm{d}P = \int \phi(X)\,\mathrm{d}P$ *pour toute fonction continue bornée* $\phi : \mathbb{R} \to \mathbb{R}$ *;*

(iii) $\lim_{n\to\infty} \varphi^{X_n}(t) = \varphi^X(t)$ *pour tout $t \in \mathbb{R}$;*

(iv) Il existe un espace probabilisé $(\Omega', \mathcal{A}', P')$ sur lequel sont définies des variables aléatoires X'_n, $n \in \mathbb{N}$ et X', telles que X_n et X'_n ont même loi pour tout n, X et X' ont même loi, et $\lim_{n\to\infty} X'_n = X'$ *p.s.*

On note alors $X_n \xrightarrow{\mathcal{L}} X$ ou $X_n \xrightarrow{d} X$ (pour X_n converge « en distribution » vers X).

On prendra garde au fait que, dans la définition-théorème V.4.1.iv, la convergence presque sûre a lieu pour les nouvelles variables X'_n, X'. En particulier, nous verrons à l'exemple V.4.2.iv que la convergence en loi n'implique pas la convergence presque sûre. Dans le même esprit, si X'_n et X_n ont même loi pour tout n, il n'en est rien en général de (X'_n, X'_{n+1}) et (X_n, X_{n+1}) ou de tout autre vecteur formé à l'aide d'éléments des suites $(X'_n)_{n\in\mathbb{N}}$ et $(X_n)_{n\in\mathbb{N}}$.

Avant de démontrer l'équivalence entre les points (i)–(iv) de cette définition, montrons que la convergence faible est bien le mode le plus faible de convergence que nous avons introduit jusqu'à présent.

Exemples V.4.2. (i) Si X_n converge p.s. vers X, alors X_n converge en loi vers X. Cela se déduit par exemple du théorème de convergence dominée (II.2.8) et du point (ii) de la définition.

(ii) Si X_n converge en probabilité vers X, alors X_n converge en loi vers X. En effet, pour tout $\varepsilon > 0$, et tout t,

$$\begin{aligned} F^{X_n}(t) &= P\{\, X_n \leq t \,\} \\ &\leq P\{\, X \leq t + \varepsilon \,\} + P\{\, |X_n - X| \geq \varepsilon \,\} \\ &\leq F^X(t+\varepsilon) + o(1) \qquad (n \to \infty) \end{aligned}$$

et

$$\begin{aligned} F^{X_n}(t) &\geq P\{\, X \leq t - \varepsilon \,\} - P\{\, |X_n - X| \geq \varepsilon \,\} \\ &\geq F^X(t-\varepsilon) + o(1) \qquad (n \to \infty) \end{aligned}$$

et l'on conclut à l'aide de V.4.1.i puisque $\varepsilon > 0$ est arbitraire. Rappelons que la convergence dans L^p, $p > 0$, entraîne la convergence en probabilité, et donc la convergence en loi.

(iii) Soit X de loi $\mathcal{N}(0,1)$ et $X_n = (-1)^n X$. Alors X_n converge en loi vers X (puisque, par symétrie de la loi normale centrée, X_n a même loi que X pour tout n), mais ne converge pas p.s. vers X et ne converge pas en probabilité vers X. Le couple (X, X_n) ne converge pas non plus en loi.

(iv) Nous notons cependant le résultat suivant, utile en statistique. Si X_n converge en loi vers une variable constante c, alors X_n converge en probabilité vers c. En effet, pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{\, -\varepsilon < X_n - c \leq \varepsilon \,\} = \lim_{n\to\infty} \big(F^{X_n}(c+\varepsilon) - F^{X_n}(c-\varepsilon) \big) = 1\,,$$

et donc $\lim_{n\to\infty} P\{\, |X_n - c| > \varepsilon \,\} = 0$. Le même argument donne un petit résultat aussi utile en statistique. Si X_n et Y_n convergent en loi vers X et Y, on ne peut rien dire en général sur la convergence du couple (X_n, Y_n) (voir par exemple (iii) ci-dessus). Par contre, si Y_n converge en loi vers une constante c, alors le couple (X_n, Y_n) converge en loi vers le couple (X, c). En particulier, $X_n Y_n$ converge en loi vers cX et $X_n + Y_n$ vers $X + c$.

(v) Soient X_n, $n \in \mathbb{N}$, et X, des variables aléatoires à valeurs entières ; alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\lim_{n\to\infty} P\{\, X_n = k \,\} = P\{\, X_n = k \,\}$ pour tout $k \in \mathbb{N}$. Il suffit pour s'en convaincre de choisir, pour chaque $k \in \mathbb{N}$, des points s et t de continuité pour F^X tels que $k - 1 < s < k < t < k + 1$. Alors

$$P\{\, X_n = k \,\} = F^{X_n}(t) - F^{X_n}(s) \to F^X(t) - F^X(s) = P\{\, X = k \,\}\,.$$

De la même façon,

$$F^{X_n}(t) = F^{X_n}(k) = \sum_{0 \leq l \leq k} P\{\, X_n = l \,\} \to \sum_{0 \leq l \leq k} P\{\, X = l \,\} = F^X(k) = F^X(t)\,.$$

Démonstration du théorème V.4.1.. (iv)⇒(iii) et (iv)⇒(ii) sont immédiats par convergence dominée (II.2.8) et en remarquant que $\varphi^{X'_n} = \varphi^{X_n}$ et $\varphi^{X'} = \varphi^X$, respectivement $E(\phi(X'_n)) = E(\phi(X_n))$ et $E(\phi(X')) = E(\phi(X))$.
(ii)⇒(iii) est clair. Il suffit de poser $\phi(x) = \cos(tx)$, puis $\phi(x) = \sin(tx)$.
(iv)⇒(i). Notons $F^{X_n} = F_n$ et $F^X = F$ dans ce qui suit. Soit t un point de continuité de F. Ainsi, $P\{\, X' \neq t \,\} = 1$. Donc, presque sûrement, $\lim_{n\to\infty} \mathbb{1}_{]-\infty,t]}(X'_n) = \mathbb{1}_{]-\infty,t]}(X')$. En utilisant le théorème de convergence dominée II.2.8, on en déduit que $\lim_{n\to\infty} F_n(t) = F(t)$.

(i)⇒(iv). Soit l'espace probabilisé $(\Omega', \mathcal{A}', P') = \big(\,]0,1[\,, \mathcal{B}(\,]0,1[\,), \lambda\big)$ où λ est la mesure de Lebesgue. Soit U une variable aléatoire uniforme sur $]0,1[$, définie sur $(\Omega', \mathcal{A}', P')$ (par exemple la fonction identité !). Alors (proposition III.2.7), $X'_n = F_n^{\leftarrow}(U)$ et $X' = F^{\leftarrow}(U)$ ont respectivement mêmes lois que X_n et X. Il suffit donc de montrer que $\lim_{n\to\infty} F_n^{\leftarrow}(u) = F^{\leftarrow}(u)$, sauf peut-être sur un ensemble de mesure de Lebesgue nulle.

Pour cela, soit $u \in\,]0,1[$ et $t = F^{\leftarrow}(u)$. Soient de plus $\varepsilon > 0$ arbitraire et $t_\varepsilon^+, t_\varepsilon^-$ des points de continuité de F tels que $t_\varepsilon^- < t < t_\varepsilon^+$ et $|t_\varepsilon^+ - t_\varepsilon^-| \leq \varepsilon$ (de tels points existent d'après la propriété III.2.4). D'après les propriétés de $F^{\leftarrow}$ utilisées lors de la démonstration de la proposition III.2.7, $F(t_\varepsilon^-) \leq u$ et $F(t_\varepsilon^+) \geq u$. Ainsi, par l'hypothèse (i), pour tout $\eta > 0$ tel que $0 < u - \eta < u + \eta < 1$, $F_n(t_\varepsilon^-) < u + \eta$ et $F_n(t_\varepsilon^+) > u - \eta$ pour tout n assez grand. D'après les mêmes propriétés appliquées à $F_n^{\leftarrow}$,

$$F_n^{\leftarrow}(u+\eta) > t_\varepsilon^- \geq t - \varepsilon \qquad \text{et} \qquad F_n^{\leftarrow}(u-\eta) \leq t_\varepsilon^+ \leq t + \varepsilon\,.$$

Ainsi, puisque $\varepsilon > 0$ est arbitraire,

$$\liminf_{n\to\infty} F_n^{\leftarrow}(u+\eta) \geq t = F^{\leftarrow}(u) \qquad \text{et} \qquad \limsup_{n\to\infty} F_n^{\leftarrow}(u-\eta) \leq t = F^{\leftarrow}(u)\,.$$

En remplaçant u par $u - \eta$ dans la première limite et en faisant tendre η vers 0, il vient $\liminf_{n\to\infty} F_n^{\leftarrow}(u) \geq F^{\leftarrow}(u-)$. De la même façon, en remplaçant u par $u + \eta$ dans la seconde limite, il vient $\limsup_{n\to\infty} F_n^{\leftarrow}(u) \leq F^{\leftarrow}(u+) = F^{\leftarrow}(u)$ puisque $F^{\leftarrow}$ est continue à droite. Ainsi, pour tout $u \in\,]0,1[$,

$$F^{\leftarrow}(u-) \leq \liminf_{n\to\infty} F_n^{\leftarrow}(u) \leq \limsup_{n\to\infty} F_n^{\leftarrow}(u) \leq F^{\leftarrow}(u)\,.$$

Or les points de discontinuité de $F^{\leftarrow}$, en nombre dénombrable d'après la remarque suivant la proposition III.2.7, sont de mesure de Lebesgue nulle. C'est la démonstration.

(ii)⇒(i) : Observons que pour tous $\varepsilon > 0$, $t \in \mathbb{R}$, $x \in \mathbb{R}$,

$$\mathbb{1}_{]-\infty,t-\varepsilon]}(x) \leq \frac{(t-x)^+}{\varepsilon} \wedge 1 \leq \mathbb{1}_{]-\infty,t]}(x) \leq \frac{(t+\varepsilon-x)^+}{\varepsilon} \wedge 1 \leq \mathbb{1}_{]-\infty,t+\varepsilon]}(x)\,.$$

Donc, si (ii) a lieu, nous en déduisons que

$$\begin{aligned}
F(t-\varepsilon) = E\big(\mathbb{1}_{]-\infty,t-\varepsilon]}(X)\big) &\leq E\bigg(\frac{(t-X)^+}{\varepsilon} \wedge 1\bigg) \\
&= \liminf_{n\to\infty} E\bigg(\frac{(t-X_n)^+}{\varepsilon} \wedge 1\bigg) \\
&\leq \liminf_{n\to\infty} E\big(\mathbb{1}_{]-\infty,t]}(X_n)\big) \\
&= \liminf_{n\to\infty} F_n(t)\,.
\end{aligned}$$

De la même façon,

$$\begin{aligned}
F(t+\varepsilon) = E\big(\mathbb{1}_{]-\infty,t+\varepsilon]}(X)\big) &\geq E\bigg(\frac{(t+\varepsilon-X)^+}{\varepsilon}\wedge 1\bigg)\\
&\geq \limsup_{n\to\infty} E\bigg(\frac{(t+\varepsilon-X_n)^+}{\varepsilon}\wedge 1\bigg)\\
&\geq \limsup_{n\to\infty} E\big(\mathbb{1}_{]-\infty,t]}(X_n)\big)\\
&= \limsup_{n\to\infty} F_n(t)\,,
\end{aligned}$$

ce qui implique (i) puisque ε est arbitraire.

Il ne nous reste plus qu'à montrer (iii)⇒(i) par exemple. L'idée de la démonstration est d'utiliser la formule d'inversion de Fourier III.5.4, et d'intégrer dans la convergence. Pour cela il est nécessaire que X_n admette une densité et que l'on puisse passer à la limite en n dans les inversions des fonctions caractéristiques (ce qui impose a priori que les φ^{X_n} soient équiintégrables). L'astuce pour obtenir (iii)⇒(i) sans autre hypothèse consiste à multiplier φ^{X_n} par une fonction caractéristique intégrable, rendant ainsi la suite équiintégrable, ce qui revient à ajouter à X_n un petit « bruit ». Il est commode de prendre ce bruit gaussien. On notera donc N une variable aléatoire de loi $\mathcal{N}(0,1)$. Si Z est une variable aléatoire, observons que pour tout $t\in\mathbb{R}$ et tout $\varepsilon>0$,

$$\begin{aligned}
F^Z(t) = P\{\,Z\leq t\,\} &\leq P\big\{\,Z\leq t\;;\;\varepsilon^2\,|N|\leq\varepsilon\,\big\} + P\big\{\,\varepsilon|N|\geq 1\,\big\}\\
&\leq P\big\{\,Z+\varepsilon^2N\leq t+\varepsilon\,\big\} + P\big\{\,\varepsilon|N|\geq 1\,\big\} \qquad (1)\\
&\leq F^{Z+\varepsilon^2N}(t+\varepsilon) + \varepsilon E\big(|N|\big)
\end{aligned}$$

où nous avons utilisé l'inégalité de Markov (III.4.9). De même,

$$\begin{aligned}
F^Z(t) &\geq P\big\{\,Z\leq t\,;\,\varepsilon^2|N|\leq\varepsilon\,\big\}\\
&\geq P\big\{\,Z+\varepsilon^2N\leq t-\varepsilon\,;\,\varepsilon^2|N|\leq\varepsilon\,\big\} \qquad (2)\\
&\geq F^{Z+\varepsilon^2N}(t-\varepsilon) - \varepsilon E\big(|N|\big)\,.
\end{aligned}$$

Soient donc, pour chaque n, X_n de fonction de répartition F^{X_n} et X de fonction de répartition F^X, et prenons N indépendante de X_n et X. Alors $\varphi^{X_n+\varepsilon^2N}(t) = \varphi^{X_n}(t)e^{-\varepsilon^4t^2/2}$ (proposition IV.2.3), et de même pour X. D'après le théorème d'inversion de Fourier III.5.4, $X_n+\varepsilon^2N$ et $X+\varepsilon^2N$ admettent des densités, $f^{X_n+\varepsilon^2N}$ et $f^{X+\varepsilon^2N}$, telles que, pour tout $x\in\mathbb{R}$,

$$f^{X_n+\varepsilon^2N}(x) - f^{X+\varepsilon^2N}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\Big(\varphi^{X_n}(t)-\varphi^X(t)\Big)e^{-\varepsilon^4t^2/2}\,\mathrm{d}t\,.$$

Puisque φ_n converge vers φ simplement et que $\left|e^{-itx}\big(\varphi^{X_n}(t)-\varphi^X(t)\big)\right|\le 2$, le théorème de convergence dominée (II.2.8) implique que

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\left|f^{X_n+\varepsilon^2N}(x)-f^{X+\varepsilon^2N}(x)\right|=0\,. \tag{3}$$

Soit a assez grand tel que $P\{\,|X+\varepsilon^2N|>a\,\}\le\varepsilon$. En intégrant sur le compact $[\,-a,a\,]$, nous déduisons de (3) que

$$\begin{aligned}\limsup_{n\to\infty}P\big\{\,|X_n+\varepsilon^2N|\ge a\,\big\}&=1-\liminf_{n\to\infty}P\big\{\,X_n+\varepsilon^2N\in[\,-a,a\,]\,\big\}\\&=1-\liminf_{n\to\infty}\int_{-a}^{a}f^{X_n+\varepsilon^2N}(x)\,\mathrm{d}x\\&=1-\int_{-a}^{a}f^{X+\varepsilon^2N}(x)\,\mathrm{d}x\\&=1-P\big\{\,X+\varepsilon^2N\in[\,-a,a\,]\,\big\}\le\varepsilon\,.\end{aligned}\tag{4}$$

En faisant usage de (1), il vient pour tout $t\in\mathbb{R}$,

$$\begin{aligned}\limsup_{n\to\infty}F^{X_n}(t)&\le\limsup_{n\to\infty}F^{X_n+\varepsilon^2N}(t+\varepsilon)+\varepsilon E\big(|N|\big)\\&\le\limsup_{n\to\infty}P\big\{\,X_n+\varepsilon^2N\in[\,-a,t+\varepsilon\,]\,\big\}\\&\qquad+\limsup_{n\to\infty}P\big\{\,|X_n+\varepsilon^2N|>a\,\big\}+\varepsilon E\big(|N|\big)\,.\end{aligned}$$

D'après la convergence uniforme des densités (3),

$$\begin{aligned}\limsup_{n\to\infty}P\big\{\,X_n+\varepsilon^2N\in[\,-a,t+\varepsilon\,]\,\big\}&=P\big\{\,X+\varepsilon^2N\in[\,-a,t+\varepsilon\,]\,\big\}\\&\le F^{X+\varepsilon^2N}(t+\varepsilon)\,.\end{aligned}$$

Ainsi, par (2) et (4), on conclut que

$$\limsup_{n\to\infty}F^{X_n}(t)\le F^X(t+2\varepsilon)+\varepsilon+2\varepsilon E\big(|N|\big)\,.$$

De façon analogue,

$$\begin{aligned}\liminf_{n\to\infty}F^{X_n}(t)&\ge\liminf_{n\to\infty}F^{X_n+\varepsilon^2N}(t-\varepsilon)-\varepsilon E\big(|N|\big)\\&\ge\liminf_{n\to\infty}P\big\{\,X_n+\varepsilon^2N\in[\,-a,t-\varepsilon\,]\,\big\}-\varepsilon-\varepsilon E\big(|N|\big)\\&\ge F_X(t-2\varepsilon)-\varepsilon-2\varepsilon E\big(|N|\big)\,.\end{aligned}$$

Puisque $\varepsilon > 0$ est arbitraire et F est continue à droite et admet en tout point une limite à gauche, nous obtenons,

$$F^X(t-) \leq \liminf_{n\to\infty} F^{X_n}(t) \leq \limsup_{n\to\infty} F^{X_n}(t) \leq F(t) \, ,$$

ce qui est (i). La démonstration du théorème est terminée. □

La démonstration du théorème V.4.1 se généralise à $\mathbb{R}^d$, et seule notre démonstration de (i)⇒(iv) doit être substantiellement modifiée. On pourra se référer par exemple au livre de Pollard (1984) pour une démonstration de l'équivalence entre convergence en loi et convergence presque sûre d'une version bien choisie des vecteurs aléatoires.

Notre démonstration de (iii)⇒(i) aurait pu être remplacée par une preuve de l'implication (iii)⇒(ii) laquelle peut être obtenue en introduisant un point de vue plus proche de l'analyse fonctionnelle, important dans certains aspects théoriques des probabilités et statistiques.

De façon générale, lorsque l'on dispose d'une suite $(z_n)_{n\in\mathbb{N}}$ dans un espace topologique, sa convergence peut être démontrée en deux étapes. On commence par montrer qu'elle est relativement compacte, c'est-à-dire que l'ensemble $\{\, z_n \; : \; n \in \mathbb{N} \,\}$ est d'adhérence compacte ; puis on montre que l'ensemble des valeurs d'adhérence contient un unique point, que l'on identifie éventuellement. Par exemple, si $(z_n)_{n\in\mathbb{N}}$ est une suite dans $(\mathrm{C}[\,0,1\,], \|\cdot\|_\infty)$, le théorème d'Arzela-Ascoli donne un critère de compacité relative de la suite $(z_n)_{n\in\mathbb{N}}$ (il faut et il suffit qu'elle soit bornée et équicontinue) et on peut identifier la limite éventuelle en démontrant une convergence ponctuelle de z_n. Dans l'espace L^1, le théorème V.3.5 est l'exact analogue du théorème d'Arzela-Ascoli. Une suite $(z_n)_{n\in\mathbb{N}}$ dans L^1 est relativement compacte si elle est bornée et équiintégrable, et on peut identifier sa limite en démontrant alors seulement une convergence en probabilité. La notion d'équiintégrabilité est donc aussi une notion de compacité relative. Dans un espace de mesures, la notion correspondante est celle d'équitension, laquelle assure que la masse des mesures ne s'échappe pas à l'infini. Elle est définie comme suit.

Définition V.4.3. Soit Ω un espace topologique muni de sa tribu borélienne et d'une probabilité P. On dit que P est tendue si pour tout $\varepsilon > 0$, il existe un compact K tel que $P(K) \geq 1 - \varepsilon$.

Si $(P_n)_{n\in\mathbb{N}}$ est une suite de probabilités, on dit que la suite est équitendue (ou parfois tendue, ou uniformément tendue) si, pour tout $\varepsilon > 0$, il existe un compact K, tel que $P_n(K) \geq 1 - \varepsilon$ pour tout n.

Si $(X_n)_{n\in\mathbb{N}}$ sont des variables de lois $(P_n)_{n\in\mathbb{N}}$ équitendues, on dit aussi que $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires équitendues (ou tendues, ou uniformément tendues).

Sur $\mathbb{R}^d$, toute loi de probabilité est tendue puisque $\mathbb{R}^d$ est la réunion des compacts $K_M = \{ x \in \mathbb{R}^d : |x| \leq M \}$ pour $M = 1, 2, \ldots$ et qu'une mesure vérifie la proposition I.4.3.iii.

Théorème V.4.4. *Toute suite de lois équitendue sur $\mathbb{R}^d$ admet une sous-suite convergeant étroitement.*

Démonstration. (Esquissée ; pour plus de détails, voir Rudin (1975), Dudley (1989)). La preuve s'appuie sur un argument de dualité. Le dual de l'espace des fonctions continues bornées sur $\mathbb{R}^d$ contenant des objets qui ne sont pas des mesures, il nous faut travailler avec l'espace $\mathrm{C}_0(\mathbb{R}^d)$ des fonctions continues sur $\mathbb{R}^d$ tendant vers 0 à l'infini. Muni de la norme supremum $\|\cdot\|_\infty$, c'est un espace espace de Banach séparable. Son dual s'identifie avec l'espace $\mathcal{M}(\mathbb{R}^d)$ des mesures signées bornées $\mu = \mu^+ - \mu^-$, où μ^+ et μ^- sont des mesures positives et bornées, à l'aide de la dualité

$$\langle \mu, \phi \rangle = \int \phi \, \mathrm{d}\mu = \int \phi \, \mathrm{d}\mu^+ - \int \phi \, \mathrm{d}\mu^- \, .$$

On peut munir $\mathcal{M}(\mathbb{R}^d)$ de la topologie faible* dont une base de voisinages est donnée par les ensembles

$$\Big\{ \nu \in \mathcal{M}(\mathbb{R}^d) \ : \ \Big| \int \phi_i \, \mathrm{d}\nu - \int \phi_i \, \mathrm{d}\mu \Big| \leq \varepsilon \, , \, 1 \leq i \leq k \Big\}$$

pour $\mu \in \mathcal{M}(\mathbb{R}^d)$, $\phi_i \in \mathrm{C}_0(\mathbb{R}^d)$, $1 \leq i \leq k$, $\varepsilon > 0$. En particulier, une suite $(\mu_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{M}(\mathbb{R}^d)$ converge faiblement* vers μ si pour toute $\phi \in \mathrm{C}_0(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \int \phi \, \mathrm{d}\mu_n = \int \phi \, \mathrm{d}\mu \, .$$

On munit l'espace $\mathcal{M}(\mathbb{R}^d)$ de la norme duale

$$\|\mu\| = \sup \Big\{ \int \phi \, \mathrm{d}\mu \ : \ \|\phi\|_\infty \leq 1 \Big\} \, .$$

Le théorème de Banach-Alaoglu, conséquence du théorème de Tyckhonov, indique que la boule unité de $\mathcal{M}(\mathbb{R}^d)$, c'est-à-dire $\{ \mu \in \mathcal{M}(\mathbb{R}^d) \ : \ \|\mu\| \leq 1 \}$, est compacte et métrisable pour la topologie faible*. (Cette structure métrique est importante car elle permet l'extraction de sous-suites convergentes.)

Soit donc $(P_n)_{n\in\mathbb{N}}$ une suite de lois équitendue sur $\mathbb{R}^d$. Elle est dans la boule unité de $\mathcal{M}(\mathbb{R}^d)$. C'est donc une suite relativement compacte dans $\mathcal{M}(\mathbb{R}^d)$ et elle admet une sous-suite $(P_{n'})_{n'\in\mathbb{N}}$ convergeant vers une mesure Q, a priori signée mais de norme inférieure ou égale à 1. Cette valeur d'adhérence Q de la suite $(P_n)_{n\in\mathbb{N}}$ est nécessairement une mesure positive. En effet, sinon on montre l'existence d'une fonction positive ϕ de $\mathrm{C}_0(\mathbb{R}^d)$ telle que $\int \phi \,\mathrm{d}Q < 0$, et alors $\int \phi \,\mathrm{d}P_{n'} < 0$ pour n' assez grand, ce qui contredit la positivité des lois P_n. Il ne nous reste plus qu'à montrer que Q est une probabilité et que $P_{n'}$ converge étroitement vers Q. Il suffit de démontrer à cet effet que pour toute $\phi : \mathbb{R}^d \to \mathbb{R}$ continue telle que $0 \leq \phi \leq 1$,

$$\lim_{n'\to\infty} \int \phi \,\mathrm{d}P_{n'} = \int \phi \,\mathrm{d}Q\,.$$

Ceci est une conséquence de l'équitension de la suite $(P_n)_{n\in\mathbb{N}}$. Pour tout $M > 0$, soit K_M le compact de $\mathbb{R}^d$ donné par $K_M = \{\, x \in \mathbb{R}^d : \|x\| \leq M \,\}$. Pour tout $\varepsilon > 0$, il existe M tel que $P_n(K_M) \geq 1 - \varepsilon$ pour tout n. Soit $\psi_M(x) = (M + 1 - \|x\|)^+ \wedge 1$, $x \in \mathbb{R}^d$. Pour tout n,

$$\int \phi \,\mathrm{d}P_n \leq \int_{K_M} \phi \,\mathrm{d}P_n + \varepsilon \leq \int \psi_M \phi \,\mathrm{d}P_n + \varepsilon$$

et donc, puisque $\psi_M \phi \in \mathrm{C}_0(\mathbb{R}^d)$,

$$\limsup_{n'\to\infty} \int \phi \,\mathrm{d}P_{n'} \leq \int \psi_M \phi \,\mathrm{d}Q + \varepsilon \leq \int \phi \,\mathrm{d}Q + \varepsilon\,.$$

En particulier, si ϕ est constante égale à 1, $Q(\mathbb{R}^d) \geq 1 - \varepsilon$, et donc $Q(\mathbb{R}^d) = 1$ puisque Q est dans la boule unité de $\mathcal{M}(\mathbb{R}^d)$. En particulier, l'on peut également choisir M tel que $Q(K_M) \geq 1 - \varepsilon$. Ainsi

$$\int \phi \,\mathrm{d}Q \leq \int_{K_M} \phi \,\mathrm{d}Q + \varepsilon \leq \int \psi_M \phi \,\mathrm{d}Q + \varepsilon$$

et donc, comme précédemment,

$$\int \phi \,\mathrm{d}Q - \varepsilon \leq \liminf_{n'\to\infty} \int \psi_M \phi \,\mathrm{d}P_{n'} \leq \liminf_{n'\to\infty} \int \phi \,\mathrm{d}P_{n'}\,.$$

La conclusion s'ensuit, et le théorème V.4.4 est établi de cette façon. □

Nous pouvons maintenant donner une démonstration directe de l'implication (iii)⇒(ii) dans le théorème V.4.1. Montrons que la suite $(X_n)_{n\in\mathbb{N}}$ (de variables aléatoires réelles) est uniformément tendue. Il va suffire de montrer que pour toute variable aléatoire réelle Z et tout $u > 0$,

$$P\{\, |Z| \geq 1/u \,\} \leq \frac{7}{u} \int_0^u \big(1 - \Re e\, \varphi^Z(t)\big) \,\mathrm{d}t\,.$$

Pour cela, notons d'abord que l'inégalité $(\sin x)/x \leq \sin 1$ est vraie pour tout $|x| \geq 1$. Puis, par le théorème de Fubini-Tonelli (II.5.1),

$$\begin{aligned}
\frac{1}{u}\int_0^u \big(1 - \Re e\,\varphi^Z(t)\big)\,\mathrm{d}t &= \frac{1}{u}\int_0^u \int_{\mathbb{R}} \big(1 - \cos(tx)\big)\,\mathrm{d}P^Z(x)\,\mathrm{d}t \\
&= \int_{\mathbb{R}} \left(1 - \frac{\sin(ux)}{ux}\right)\mathrm{d}P^Z(x) \\
&\geq \big(1 - \sin(1)\big)P\{\,|Z| \geq 1/u\,\}\,.
\end{aligned}$$

Enfin, $\sin(1) \leq 6/7$.

En conséquence de cette inégalité appliquée à X_n pour tout n,

$$\begin{aligned}
\limsup_{n\to\infty} P\{\,|X_n| \geq M\,\} &\leq \limsup_{n\to\infty} 7M\int_0^{1/M} \big(1 - \Re e\,\varphi^{X_n}(t)\big)\,\mathrm{d}t \\
&= 7M\int_0^{1/M} \big(1 - \Re e\,\varphi^X(t)\big)\,\mathrm{d}t
\end{aligned}$$

et le majorant peut être rendu arbitrairement petit en prenant M arbitrairement grand puisque $\lim_{t\to 0} \Re e\,\varphi^X(t) = 1$ et φ^X est continue.

Ainsi, d'après le théorème V.4.4, il existe une sous-suite $(P_{n'})_{n'\in\mathbb{N}}$ de la suite $(P_n)_{n\in\mathbb{N}}$ des lois des variables aléatoires $(X_n)_{n\in\mathbb{N}}$ qui converge étroitement vers une mesure de probabilité Q. Cette mesure Q est la loi de X puisque $\varphi^{X_{n'}}$ converge vers φ^X et que φ^X détermine la loi de X (théorème III.5.2). La conclusion s'ensuit.

Exemples V.4.5. (i) Une suite $(x_n)_{n\in\mathbb{N}}$ de réels converge vers x si et seulement si δ_{x_n} converge vers δ_x étroitement (d'après V.4.1.ii).

(ii) Si X est une variable aléatoire, alors X/n converge en loi vers 0 lorsque $n \to \infty$. Par contre, sauf si $X = 0$ p.s., la suite des lois de $(nX)_{n\in\mathbb{N}}$ n'est pas tendue. Donc nX ne peut pas converger en loi.

(iii) Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes, de loi exponentielle $\mathcal{E}xp(1)$. Nous avons vu à l'exemple V.1.3.ii que

$$M_n = \max_{1\leq i\leq n} X_i = \ln n + o(\ln n) \text{ p.s.}$$

Nous pouvons maintenant préciser un peu le contenu du terme $o(\ln n)$. En effet, montrons que $Z_n = M_n - \ln n$ converge en loi vers une variable Z de loi $F^Z(t) = \exp(-e^{-t})$, $t \in \mathbb{R}$. En effet (voir plus précisément la démonstration du théorème V.5.4),

$$\begin{aligned} F^{Z_n}(t) &= P\{\max_{1\leq i\leq n} X_i \leq t + \ln n\} \\ &= \bigl(1-\exp(-t-\ln n)\bigr)^n \\ &= \exp\left(n\ln\left(1-\frac{e^{-t}}{n}\right)\right) = \exp(-e^{-t}) + o(1) \end{aligned}$$

lorsque $n \to \infty$.

(iv) Si X_n est de loi binomiale $\mathcal{B}(n,p)$ et $Z_n = (X_n - np)/\sqrt{np(1-p)}$, alors Z_n converge en loi vers une variable aléatoire gaussienne $\mathcal{N}(0,1)$. En effet,

$$\begin{aligned} \varphi^{Z_n}(t) &= e^{-it\sqrt{\frac{np}{1-p}}}\Bigl(1-p+pe^{it/\sqrt{np(1-p)}}\Bigr)^n \\ &= e^{-it\sqrt{\frac{np}{1-p}}}\left(1+\frac{itp}{\sqrt{np(1-p)}} - \frac{t^2}{2n(1-p)} + o(n^{-1})\right)^n \\ &= e^{-t^2/2} + o(1) \qquad (n\to\infty) \end{aligned}$$

et donc $\varphi^{Z_n}(t)$ converge vers la fonction caractéristique d'une loi $\mathcal{N}(0,1)$. D'un point de vue pratique, cela signifie que pour n assez grand, on peut approximer $P\{Z_n \leq t\}$ par $(2\pi)^{-1/2}\int_{-\infty}^t e^{-x^2/2}\,\mathrm{d}x$ (on applique ici l'équivalence entre V.4.1.iii et V.4.1.i.) C'est très utile en pratique puisque le calcul de $P\{Z_n \leq t\}$ nécessite d'évaluer une somme pondérée de coefficients binomiaux (numériquement difficile pour n grand), tandis que l'approximation gaussienne ne nécessite que l'évaluation d'une intégrale. Cet exemple est un cas particulier du théorème limite central que nous allons voir dans le paragraphe suivant.

V.5. Les lois faible et forte des grands nombres, le théorème limite central

Dans toute cette section, nous désignons par $(X_i)_{i\geq 1}$ une suite de variables aléatoires réelles, indépendantes et de même loi qu'une variable X (X est une simple notation pour décrire commodément la loi commune des X_i). Pour tout $n \geq 1$, on pose $S_n = \sum_{1\leq i\leq n} X_i$ et l'on s'intéresse aux propriétés asymptotiques de S_n.

Observons que S_n/n est simplement la moyenne (dite empirique) des X_i. On conçoit alors que l'étude des sommes de variables aléatoires soit importante pour les applications. Par exemple si X_i modélise le fait qu'un individu vote ($X_i = 1$) ou ne vote pas ($X_i = 0$) pour un candidat dans une élection, S_n/n est la proportion de personnes votant pour le candidat dans un sondage de n personnes tirées au hasard dans la population. Nous avons vu dans ce cas particulier que S_n/n converge en probabilité vers l'espérance $E(X) = P\{ X = 1 \}$ (exemple V.2.2.i) et avons de plus montré que la loi de $(S_n - E(S_n))/\sqrt{n}$ ressemble à une loi normale lorsque n est assez grand (exemple V.4.5.iv).

De façon générale, une population est souvent décrite statistiquement, ou résumée par la donnée d'une statistique moyenne (âge moyen, poids moyen, taille moyenne, prix moyen etc). Un sondage est donc un exemple typique où l'on est naturellement conduit à des sommes de variables aléatoires.

Notre premier résultat est la loi des grands nombres. Elle montre d'une part que l'intuition est correcte, c'est-à-dire, par exemple, que si l'on jette une pièce équilibrée un grand nombre de fois, la proportion de piles tend à se stabiliser vers $1/2$, d'autre part que la théorie des probabilités que nous avons construite ne conduit pas à une modélisation absurde du réel, et enfin qu'une certaine régularité apparaît dans les phénomènes aléatoires.

***Théorème V.5.1* (loi faible des grands nombres).** *Si $E(|X|) < \infty$, alors S_n/n converge en probabilité vers $E(X)$ lorsque $n \to \infty$.*

Démonstration. Quitte à centrer les variables X_i, on peut supposer que $E(X_i) = 0$. Puisque $X \in \mathrm{L}^1$, la fonction caractéristique φ^X est dérivable (proposition III.5.6.i) et de plus $(\varphi^X)'(0) = iE(X) = 0$. La formule de Taylor donne $\varphi^X(t) = 1 + o(t)$. Donc $\varphi^{S_n/n}(t) = (\varphi^X(t/n))^n = (1 + o(n^{-1}))^n = 1 + o(1)$. Or 1 est la fonction caractéristique de δ_0. Donc S_n/n converge en loi vers la constante 0, donc en probabilité vers 0 (voir exemple V.4.2.iv). □

En fait, cette loi des grands nombres peut être considérablement renforcée par le résultat suivant qui nécessite exactement les mêmes hypothèses.

***Théorème V.5.2* (loi forte des grands nombres).** *Les deux conditions suivantes sont équivalentes :*

(i) $E(|X|) < \infty$;

(ii) $\lim_{n\to\infty} S_n/n = E(X)$ *p.s.*

Démonstration. (ii)⇒(i). Si la suite S_n/n converge p.s., alors X_n/n converge p.s. vers 0. D'après le lemme de Borel-Cantelli (proposition V.1.2.ii) et puisque les X_i sont indépendantes et toutes de même loi (que X), pour tout (ou seulement un) $\varepsilon > 0$,

$$\sum_{n\geq 1} P\{ |X| \geq \varepsilon n \} = \sum_{n\geq 1} P\{ |X_n| \geq \varepsilon n \} < \infty .$$

On conclut à l'aide de la proposition III.4.8.

(i)⇒(ii). La démonstration consiste à prouver dans un premier temps le résultat sous l'hypothèse plus forte que $E(|X|^4) < \infty$ et $E(X) = 0$. Dans ce cas, dont on peut se contenter en première lecture, $P\{\,|S_n/n| \geq \varepsilon\,\}$ peut être majoré en utilisant l'inégalité de Markov. La borne ainsi obtenue est le terme général d'une série convergente, ce qui permet de conclure grâce au lemme de Borel-Cantelli. Sous l'hypothèse plus faible du théorème, on approxime toute variable de L^1 par des variables de L^4 (ou même des variables bornées), puis on se ramène au cas traité.

Comme dans le théorème V.5.1, on peut remplacer X_i par $X_i - E(X_i)$ et supposer les variables aléatoires centrées. Commençons donc par montrer le résultat lorsque $E(|X|^4) < \infty$ et $E(X) = 0$. Dans ce cas, l'inégalité de Markov (III.4.10.i) montre que pour tout $n \geq 1$ et tout $\delta > 0$,

$$P\big\{\,|S_n| \geq \delta n\,\big\} \leq \frac{1}{\delta^4 n^4}\,E\big(S_n^4\big)\,.$$

Observons que

$$\begin{aligned} S_n^4 = \sum_{1\leq i\leq n} X_i^4 + 4 \sum_{1\leq i\neq j\leq n} X_i^3 X_j + 3 \sum_{1\leq i\neq j\leq n} X_i^2 X_j^2 \\ + 6 \sum_{\substack{1\leq i,j,k \\ \text{distincts}\leq n}} X_i X_j X_k^2 + \sum_{\substack{1\leq i,j,k,l \\ \text{distincts}\leq n}} X_i X_j X_k X_l\,. \end{aligned}$$

Donc, par linéarité de l'espérance, indépendance et centrage des X_i,

$$\begin{aligned} E(S_n^4) &= \sum_{1\leq i\leq n} E(X_i^4) + 4 \sum_{1\leq i\neq j\leq n} E(X_i^3)E(X_j) + 3 \sum_{1\leq i\neq j\leq n} E(X_i^2)E(X_j^2) \\ &\quad + 6 \sum_{\substack{1\leq i,j,k \\ \text{distincts}\leq n}} E(X_i)E(X_j)E(X_k^2) + \sum_{\substack{1\leq i,j,k,l \\ \text{distincts}\leq n}} E(X_i)E(X_j)E(X_k)E(X_l) \\ &= nE\big(X^4\big) + 3n(n-1)\big(E(X^2)\big)^2\,. \end{aligned}$$

Donc $\sum_{n\geq 1} P\{\,|S_n| > \delta n\,\} < \infty$, ce qui démontre la loi forte des grands nombres dans ce cas d'après le lemme de Borel-Cantelli V.1.2.

Supposons maintenant X intégrable et centrée, sans autre hypothèse. Soit $\varepsilon > 0$ fixé. Il existe, pour tout $i \geq 1$, des variables Y_i étagées, centrées, indépendantes et de même loi, telles que $E(|X_i - Y_i|) \leq \varepsilon$. Si $T_n = \sum_{1\leq i\leq n} Y_i$, nous avons

$$\frac{1}{n}\,|S_n| \leq \frac{1}{n} \sum_{1\leq i\leq n} |X_i - Y_i| + \frac{1}{n}\,|T_n|\,. \tag{1}$$

Puisque T_n/n converge p.s. vers 0 d'après le point précédent, il suffit de montrer que

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{1\leq i\leq n} |X_i - Y_i|$$

peut être rendu arbitrairement petit en prenant ε arbitrairement petit.

Notons Z_i, $i \geq 1$, des variables aléatoires positives, indépendantes et de même loi qu'une variable Z intégrable. On considèrera $Z_i = |X_i - Y_i|$ pour conclure la démonstration. Nous voulons borner

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{1\leq i\leq n} Z_i \,.$$

Pour cela, nous utilisons un argument de bloc. Pour tout k et tout $\delta > 0$, la probabilité

$$P\Big\{ \max_{2^k<n\leq 2^{k+1}} \frac{1}{n} \sum_{1\leq i\leq n} Z_i \geq 2E(Z) + \delta \Big\}$$

est majorée par

$$\begin{aligned} &P\Big\{ \exists i \in \{\,1,2,\ldots,2^{k+1}\,\} \,:\, Z_i > 2^k \Big\} \\ &\qquad\qquad + P\Big\{ \max_{2^k<n\leq 2^{k+1}} \frac{1}{n} \sum_{1\leq i\leq n} Z_i \mathbb{1}_{[0,2^k]}(Z_i) \geq 2E(Z) + \delta \Big\}\,, \end{aligned}$$

et donc aussi par

$$\begin{aligned} &2^{k+1}P\{\,Z > 2^k\,\} + P\Big\{ \sum_{1\leq i\leq 2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) \geq 2^{k+1}E(Z) + \delta 2^k \Big\} \\ &\quad \leq 2^{k+1}P\{\,Z > 2^k\,\} + P\Big\{ \sum_{1\leq i\leq 2^{k+1}} Z_i \mathbb{1}_{[0,2^k]}(Z_i) - E\big(Z_i \mathbb{1}_{[0,2^k]}(Z_i)\big) \geq \delta 2^k \Big\}\,. \end{aligned}$$

En appliquant maintenant l'inégalité de Tchebitchev (III.4.10.ii) au deuxième terme du majorant précédent, il vient

$$\begin{aligned} &P\Big\{ \max_{2^k<n\leq 2^{k+1}} \frac{1}{n} \sum_{1\leq i\leq n} Z_i \geq 2E(Z) + \delta \Big\} \\ &\qquad\qquad \leq 2^{k+1}P\{\,Z > 2^k\,\} + \frac{1}{\delta^2 2^{2k}}\, 2^{k+1} E\big(Z^2 \mathbb{1}_{[0,2^k]}(Z)\big) \\ &\qquad\qquad \leq 2^{k+1}P\{\,Z > 2^k\,\} + \frac{2}{\delta^2 2^k}\, E\big(Z^2 \mathbb{1}_{[0,2^k]}(Z)\big)\,. \end{aligned}$$

Puisque pour tout $k \in \mathbb{N}$,

$$\int_{2^k}^{2^{k+1}} P\{\, Z > t \,\}\, \mathrm{d}t \geq 2^k P\{\, Z > 2^{k+1} \,\}\,,$$

la démonstration de la proposition III.4.8 implique

$$\sum_{k \geq 0} 2^{k+1} P\{\, Z > 2^k \,\} \leq 4E(Z)\,.$$

De plus

$$\sum_{k \geq 0} 2^{-k} E\big(Z^2 \mathbb{1}_{[0,2^k]}(Z)\big) = E\Big(Z^2 \sum_{k \geq 0} 2^{-k} \mathbb{1}_{[0,2^k]}(Z)\Big) \leq 4E(Z)$$

puisque si $2^l < Z \leq 2^{l+1}$ pour un $l \geq 0$, alors

$$Z^2 \sum_{k \geq 0} 2^{-k} \mathbb{1}_{[0,2^k]}(Z) \leq 2^{2l+2} \sum_{k \geq l+1} 2^{-k} \leq 4Z$$

(et de même si $0 \leq Z \leq 1$). Donc, finalement,

$$\sum_{k \geq 0} P\Big\{ \max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \geq 2E(Z) + \delta \Big\} \leq 4(1 + 2\delta^{-2})E(Z)\,.$$

D'après le lemme de Borel-Cantelli (proposition IV.3.5.i), presque sûrement pour tout k assez grand

$$\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i < 2E(Z) + \delta\,. \tag{2}$$

Puisque $\delta > 0$ est arbitraire,

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i \leq 2E(Z) \text{ p.s.}$$

Nous pouvons maintenant finir la démonstration en choisissant $Z_i = |X_i - Y_i|$. De (1) et (2) nous déduisons en effet

$$\begin{aligned}\limsup_{n \to \infty} \frac{1}{n} |S_n| &\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{1 \leq i \leq n} Z_i + \limsup_{n \to \infty} \frac{1}{n} T_n \\ &\leq 2E\big(|X_1 - Y_1|\big) \leq 2\varepsilon \ \text{ p.s.}\end{aligned}$$

Puisque ε est arbitraire, ceci conclut la démonstration. $\square$

À noter qu'en travaillant avec une sous-suite $(\lfloor \rho^k \rfloor)_{k\in\mathbb{N}}$ pour un $\rho > 1$ bien choisi en fonction de $\delta > 0$ en lieu et place de la sous-suite $(2^k)_{k\in\mathbb{N}}$, on démontre directement que $\limsup_{n\to\infty} n^{-1} \sum_{1\leq i\leq n} Z_i \leq E(Z)$ p.s., et du même coup le théorème.

***Exemple* V.5.3.** Reprenons les notations de l'exemple IV.3.6.iii. Puisque les variables U_i sont indépendantes et de loi de Bernoulli $\mathcal{B}(1, 1/2)$, la loi forte s'applique et

$$P\Big\{ \omega : \lim_{n\to\infty} \frac{1}{n} \sum_{1\leq i\leq n} U_i(\omega) = 1/2 \Big\} = 1 \,.$$

Autrement dit, presque tout nombre de $[\,0, 1\,]$ admet en moyenne autant de 0 et de 1 dans son développement dyadique. Nous avons de plus montré en IV.3.6.iii que

$$P\big\{ \omega : \lim_{n\to\infty} N_n(\omega)/n = 1/2 \big\} = 1 \,,$$

i.e. presque tout nombre a de l'ordre de $n/2$ blocs de 0 et de 1 dans ses n premiers chiffres.

Une autre façon d'énoncer la loi forte des grands nombres est de dire que si $E(|X|) < \infty$, alors $S_n/n = E(X) + o(1)$ p.s. lorsque $n \to \infty$. En un certain sens, le théorème limite central donne un terme de plus dans le développement asymptotique de S_n/n, précisant le comportement limite en loi du terme $o(1)$ (modulo une hypothèse supplémentaire sur la loi des X_i). Il permet d'approximer la loi de S_n/n lorsque n est grand. Le fait remarquable est que sous la condition $E(X^2) < \infty$, la loi limite de $\sqrt{n}(S_n/n - E(X))$ ne dépend que de la variance des X_i.

***Théorème* V.5.4 (limite central).** *(i) Si $E(X^2) < \infty$, alors $\dfrac{S_n - nE(X)}{\sqrt{n}}$ converge en loi vers une variable de loi $\mathcal{N}\big(0, \mathrm{Var}(X)\big)$.*
(ii) Si $S_n/\sqrt{n}$ converge en loi, alors $E(X) = 0$ et $E(X^2) < \infty$ et la loi limite est normale centrée, de variance $\mathrm{Var}(X)$.

Démonstration. (i) Si X est constante p.s., le résultat est évident puisque $\mathcal{N}(0, 0)$ est la masse de Dirac en 0. Supposons donc que X n'est pas constante p.s. Quitte à changer X_i en $(X_i - E(X_i))/\sqrt{\mathrm{Var}(X_i)}$, on peut supposer que $E(X_i) = 0$ et $\mathrm{Var}(X_i) = 1$ pour tout i. Il suffit alors de montrer que $\lim_{n\to\infty} \varphi^{S_n/\sqrt{n}}(t) = e^{-t^2/2}$ pour tout $t \in \mathbb{R}$, puisque $e^{-t^2/2}$ est la fonction caractéristique de la loi normale $\mathcal{N}(0, 1)$. Pour cela, par indépendance et équidistribution, nous avons pour tout $n \geq 1$,

$$\varphi^{S_n/\sqrt{n}}(t) = \big(\varphi^X(t/\sqrt{n})\big)^n$$

où X est une variable ayant la loi des X_i. Or X est de carré intégrable. On peut donc dériver deux fois sa fonction caractéristique φ^X (proposition III.5.6.i) et

$$(\varphi^X)'(0) = E(X) = 0, \qquad (\varphi^X)''(0) = -E\big(X^2\big) = -1\,.$$

Donc

$$\varphi^X(u) = 1 - \frac{u^2}{2} + o(u^2) \quad \text{lorsque } u \to 0\,.$$

Lorsque $n \to \infty$,

$$\varphi^{S_n/\sqrt{n}}(t) = \Big(1 - \frac{t^2}{2n} + o\Big(\frac{1}{n}\Big)\Big)^n = e^{-t^2/2} + o(1)\,.$$

La fonction caractéristique étant à valeurs complexes, pour pleinement justifier la limite précédente, nous faisons usage de la propriété suivante. Soit $(z_n)_{n\in\mathbb{N}}$ une suite de nombres complexes tendant vers 0 ; alors

$$\lim_{n\to\infty} \Big(1 + \frac{z_n}{n}\Big)^n = 1\,.$$

En effet,

$$\Big(1 + \frac{z_n}{n}\Big)^n - 1 = \sum_{1\leq k\leq n} \mathrm{C}_n^k \Big(\frac{z_n}{n}\Big)^k\,.$$

Donc

$$\begin{aligned}
\Big|\Big(1 + \frac{z_n}{n}\Big)^n - 1\Big| &\leq \sum_{1\leq k\leq n} \mathrm{C}_n^k \Big(\frac{|z_n|}{n}\Big)^k \\
&= \sum_{0\leq k\leq n-1} \mathrm{C}_n^{k+1} \Big(\frac{|z_n|}{n}\Big)^{k+1} \\
&= \sum_{0\leq k\leq n-1} \mathrm{C}_{n-1}^k \Big(\frac{|z_n|}{n}\Big)^k \frac{|z_n|}{k+1} \\
&\leq |z_n| \Big(1 + \frac{|z_n|}{n}\Big)^{n-1}
\end{aligned}$$

tend vers 0 quand $n \to \infty$.

(ii) est admis ici. On peut se reporter par exemple à Feller (1971, §IX.8). □

Exemples V.5.5. (i) Si X_i est de loi binomiale $\mathcal{B}(1,p)$ (modèle de la proportion de piles dans n jets de pile ou face, modèle des votes pour un candidat dans un sondage électoral etc.), en notant toujours $S_n = X_1 + \cdots + X_n$, nous avons pour tout $a < b$,

$$\lim_{n\to\infty} P\Big\{\, a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \,\Big\} = \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,\mathrm{d}t\,.$$

Ce théorème limite central pour des sommes de variables aléatoires de Bernoulli a été démontré initialement par de Moivre (1667–1754). La démonstration de de Moivre consiste à remarquer que S_n suit une loi binomiale $\mathcal{B}(n,p)$, puis à écrire explicitement la probabilité

$$P\left\{ a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \right\} = \sum_{k\in[np+a\sqrt{np(1-p)},np+b\sqrt{np(1-p)}]} \mathrm{C}_n^k\, p^k(1-p)^{n-k}\,.$$

Lorsque $k \in [\,np + a\sqrt{np(1-p)}, np + b\sqrt{np(1-p)}\,]$, on peut utiliser la formule de Stirling pour approximer le coefficient binomial C_n^k, ce qui conduit au résultat en approximant la somme sur k par une intégrale.

D'après l'exemple V.4.2.iv, puisque S_n/n converge vers p en probabilité d'après la loi des grands nombres, on a aussi

$$\lim_{n\to\infty} P\left\{ a \leq \frac{S_n - np}{\sqrt{n\frac{S_n}{n}(1-\frac{S_n}{n})}} \leq b \right\} = \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,\mathrm{d}t\,.$$

Autrement dit, l'intervalle aléatoire

$$\left[\, \frac{S_n}{n} - \frac{b}{\sqrt{n}}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}\,,\, \frac{S_n}{n} - \frac{a}{\sqrt{n}}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}\,\right]$$

contient p avec une probabilité voisine de $(2\pi)^{-1/2}\int_a^b e^{-t^2/2}\,\mathrm{d}t$ lorsque n est grand. Si maintenant p est inconnu, on voit qu'en observant des réalisations des X_i, on peut construire un intervalle (puisqu'alors S_n/n est observée) qui contient p avec une probabilité assez grande. Ce type de résultat est essentiel en statistique.

(ii) Montrons maintenant l'intérêt du point (iv) du théorème V.4.1 et de la proposition III.2.7. Prenons S_n une somme de variables aléatoires indépendantes, de même loi et de carré intégrable. Alors S_n vérifie le théorème limite central V.5.4. Soit maintenant N_λ une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$, c'est-à-dire telle que $P\{\,N_\lambda = k\,\} = e^{-\lambda}\lambda^k/k!$, $k \in \mathbb{N}$, indépendante de (S_n). Que peut-on-dire alors de S_{N_λ}, somme d'un nombre aléatoire de termes, lorsque $\lambda \to \infty$?

Soient

$$Z_n = \frac{S_n - nE(X_1)}{\sqrt{n}} \qquad \text{et} \qquad T_\lambda = \frac{N_\lambda - \lambda}{\sqrt{\lambda}}\,.$$

On vérifie facilement à l'aide des fonctions caractéristiques que N_λ/λ converge en loi vers la constante 1 et que T_λ converge en loi vers une variable de loi $\mathcal{N}(0,1)$. De plus, le théorème limite central V.5.4 montre que Z_n converge en loi vers

$\mathcal{N}(0, \mathrm{Var}(X_1))$. Soient maintenant U, V, deux variables aléatoires indépendantes, de loi uniforme sur $]\,0,1\,[$. Alors Z_n a même loi que $Z'_n = F^{Z_n\leftarrow}(U)$ et T_λ a même loi que $T'_\lambda = F^{T_\lambda\leftarrow}(V)$ d'après la proposition III.2.7. Il s'ensuit que S_n a même loi que $S'_n = \sqrt{n}Z'_n + nE(X_1)$ et que N_λ a même loi que $N'_\lambda = \sqrt{\lambda}T'_\lambda + \lambda$. De plus, par construction, S_{N_λ} a même loi que $S'_{N'_\lambda}$ puisque $(S'_n)_{n\geq 1}$ et N'_λ sont indépendantes. Ceci peut aussi être vérifié en écrivant que, pour tout $t \in \mathbb{R}$,

$$\begin{aligned}
P\{\, S_{N_\lambda} \leq t \,\} &= \sum_{k\in\mathbb{N}} P\{\, S_{N_\lambda} \leq t\,;\, N_\lambda = k \,\} \\
&= \sum_{k\in\mathbb{N}} P\{\, S_k \leq t\,;\, N_\lambda = k \,\} \\
&= \sum_{k\in\mathbb{N}} P\{\, S_k \leq t\}P\{N_\lambda = k \,\} \\
&= \sum_{k\in\mathbb{N}} P\{\, S'_k \leq t \,\}P\{\, N'_\lambda = k \,\} = P\{\, S'_{N'_\lambda} \leq t \,\}\,.
\end{aligned}$$

La démonstration du théorème V.4.1 ((iv)$\Rightarrow$(i)) montre que Z'_n converge p.s. vers Z de loi $\mathcal{N}(0, \mathrm{Var}(X_1))$ tandis que T'_λ converge p.s. vers T de loi $\mathcal{N}(0,1)$. Puique Z est $\sigma(U)$-mesurable et T est $\sigma(V)$-mesurable, Z et T sont indépendantes. Puisque T'_λ converge p.s. vers T, observons aussi que N'_λ converge p.s. vers $+\infty$ lorsque $\lambda \to \infty$. Il vient alors

$$\begin{aligned}
\frac{S'_{N'_\lambda} - \lambda E(X_1)}{\sqrt{\lambda}} &= \frac{S'_{N'_\lambda} - N'_\lambda E(X_1)}{\sqrt{N'_\lambda}}\sqrt{\frac{N'_\lambda}{\lambda}} + \frac{N'_\lambda - \lambda}{\sqrt{\lambda}}EX_1 \\
&= Z'_{N'_\lambda}\sqrt{\frac{N'_\lambda}{\lambda}} + T'_\lambda E(X_1)\,.
\end{aligned}$$

Puisque $N'_\lambda \to \infty$ p.s. et $\lim_{\lambda\to\infty} N'_\lambda/\lambda = 1$ p.s., $(S'_{N'_\lambda} - \lambda E(X_1))/\sqrt{\lambda}$ converge p.s. vers $Z + TE(X_1)$ qui est de loi $\mathcal{N}(0, \mathrm{Var}(X_1) + E(X_1)^2)$. Ainsi, $(S_{N_\lambda} - \lambda E(X_1))/\sqrt{\lambda}$ converge en loi vers $\mathcal{N}(0, E(X_1^2))$. Un examen attentif de cet exemple montre que tout l'intérêt du point (iv) de la définition-théorème V.4.1 est qu'il permet de transformer un problème de probabilité en un problème d'analyse ; ayant une convergence presque sûre, on peut travailler en fixant l'aléa ω, donc, en un certain sens, sur des suites déterministes.

Lorsque le paramètre $p = p_n$ de la loi binomiale $\mathcal{B}(n,p)$ étudiée à l'exemple (i) dépend de n et est de l'ordre de λ/n, $\lambda > 0$, la loi de S_n converge vers la loi de Poisson de paramètre λ.

***Théorème* V.5.6 (limite central poissonien).** *Soit S_n une variable aléatoire de loi $\mathcal{B}(n, p_n)$. Si $\lim_{n\to\infty} np_n = \lambda > 0$, S_n converge en loi vers une variable aléatoire de Poisson de paramètre λ.*

Démonstration. En vertu de l'exemple V.4.2.v, il suffit de prouver que pour tout $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P\{\, S_n = k \,\} = e^{-\lambda} \frac{\lambda^k}{k!} \,.$$

Or, pour chaque $k \in \mathbb{N}$ fixé,

$$P\{\, S_n = k \,\} = \mathrm{C}_n^k p_n^k (1 - p_n)^{n-k} = \frac{1}{k!} \cdot \frac{n!}{(n-k)!n^k} \cdot (np_n)^k \cdot \Big(1 - \frac{np_n}{n}\Big)^{n-k} \,.$$

Lorsque $n \to \infty$,

$$\frac{n!}{(n-k)!n^k} \to 1 \qquad \text{et} \qquad (np_n)^k \to \lambda^k \,,$$

et, en prenant le logarithme,

$$\Big(1 - \frac{np_n}{n}\Big)^{n-k} \to e^{-\lambda} \,.$$

La conclusion s'ensuit. □

***Commentaire* V.5.7.** Considérons encore que $E(X) = 0$ et $E(X^2) = 1$. La loi des grands nombres nous dit que S_n/n converge p.s. vers 0, mais le théorème limite central nous dit que $S_n/\sqrt{n}$ converge en loi vers $\mathcal{N}(0,1)$. Que se passe-t-il pour des normalisations de S_n entre $1/n$ (loi des grands nombres) et $1/\sqrt{n}$ (théorème limite central) ?

On peut montrer que si $\lim_{n\to\infty} a_n/\sqrt{n} = \infty$, alors $S_n/a_n \xrightarrow{P} 0$ (inégalité de Markov : $P\{\, |S_n| \geq \varepsilon a_n \,\} \leq E(S_n^2)/\varepsilon^2 a_n^2 = nE(X^2)/\varepsilon^2 a_n)$. C'est encore vrai presque sûrement si $a_n = n^{1/p}$ avec $1 < p < 2$. Mais ce n'est plus le cas si a_n est trop proche de $\sqrt{n}$. Le cas limite est obtenu pour $a_n = \sqrt{2n \ln\ln n}$ où l'on obtient la loi, dite du logarithme itéré : presque sûrement, la suite $S_n/\sqrt{2n\ln\ln n}$ est relativement compacte et l'ensemble de ses valeurs d'adhérence est l'intervalle $[-1, 1]$. En particulier,

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n\ln\ln n}} = -1 \text{ p.s.} \qquad \text{et} \qquad \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\ln\ln n}} = 1 \text{ p.s.}$$

La loi du logarithme itéré implique le fait suivant. Le théorème limite central nous dit que si $E(X^2) < \infty$, alors $Z_n = (S_n - nE(X))/\sqrt{n}$ converge en loi vers une variable aléatoire gaussienne. Existe-t-il une variable Z gaussienne telle que Z_n converge en probabilité ou p.s. vers Z ? La loi du logarithme itéré montre que non. Supposons que Z_n converge en probabilité vers Z. Quitte à extraire une sous-suite, en vertu du théorème V.2.4, nous pouvons supposer que Z_n converge presque sûrement vers Z. La loi du logarithme itéré implique $\limsup_{n\to\infty} Z_n/\sqrt{2\ln\ln n} = \big(E(X^2)\big)^{1/2}$ p.s. ce qui contredit le fait que $\lim_{n\to\infty} Z_n/\sqrt{2\ln\ln n} = 0$ p.s., puisque Z_n converge vers Z p.s. Ainsi, la convergence en loi, comme son nom l'indique et la définition le montre, ne concerne que les lois et non les variables. Elle n'en demeure pas moins extrêmement utile en pratique où l'on est souvent intéressé par les lois.

Commentaire V.5.8. Le théorème limite central peut être démontré de nombreuses autres façons. Esquissons une démonstration due à Esséen, qui ne fait pas appel aux fonctions caractéristiques, mais suppose que $E(|X|^3) < \infty$. Supposons, toujours pour simplifier que $E(X) = 0$ et $E(X^2) = 1$. Il faut et il suffit de montrer que $E(\phi(S_n/\sqrt{n}))$ converge vers $E(\phi(Z))$ où Z est de loi $\mathcal{N}(0,1)$ et ϕ est une fonction continue bornée arbitraire (*cf.* définition-théorème V.4.1). En fait, on peut montrer qu'il suffit de ne considérer que les fonctions ϕ continues, bornées, tendant vers 0 à l'infini et à dérivées bornées (puisque ces fonctions sont denses dans l'ensemble des fonctions continues, bornées, tendant vers 0 à l'infini). Soient donc ϕ une telle fonction et Z_i des copies indépendantes de Z, et indépendantes des X_i. Soient $S_{j,n} = X_1 + \cdots + X_{j-1} + Z_{j+1} + \cdots + Z_n$, $j = 1,\ldots,n$. Alors $(S_{1,n} + Z_1)/\sqrt{n}$ a même loi que Z et $S_n = X_n + S_{n,n}$. Donc

$$
\begin{aligned}
&\Big|E\big(\phi(S_n/\sqrt{n})\big) - E\big(\phi(Z)\big)\Big| \\
&\qquad \le \sum_{1\le j\le n} \left| E\left(\phi\Big(\frac{S_{j,n}+X_j}{\sqrt{n}}\Big) - \phi\Big(\frac{S_{j,n}+Z_j}{\sqrt{n}}\Big)\right)\right| \\
&\qquad = \sum_{1\le j\le n} \left| E\left(\phi\Big(\frac{S_{j,n}}{\sqrt{n}} + \frac{X_j}{\sqrt{n}}\Big) - \phi\Big(\frac{S_{j,n}}{\sqrt{n}}\Big) - \phi\Big(\frac{S_{j,n}}{\sqrt{n}} + \frac{Z_j}{\sqrt{n}}\Big) + \phi\Big(\frac{S_{j,n}}{\sqrt{n}}\Big)\right)\right| \\
&\qquad \le \sum_{1\le j\le n} \left| E\left(\frac{X_j}{\sqrt{n}}\phi'\Big(\frac{S_{j,n}}{\sqrt{n}}\Big) + \frac{X_j^2}{2n}\phi''\Big(\frac{S_{j,n}}{\sqrt{n}}\Big) + \frac{X_j^3}{6n^{3/2}}\phi''(\theta_{j,n})\right.\right. \\
&\qquad\qquad\qquad \left.\left. - \frac{Z_j}{\sqrt{n}}\phi'\Big(\frac{S_{j,n}}{\sqrt{n}}\Big) - \frac{Z_j^2}{2n}\phi''\Big(\frac{S_{j,n}}{\sqrt{n}}\Big) - \frac{Z_j^3}{6n^{3/2}}\phi''(\tau_{j,n})\right)\right|
\end{aligned}
$$

où $\theta_{j,n}$ et $\tau_{j,n}$ sont donnés par la formule de Taylor. Puisque X_j et Z_j sont indépendantes de $S_{j,n}$, de moyenne nulle et de variance 1, il vient

$$\Big|E\big(\phi(S_n/\sqrt{n})\big) - E\big(\phi(Z)\big)\Big| \leq \frac{\|\phi''\|_\infty}{n^{3/2}} \sum_{1\leq j\leq n} E\big(|X_j|^3 + |Z_j|^3\big) = o(1)$$

lorsque $n \to \infty$.

Un raffinement de la méthode permet d'obtenir le théorème limite central sous la seule condition $E(X^2) < \infty$ (voir par exemple Pollard (1984)). Il convient aussi de remarquer que la même démonstration fournit un théorème limite central pour des sommes de vecteurs aléatoires indépendants et de même loi.

Exercices

Exercice V.1. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles sur un espace probabilisé $(\Omega, \mathcal{A}, P)$; on suppose qu'il existe une suite de réels $(a_n)_{n\in\mathbb{N}}$ telle que les séries

$$\sum_n a_n \qquad \text{et} \qquad \sum_n P\{\, X_n \neq a_n \,\}$$

soient convergentes. Démontrer que la série $\sum_n X_n$ est p.s. convergente.

Exercice V.2. Soit $(X_n)_{n\in\mathbb{N}}$ une famille de variables aléatoires gaussiennes, centrées, de variance $(\sigma_n^2)_{n\in\mathbb{N}}$ convergeant en loi vers une variable aléatoire X.

a) Montrer que la suite $(\sigma_n^2)_{n\in\mathbb{N}}$ est convergente et en déduire que X suit une loi gaussienne. Étudier le cas où les X_n ne sont pas centrées.

b) On suppose que $X_n \to X$ en probabilité. Démontrer que X_n converge vers X dans tous les espaces L^p.

Exercice V.3. Montrer que pour $x > 0$,

$$e^{-x^2/2}\left(\frac{1}{x} - \frac{1}{x^3}\right) \leq \int_x^{+\infty} e^{-t^2/2}\,\mathrm{d}t \leq e^{-x^2/2}\frac{1}{x}\,.$$

Indication : intégrer par parties $t^{-1}te^{-t^2/2}$.

Soit maintenant $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes, toutes de même loi $\mathcal{N}(0,1)$. Montrer que

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \qquad \text{p.s.}$$

Montrer également que

$$\frac{\max_{1\leq i\leq n} X_i}{\sqrt{2\ln n}} \xrightarrow{P} 1\,.$$

***Exercice* V.*4*.** Soit $(X_i)_{i\in I}$ une famille de variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$; on suppose qu'il existe une fonction $G : [\,0,\infty\,[\to [\,0,\infty\,[$ vérifiant $\lim_{t\to\infty} G(t)/t = \infty$ telle que $\sup_{i\in I} E(G(|X_i|))$ est fini. Démontrer que la famille $(X_i)_{i\in I}$ est uniformément intégrable.

***Exercice* V.*5*.** Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$ convergeant en loi respectivement vers X et Y.

a) On suppose que pour tout n, X_n et Y_n sont indépendantes et que X et Y sont indépendantes. Démontrer que $X_n + Y_n$ converge en loi vers $X + Y$. Donner un exemple montrant que l'hypothèse d'indépendance est indispensable.

b) On suppose que $Y = 0$. Prouver que $X_n + Y_n$ converge en loi vers X et $X_n Y_n$ converge en loi vers 0.

***Exercice* V.*6*.** Soit $(\alpha_n)_{n\in\mathbb{N}}$ une suite de nombres appartenant à $[\,0,1\,]$; on lui associe une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ dont les lois vérifient

$$P\{\, X_n \leq t \,\} = \begin{cases} 0 & \text{si } t < 0, \\ \alpha_n + (1-\alpha_n)t^n & \text{si } t \in [\,0,1\,], \\ 1 & \text{si } t > 1. \end{cases}$$

À quelles conditions sur $(\alpha_n)_{n\in\mathbb{N}}$, la suite $(X_n)_{n\in\mathbb{N}}$ converge-t-elle en loi ? en probabilité ? presque sûrement ?

***Exercice* V.*7*.** Montrer que 4.1.i–iv sont équivalents à $\lim_{n\to\infty} \int \phi \,\mathrm{d}P_n = \int \phi \,\mathrm{d}P$ pour toute fonction ϕ infiniment différentiable, à support compact.

***Exercice* V.*8*.** Une formule d'inversion de la transformée de Laplace.

a) Soit $\mathcal{P}(\lambda) = \sum_{n\in\mathbb{N}} e^{-\lambda} \frac{\lambda^n}{n!} \delta_n$ la loi de Poisson de paramètre λ. Montrer que si X_λ est de loi $\mathcal{P}(\lambda\theta)$ alors $(X_\lambda - \lambda\theta)/\lambda$ converge en probabilité vers 0 lorsque $\lambda \to \infty$. En déduire que

$$\lim_{\lambda\to\infty} e^{-\lambda\theta} \sum_{k\leq\lambda x} \frac{\lambda\theta^k}{k!} = \begin{cases} 0 & \text{si } \theta > x, \\ 1 & \text{si } \theta < x. \end{cases}$$

b) Soit $L(t) = \int_0^\infty e^{-tx} \,\mathrm{d}P(x)$ la transformée de Laplace d'une loi P sur $\mathbb{R}^+$. Montrer que $L(t)$ est infiniment dérivable. Montrer que si P est de fonction de répartition F, alors

$$\lim_{\lambda\to\infty} \sum_{k\leq\lambda x} \frac{(-1)^k}{k!} \lambda^k L^{(k)}(\lambda) = F(x)$$

en tout point de continuité de F.

***Exercice* V.*9*.** Une formule d'inversion de la transformée de Fourier. Soient X, Y deux variables aléatoires réelles indépendantes. Notons f^X la densité de X.
a) Montrer que $E(e^{-itY}\varphi^X(Y)) = E(\varphi^Y(X-t))$, $t \in \mathbb{R}$.
b) Prendre Y de loi $\mathcal{N}(0,\sigma^2)$ et supposer φ^X intégrable par rapport à la mesure de Lebesgue. En considérant $\sigma \to \infty$, montrer la formule donnée au théorème III.5.4.
c) Montrer que pour tous x, y et $m > 0$,

$$\begin{aligned}\frac{1}{2\pi}\int_{-m}^{m} &\frac{e^{-itx}-e^{-ity}}{it}\varphi^X(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty}\left(\frac{1}{\pi}\int_0^m \frac{\sin t(z-x)}{t}\,\mathrm{d}t - \frac{1}{\pi}\int_0^m \frac{\sin t(z-y)}{t}\,\mathrm{d}t\right) f^X(z)\,\mathrm{d}z\,.\end{aligned}$$

On rappelle que $\displaystyle\int_0^\infty \frac{\sin(tx)}{t}\,\mathrm{d}t = \mathrm{signe}(x)\pi/2$.

En déduire que si x et y sont des points de continuité de F^X, alors

$$F^X(y) - F^X(x) = \lim_{m\to\infty}\frac{1}{2\pi}\int_{-m}^{m}\frac{e^{-itx}-e^{-ity}}{it}\varphi^X(t)\,\mathrm{d}t\,,$$

ce qui donne une formule d'inversion de Fourier, et montre que φ^X caractérise F^X et donc P^X.

***Exercice* V.*10*.** Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires, de loi uniforme sur $[\,0,1\,]$. Soit N_n une variable aléatoire de loi binomiale $\mathcal{B}(n,p)$. Montrer que $n\min_{1\leq i\leq N_n} X_i$ converge en loi, lorsque $n\to\infty$, vers une variable aléatoire exponentielle de moyenne $1/p$.

***Exercice* V.*11*.** Appliquer le théorème limite central à une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de même loi de Poisson de paramètre 1 pour trouver la limite de la suite

$$u_n = e^{-n}\sum_{0\leq k\leq n}\frac{n^k}{k!}\,, \qquad n\in\mathbb{N}\,.$$

Réponse : 1/2.

***Exercice* V.*12*.** Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires réelles, indépendantes et de même loi P. On appelle mesure empirique de $X_1,\ldots,X_n$ la loi de probabilité $P_n = n^{-1}\sum_{1\leq i\leq n}\delta_{X_i}$ (cette mesure est aléatoire puisque les X_i le sont). Montrer que presque sûrement P_n converge étroitement vers P.

Indication : utiliser la définition V.4.1.i et la loi forte des grands nombres. Si F_n (resp. F) est la fonction de répartition de P_n (resp. P), on prendra garde au fait que l'ensemble de mesure nulle sur lequel $\lim_{n\to\infty} F_n(t) \neq F(t)$ doit pouvoir être pris indépendant de t ; à cette fin, on peut utiliser la monotonie et la bornitude de F.

Exercice V.13. Notons $U^{(p)}$ la variable aléatoire réelle $\sum_{i\geq 1} 2^{-i} X_i$ où les X_i sont indépendantes, de loi $\mathcal{B}(1,p)$ et soit $\mathcal{L}^{(p)}$ la loi de $U^{(p)}$. Soit $x \in [\,0,1\,]$. Notons $x = \sum_{i\geq 1} 2^{-i} x_i$ son développement en base 2.

a) En utilisant la loi forte des grands nombres, montrer que sous $\mathcal{L}^{(p)}$, pour presque tout x, la proportion de 1 dans le développement en base 2 (*i.e.* $n^{-1}\sum_{1\leq i\leq n} x_i$) tend vers p. En déduire que les lois $\mathcal{L}^{(p)}$ sont étrangères les unes par rapport aux autres.

b) Montrer que $\mathcal{L}^{(1/2)}$ est la mesure de Lebesgue sur $[\,0,1\,]$ (loi uniforme sur $[\,0,1\,]$).

Indication : déterminer les mesures sous $\mathcal{L}^{(1/2)}$ des intervalles dyadiques.

Montrer que les lois $\mathcal{L}^{(p)}$ n'ont pas de parties discrètes. Donc si $p \notin \{\,0, 1/2, 1\,\}$ la fonction de répartition de $\mathcal{L}^{(p)}$ est continue, mais pas absolument continue.

Exercice V.14. Au théorème IV.3.1 nous avons vu comment construire une suite infinie de variables aléatoires indépendantes. Donnons ici une construction plus explicite sur $\mathbb{R}$. Soient X_n, $n \geq 1$, les variables aléatoires de loi $\mathcal{B}(1,1/2)$ construites à l'exemple IV.1.7.ii. En utilisant l'exercice V.13 et l'exemple V.1.3.i, montrer qu'on peut construire une suite $(U_n)_{n\geq 1}$ de variables aléatoires uniformes sur $[\,0,1\,]$, indépendantes.

Indication : considérer la construction en triangle

$$
\begin{aligned}
&U_1 = 2^{-1}X_1 + 2^{-2}X_2 + 2^{-3}X_4 + 2^{-4}X_7 + \cdots\\
&U_2 = 2^{-1}X_3 + 2^{-2}X_5 + 2^{-3}X_8 + \cdots\\
&U_3 = 2^{-1}X_6 + 2^{-2}X_9 + \cdots\\
&U_4 = 2^{-1}X_{10} + \cdots\\
&\quad\vdots
\end{aligned}
$$

Montrer alors que si l'on se donne une famille de loi P_i, $i \in \mathbb{N}$, sur $\mathbb{R}$, on peut construire une suite de variables aléatoires réelles $(Z_i)_{i\in\mathbb{N}}$, indépendantes, telles que Z_i est de loi P_i. Nous avons donc dans ce cas une preuve constructive du théorème de Kolmogorov IV.3.1.

Exercice V.15. On considère une marche aléatoire sur $\mathbb{Z}$, partant de l'origine, représentée par une suite $(X_n)_{n\geq 1}$ de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes, et de même loi de Bernoulli sur $\{-1, 1\}$ de paramètre $0 < p < 1$ (autrement dit $P\{X_n = 1\} = 1 - P\{X_n = -1\} = p$ pour tout n). On pose $S_n = X_1 + \cdots + X_n$, $n \geq 1$, et par convention $S_0 = 0$. La variable aléatoire S_n représente donc la position au temps n du marcheur parti de 0. On s'intéresse à la probabilité de revenir une infinité de fois à son point de départ, c'est-à-dire à la probabilité de l'événement

$$A = \{ S_n = 0 \text{ pour une infinité de } n \}.$$

a) Démontrer que S_n/n converge presque sûrement vers une limite que l'on précisera.
b) Déduire de la question précédente que $P(A) = 0$ si $p \neq 1/2$.
c) On suppose à présent que $p = 1/2$.
i) Pour tout $k \geq 0$, soit $Z_k = (S_{2^{k+1}} - S_{2^k})/\sqrt{2^k}$. Prouver que Z_k a même loi que $S_{2^k}/\sqrt{2^k}$. En déduire, en faisant usage du théorème limite central, que pour tout réel M,

$$\sum_{k\geq 0} P\{ Z_k \geq M \} = \infty .$$

ii) Conclure de la question précédente que $P\{\sup_k Z_k \geq M\} = 1$ pour tout M, puis que $P\{\sup_k |Z_k| = \infty\} = 1$. En déduire que

$$P\Big\{ \sup_{n\geq 1} \Big| \frac{S_n}{\sqrt{n}} \Big| = \infty \Big\} = 1 .$$

iii) Démontrer avec la loi du 0–1 que l'événement $B^+ = \{\sup_{n\geq 1} S_n/\sqrt{n} = +\infty\}$ est de probabilité 0 ou 1. Soit $B^- = \{\inf_{n\geq 1} S_n/\sqrt{n} = -\infty\}$. Démontrer que $P(B^+) = P(B^-)$. Conclure, à l'aide de la question ii), que $P(B^+) = P(B^-) = 1$.
iv) Déduire de ce qui précède que $P(A) = 1$.

Exercice V.16. Soient μ et ν deux mesures de probabilité sur un espace mesurable $(E, \mathcal{B})$. On appelle distance en variation totale la quantité

$$\|\mu - \nu\| = \sup_{B\in\mathcal{B}} |\mu(B) - \nu(B)| .$$

Soient X et Y deux variables aléatoires sur $(\Omega, \mathcal{A}, P)$ de lois respectives P^X et P^Y.
a) Montrer l'inégalité $\|P^X - P^Y\| \leq P\{X \neq Y\}$.

b) Soient Y et ε deux variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, Y de loi de Poisson de paramètre $0 < p < 1$ et ε de loi de Bernoulli de paramètre $1 - (1-p)e^p$. Soit $X = 1 - \mathbb{1}_{\{\varepsilon = Y = 0\}}$. Calculer la loi de X et démontrer que l'on a $P\{\, X \neq Y \,\} \leq p^2$.

c) Soit S une variable aléatoire de même loi qu'une somme de n variables aléatoires indépendantes de lois de Bernoulli de paramètre p_i, $0 < p_i < 1$, $i = 1, \ldots, n$. Démontrer qu'il existe une variable aléatoire Z suivant une loi de Poisson de paramètre $\lambda = \sum_{1 \leq i \leq n} p_i$ telle que

$$\|P^S - P^Z\| \leq \sum_{1 \leq i \leq n} p_i^2 \,.$$

d) Retrouver le théorème V.5.6 pour $p_i = \lambda/n$, $\lambda > 0$, $1 \leq i \leq n$ $(n \geq \lambda)$.

VI

PROBABILITÉS ET ESPÉRANCES CONDITIONNELLES

Commençons par un exemple. Dans un jeu de dé, à chaque jet, chacune des six faces sont équiprobables. On peut modéliser le jet en se donnant l'espace $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$, la tribu de ses parties $\mathcal{P}(\Omega)$ et la loi de probabilité P définie par $P(\{ \omega \}) = 1/6$ pour tout $\omega \in \Omega$. La variable aléatoire X, représentant le résultat du jet, peut être prise comme étant l'identité de Ω sur lui-même. Imaginons maintenant que nous lancions le dé sans le regarder, et qu'un spectateur nous dise que nous avons obtenu un chiffre pair. Étant donnée cette information, nous pouvons réévaluer nos chances d'obtenir un certain $\omega \in \Omega$. Clairement, si ω est impair, cette chance est nulle, et si ω est pair, elle est $1/3$. Notons $\Omega_{\text{pair}} = \{ 2, 4, 6 \}$. La façon dont nous évaluons la probabilité de ω sachant que $\omega \in \Omega_{\text{pair}}$ consiste à évaluer $P(\{ \omega \} \cap \Omega_{\text{pair}})/P(\Omega_{\text{pair}})$, ou, ce qui revient au même ici, le nombre de façon d'obtenir ω dans Ω_{pair}, divisé par le cardinal de Ω_{pair}.

De façon plus générale, sur un espace $(\Omega, \mathcal{A}, P)$, si l'on sait effectivement qu'un événement B est réalisable (*i.e.* $P(B) > 0$), on peut construire une nouvelle mesure de probabilité

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}\,, \qquad A \in \mathcal{A}\,,$$

appelée probabilité conditionnelle de A sachant B. Observons que $P(B \mid B) = 1$ et si $A \cap B = \emptyset$ ou $P(A \cap B) = 0$, alors $P(A \mid B) = 0$.

On peut imaginer des situations plus compliquées où l'on souhaite naturellement conditionner par un événement de mesure nulle. Par exemple, si on admet que le poids d'un individu est une variable aléatoire continue, on pourrait chercher la loi de la taille étant donné le poids. Nous verrons dans ce chapitre comment formaliser cela. Mais commençons par le cas simple où l'on conditionne par un événement de probabilité positive.

VI.1. Conditionnement discret

Définition VI.1.1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, et soit $B \in \mathcal{A}$ tel que $P(B) > 0$.

(i) On appelle probabilité conditionnelle de l'événement A sachant B le nombre $P(A \cap B)/P(B)$, noté $P(A \mid B)$.

(ii) On appelle loi conditionnelle sachant B, la mesure de probabilité définie par $A \in \mathcal{A} \mapsto P(A \cap B)/P(B)$, notée $P(\,\cdot \mid B)$.

Observons que si et seulement si A et B sont indépendants, $P(A \mid B) = P(A)$, c'est-à-dire que la connaissance de l'événement B n'apporte aucune information sur la réalisation ou non de A. Si $P(A) > 0$, remarquons que

$$P(A \mid B)P(B) = P(B \mid A)P(A) = P(A \cap B)\,.$$

Notons également que si X est une variable aléatoire réelle intégrable sur $(\Omega, \mathcal{A}, P)$, son intégrale par rapport à la probabilité conditionnelle $P(\,\cdot \mid B)$ est égale à

$$\int_\Omega X \,\mathrm{d}P(\,\cdot \mid B) = \frac{1}{P(B)} \int_B X \,\mathrm{d}P\,.$$

Comme pour la construction usuelle de l'intégrale, ceci se vérifie d'abord sur les variables indicatrices et étagées, puis se prolonge.

Enfin, dans le cas où $\Omega = \mathbb{R}^d$, puisque la probabilité conditionnelle $P(\,\cdot \mid B)$ est une probabilité, on peut définir une fonction de répartition conditionnelle sachant B, $x \in \mathbb{R}^d \mapsto P(\,]-\infty, x\,] \mid B) \in [\,0,1\,]$, une fonction caractéristique conditionnelle $t \in \mathbb{R}^d \mapsto \int e^{i\langle t,x\rangle} \,\mathrm{d}P(x \mid B)$, lesquelles caractérisent la loi conditionnelle $P(\,\cdot \mid B)$.

Exemples VI.1.2. (i) Soit X une variable aléatoire de loi exponentielle de paramètre 1. Observons que pour tous $s, t > 0$, la probabilité que X dépasse $s + t$ sachant que X dépasse s est donnée par

$$\begin{aligned} P\{\,X \geq s+t \mid X \geq s\,\} &= \frac{P\{\,X \geq s+t\,;\, X \geq s\,\}}{P\{\,X \geq s\,\}} = \frac{P\{\,X \geq s+t\,\}}{P\{\,X \geq s\,\}} \\ &= \frac{e^{-(s+t)}}{e^{-s}} = e^{-t}\,. \end{aligned}$$

On constate que cette probabilité conditionnelle est égale à $P\{X \geq t\}$. Cette propriété caractéristique de la loi exponentielle est traditionnellement appelée l'absence de mémoire (penser à X comme mesurant un temps aléatoire).

(ii) Soit $(U_1, \ldots, U_n)$ un vecteur de variables aléatoires indépendantes, toutes de loi uniforme sur $[0,1]$. Soit $M_n = \max_{1 \leq i \leq n} U_i$ et $m_n = \min_{1 \leq i \leq n} U_i$. Alors pour tous $u_1, \ldots, u_n \in [0,1]$, et tous $0 \leq a < b \leq 1$,

$$
\begin{aligned}
P\{U_1 \leq u_1, \ldots, U_n \leq u_n \mid a \leq m_n \leq M_n \leq b\} \\
&= \frac{P\{U_1 \leq u_1, \ldots, U_n \leq u_n\,;\, a \leq m_n\,;\, M_n \leq b\}}{P\{a \leq m_n\,;\, M_n \leq b\}} \\
&= \frac{P\{a \leq U_1 \leq u_1 \wedge b\,;\, \ldots\,;\, a \leq U_n \leq u_n \wedge b\}}{P\{a \leq U_1 \leq b\,;\, \ldots\,;\, a \leq U_n \leq b\}} \\
&= \frac{\prod_{1 \leq i \leq n} P\{a \leq U_i \leq u_i \wedge b\}}{\prod_{1 \leq i \leq n} P\{a \leq U \leq b\}} \\
&= \frac{\prod_{1 \leq i \leq n} (u_i \wedge b - a)}{\prod_{1 \leq i \leq n} (b-a)} \\
&= \prod_{1 \leq i \leq n} \frac{u_i \wedge b - a}{b-a}\,.
\end{aligned}
$$

On constate que cette probabilité est égale à $P\{V_1 \leq u_1, \ldots, V_n \leq v_n\}$ où $V_1, \ldots, V_n$ sont des variables aléatoires indépendantes de même loi $\mathcal{U}_{[a,b]}$. Autrement dit,

$$
\mathcal{U}_{[0,1]}^{\otimes n}\big(\cdot \,\big|\, a \leq m_n \leq M_n \leq b\big) = \mathcal{U}_{[a,b]}^{\otimes n}(\cdot)\,.
$$

L'objectif que nous poursuivons maintenant est d'essayer de généraliser cette définition dans deux directions. On essayera ainsi de remplacer A par une variable aléatoire, suivant le schéma $E(\mathbb{1}_A) = P(A)$, et B par une sous-tribu de Ω. Comme annoncé, nous débutons par une situation discrète.

Définition VI.1.3. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Une famille d'événements $(B_i)_{i \in I}$, $I \subset \mathbb{N}$, forme un système complet d'événements si les B_i sont disjoints et $P(\bigcup_{i \in I} B_i) = \sum_{i \in I} P(B_i) = 1$.

Observons que, quitte à ajouter l'événement de mesure nulle $N = \Omega \setminus \bigcup_{i \in I} B_i$, la famille $(B_i)_{i \in I}$ forme une partition de Ω.

Proposition VI.1.4. *Soit $(B_i)_{i \in I}$ un système complet d'événements sur $(\Omega, \mathcal{A}, P)$, et soit $I^* = \{i \in I : P(B_i) > 0\}$. Pour tout $A \in \mathcal{A}$,*

(i) $P(A) = \sum_{i \in I^*} P(A \mid B_i) P(B_i)$ *(formule des probabilités totales).*
(ii) De plus, si $P(A) > 0$*, pour tout* k *tel que* $P(B_k) > 0$*, on a la règle, dite de Bayes,*

$$P(B_k \mid A) = \frac{P(A \mid B_k) P(B_k)}{\sum_{i \in I^*} P(A \mid B_i) P(B_i)}\,.$$

Démonstration. C'est immédiat puisque $P(A) = \sum_{i \in I^*} P(A \cap B_i)$ et, pour tout k,

$$P(B_k \mid A) P(A) = P(B_k \cap A) = P(A \mid B_k) P(B_k)\,. \qquad \square$$

L'intérêt de la règle de Bayes est qu'elle exprime $P(B_k \mid A)$ en fonction des $P(A \mid B_k)$, et donc renverse les conditionnements.

Exemple VI.1.5. Voici une application de la règle de Bayes qui justifie pleinement l'utilisation des questionnaires à choix multiple aux examens.

Considérons des questions où m réponses possibles sont proposées et supposons qu'un candidat a une probabilité p de connaître la réponse à une question prise au hasard parmi un ensemble fini de questions. Sachant que le candidat a répondu correctement à la question, quelle est la probabilité qu'il sache effectivement la réponse ? On suppose qu'un candidat ne sachant pas la réponse répond « au hasard », et donc que chacune des m réponses possibles sont équiprobables. Soit A l'événement « le candidat répond correctement » et B l'événement « le candidat connaît la réponse ». Appliquons la règle de Bayes,

$$\begin{aligned}
P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} &= \frac{P(A \mid B)P(B)}{P(A \cap B) + P(A \cap B^c)} \\
&= \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)} = \frac{1 \cdot p}{1 \cdot p + \frac{1}{m}(1-p)} \\
&= \frac{mp}{mp + 1 - p}\,.
\end{aligned}$$

Donc, plus m est grand, plus $P(B \mid A)$ est grand ; c'est assez intuitif ; il est probable que le candidat connaisse la réponse s'il a donné une bonne réponse parmi de nombreuses proposées. Remarquons que pour $m = 3$ et $p = 1/2$, $P(B \mid A) = 3/4$, ce qui est somme toute assez grand. On conçoit donc qu'un questionnaire d'une trentaine de questions, chacune à trois ou quatre réponses possibles, soit à même de rendre compte du savoir d'un étudiant !

Observons maintenant que la tribu $\mathcal{B}$ engendrée par une partition $(B_i)_{i \in I}$, $I \subset \mathbb{N}$, d'événements de $\mathcal{A}$ est décrite comme la collection de toutes les unions possibles d'événements B_i et de leurs complémentaires. De ce point de vue, tout ensemble $A \in \mathcal{B}$ peut être fractionné sur les ensembles élémentaires B_i. Ceci conduit à la définition suivante.

Définition VI.1.6. Soit $\mathcal{B}$ une tribu. Un événement $B \in \mathcal{B}$ est appelé un atome de $\mathcal{B}$ si pour tout événement $C \in \mathcal{B}$ qui est inclus dans B, soit $C = \emptyset$, soit $C = B$.

Exemples VI.1.7. (i) Si $(B_i)_{i \in I}$ est une partition mesurable de $(\Omega, \mathcal{A})$, les B_i sont les atomes de la tribu $\sigma(B_i : i \in I)$ engendrée par les B_i.

(ii) Soit $E = \{ b_i : i \in I \subset \mathbb{N} \}$ un ensemble fini ou dénombrable et soit $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{P}(E))$ une variable aléatoire discrète. Les événements

$$Y^{-1}(\{ b_i \}) = \{ Y = b_i \} = \{ \omega \in \Omega : Y(\omega) = b_i \}$$

forment un système complet. Ce sont les atomes de la tribu engendrée par Y. Réciproquement, si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$ engendrée par un système complet d'événements $(B_i)_{i \in I}$, $I \subset \mathbb{N}$, une variable aléatoire réelle Y, définie sur $(\Omega, \mathcal{A}, P)$ et $\mathcal{B}$-mesurable, est nécessairement constante sur chaque atome de $\mathcal{B}$, et donc de la forme $Y = \sum_{i \in I} \alpha_i \mathbb{1}_{B_i}$ p.s. En effet, supposons au contraire qu'il existe $\omega, \omega' \in B_i$ tels que $Y(\omega) \neq Y(\omega')$. Il existe alors des boréliens C, C' disjoints tels que $Y(\omega) \in C$ et $Y(\omega') \in C'$. Or $Y^{-1}(C) = \bigcup_{j \in J} B_j$ et $Y^{-1}(C') = \bigcup_{j \in J'} B_j$ pour certains J et J'. Nécessairement, les ensembles d'indices J et J' contiennent tous les deux l'indice i puisque $\omega \in Y^{-1}(C)$ et $\omega' \in Y^{-1}(C')$, et $\omega, \omega' \in B_i$. Ainsi, l'ensemble $Y^{-1}(C) \cap Y^{-1}(C')$ n'est pas vide, ce qui est impossible puisque $Y^{-1}(C) \cap Y^{-1}(C') = Y^{-1}(C \cap C') = Y^{-1}(\emptyset) = \emptyset$.

Définition VI.1.8. Soit $\mathcal{B}$ une sous-tribu dans $(\Omega, \mathcal{A}, P)$, engendrée par un système complet d'événements $(B_i)_{i \in I}$, $I \subset \mathbb{N}$. Soit $I^* = \{ i \in I : P(B_i) > 0 \}$. On appelle probabilité conditionnelle de $A \in \mathcal{A}$ sachant $\mathcal{B}$ la variable aléatoire $\sum_{i \in I^*} P(A \mid B_i) \mathbb{1}_{B_i}$, notée aussi $P(A \mid \mathcal{B})$.

La probabilité conditionnelle de A sachant une sous-tribu $\mathcal{B}$ est donc une variable aléatoire, constante sur les atomes de cette sous-tribu, et donc mesurable par rapport à $\mathcal{B}$. Pour tout $\omega \in \bigcup_{i \in I} B_i$, l'application $A \in \mathcal{A} \mapsto P(A \mid \mathcal{B})(\omega)$ est une mesure de probabilité telle que $P(B_i \mid \mathcal{B})(\omega) = 1$ si $\omega \in B_i$, et $P(A \mid \mathcal{B})(\omega) = 0$ si $\omega \in B_i$ et $P(A \cap B_i) = 0$.

Il convient enfin de remarquer que $P(A \mid \mathcal{B})$ est « proche » de la fonction $\mathbb{1}_A$ « sur $\mathcal{B}$ », au sens où, pour tout $B \in \mathcal{B}$ tel que $P(B) > 0$,

$$\int_B P(A \mid \mathcal{B}) \, \mathrm{d}P = P(A \cap B) = \int_B \mathbb{1}_A \, \mathrm{d}P \,.$$

En effet, puisque $B = \bigcup_{i \in J} B_i$ pour un certain ensemble d'indices J, il suffit de montrer l'identité pour un atome ; c'est alors une conséquence de la définition VI.1.1, puisque

$$\begin{aligned}\int_{B_j} P(A \mid B)(\omega)\,\mathrm{d}P(\omega) &= \int_{B_j} \sum_{i \in I} P(A \mid B_i)\mathbb{1}_{B_i}(\omega)\,\mathrm{d}P(\omega)\\ &= P(A \mid B_j) \int_{B_j} \mathbb{1}_{B_j}(\omega)\,\mathrm{d}P(\omega)\\ &= P(A \mid B_j)P(B_j) = P(A \cap B_j)\,.\end{aligned}$$

Il est important de comprendre cette notion de proximité. À titre de comparaison, deux variables aléatoires réelles intégrables X et Y sur $(\Omega, \mathcal{A}, P)$ sont égales p.s. si et seulement si $\int_A X\,\mathrm{d}P = \int_A Y\,\mathrm{d}P$ pour tout $A \in \mathcal{A}$. En effet, prenant $A = \{X \geq Y\}$, il vient

$$0 = \int_{\{X \geq Y\}} (X - Y)\,\mathrm{d}P = \int (X - Y)^+\,\mathrm{d}P\,,$$

d'où $(X - Y)^+ = 0$ p.s. ; et de la même façon, $(Y - X)^+ = 0$, d'où $X = Y$ p.s. Dans le cas de la probabilité conditionnelle $P(A \mid \mathcal{B})$, nous avons pour tout $B \in \mathcal{B} \subset \mathcal{A}$, $\int_B Y\,\mathrm{d}P = \int_B X\,\mathrm{d}P$ avec $X = \mathbb{1}_A$, $Y = P(A \mid \mathcal{B})$.

Observons que l'on peut réécrire

$$P(A \mid \mathcal{B}) = \sum_{i \in I^*} \frac{P(A \cap B_i)}{P(B_i)} \mathbb{1}_{B_i} = \sum_{i \in I^*} \left(\frac{1}{P(B_i)} \int_{B_i} \mathbb{1}_A\,\mathrm{d}P \right) \mathbb{1}_{B_i}\,.$$

Donc $P(A \mid \mathcal{B})$ est la variable aléatoire étagée obtenue en moyennant $\mathbb{1}_A$ sur les atomes de $\mathcal{B}$. On peut alors remplacer $\mathbb{1}_A$ par une variable aléatoire, ce qui conduit à la définition suivante.

Définition VI.1.9. Soit X une variable aléatoire réelle intégrable sur $(\Omega, \mathcal{A}, P)$ et soit $\mathcal{B}$ une sous-tribu engendrée par un système complet d'événements $(B_i)_{i \in I}$, $I \subset \mathbb{N}$. Soit $I^* = \{\, i \in I : P(B_i) > 0 \,\}$. On appelle espérance conditionnelle de X sachant $\mathcal{B}$, notée $E(X \mid \mathcal{B})$, la variable aléatoire $\mathcal{B}$-mesurable

$$\sum_{i \in I^*} \frac{1}{P(B_i)} \left(\int_{B_i} X\,\mathrm{d}P \right) \mathbb{1}_{B_i}\,.$$

Il est tout à fait important de remarquer que dans la définition de l'espérance conditionnelle, la somme est sur l'ensemble d'indices I^*. En conséquence, sa valeur en tout point $\omega \in \bigcup_{I \setminus I^*} B_i$ n'est pas définie. On pourrait aussi altérer les événements B_i et leur adjoindre ou retrancher des événements de mesure nulle. Ceci changerait éventuellement l'espérance conditionnelle, mais seulement sur un ensemble de mesure nulle. Il convient donc de bien comprendre que l'espérance conditionnelle n'est définie que P-presque sûrement.

De même que $E(\mathbb{1}_A) = P(A)$, par construction, $E(\mathbb{1}_A \mid \mathcal{B}) = P(A \mid \mathcal{B})$. Remarquons aussi que si $B \in \mathcal{B}$,

$$\int_B E(X \mid \mathcal{B}) \, \mathrm{d}P = \int_B X \, \mathrm{d}P \, ,$$

et que si X est $\mathcal{B}$-mesurable, $E(X \mid \mathcal{B}) = X$ p.s.

***Notation* VI.1.10.** *Si $\mathcal{B}$ est engendrée par une variable aléatoire discrète Y, on note $E(X \mid Y) = E(X \mid \mathcal{B})$.*

***Exemple* VI.1.11.** Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, P)$ suivant une loi de Poisson de paramètre $\lambda > 0$. Soit $Y = 2\lfloor X/2 \rfloor$, où $\lfloor \cdot \rfloor$ est la fonction partie entière. Calculons les espérances conditionnelles $E(X \mid Y)$ et $E(Y \mid X)$. Puisque Y est X-mesurable, $E(Y \mid X) = Y$ p.s. Pour évaluer $E(X \mid Y)$, nous étudions les atomes de $\sigma(Y)$; ce sont les ensembles $B_n = \{ Y = 2n \}$, $n \geq 0$. On évalue

$$\begin{aligned}
\int_{B_n} X \, \mathrm{d}P &= \int_{\{ X = 2n \}} X \, \mathrm{d}P + \int_{\{X = 2n+1\}} X \, \mathrm{d}P \\
&= 2nP\{ X = 2n \} + (2n+1)P\{ X = 2n+1 \} \\
&= 2ne^{-\lambda} \frac{\lambda^{2n}}{(2n)!} + (2n+1)e^{-\lambda} \frac{\lambda^{2n+1}}{(2n+1)!} \, .
\end{aligned}$$

De même,

$$\begin{aligned}
P(B_n) = P\{ Y = 2n \} &= P\{ X = 2n \} + P\{ X = 2n+1 \} \\
&= e^{-\lambda} \frac{\lambda^{2n}}{(2n)!} + e^{-\lambda} \frac{\lambda^{2n+1}}{(2n+1)!} \, .
\end{aligned}$$

D'où

$$\frac{1}{P(B_n)} \int_{B_n} X \, \mathrm{d}P = \frac{(2n+\lambda)(2n+1)}{2n+1+\lambda} \, .$$

Ainsi

$$E(X \mid Y) = \sum_{n \geq 0} \frac{(2n+\lambda)(2n+1)}{2n+1+\lambda} \mathbb{1}_{\{Y=2n\}} = \frac{(Y+\lambda)(Y+1)}{Y+1+\lambda} \, .$$

On voit sur cet exemple, comme dans la situation générale, que l'espérance conditionnelle est définie seulement p.s., puisqu'elle peut prendre n'importe quelle valeur sur les atomes de mesure nulle.

VI.2. Conditionnement (général)

Nous pouvons généraliser les exemples précédents, en remplaçant une tribu engendrée par un système complet d'événements par une tribu arbitraire. L'intérêt de cette généralisation est que nous pourrons alors conditionner par la tribu engendrée par une variable aléatoire. Ainsi, l'espérance conditionnelle par rapport à une variable aléatoire continue pourra être définie, généralisant la définition VI.1.9 et la notation VI.1.10.

Dans les définitions précédentes de l'espérance conditionnelle, nous avons vu que celle-ci n'est définie que presque partout. Donc si A est un événement de mesure nulle, on souhaite que si deux versions de l'espérance conditionnelle coïncident sur une partie de A, elles soient encore considérées comme étant égales d'un point de vue probabiliste.

Définition et théorème VI.2.1. *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, et soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Soit de plus X une variable aléatoire réelle sur $(\Omega, \mathcal{A}, P)$, intégrable. Alors il existe une unique (p.s.) variable aléatoire, appelée espérance conditionnelle de X sachant $\mathcal{B}$, notée $E(X \mid \mathcal{B})$, telle que*
(i) $\omega \mapsto E(X \mid \mathcal{B})(\omega)$ est $\mathcal{B}$-mesurable ;
(ii) pour tout $B \in \mathcal{B}$, $\int_B E(X \mid \mathcal{B}) \,\mathrm{d}P = \int_B X \,\mathrm{d}P$.

Démonstration. i) Unicité. Nous reprenons le raisonnement utilisé à la suite de la définition VI.1.8. Soient Z_1, Z_2, $\mathcal{B}$-mesurables, telles que pour tout $B \in \mathcal{B}$, $\int_B Z_i \,\mathrm{d}P = \int_B X \,\mathrm{d}P$, $i = 1, 2$. Alors, puisque $\{ Z_1 \geq Z_2 \}$ est $\mathcal{B}$-mesurable,

$$0 = \int_{\{ Z_1 \geq Z_2 \}} (Z_1 - Z_2) \,\mathrm{d}P = \int (Z_1 - Z_2)^+ \,\mathrm{d}P,$$
$$0 = \int_{\{ Z_2 \geq Z_1 \}} (Z_2 - Z_1) \,\mathrm{d}P = \int (Z_2 - Z_1)^+ \,\mathrm{d}P\,,$$

et donc $Z_1 = Z_2$ p.s.

ii) Existence. Montrons-la d'abord en supposant X de carré intégrable. Alors X est un élément de l'espace de Hilbert $\mathrm{L}^2(\Omega, \mathcal{A}, P)$. L'espace $\mathrm{L}^2(\Omega, \mathcal{B}, P)$ est fermé dans $\mathrm{L}^2(\Omega, \mathcal{A}, P)$. On peut donc parler de la projection QX de X sur $\mathrm{L}^2(\Omega, \mathcal{B}, P)$. Cette projection vérifie

$$\forall\, U \in \mathrm{L}^2(\Omega, \mathcal{B}, P)\ ,\ \langle X - QX, U \rangle = \int_\Omega (X - QX) U \,\mathrm{d}P = 0\,.$$

Puisque QX est une classe d'équivalence de fonctions de $\mathcal{L}^2(\Omega, \mathcal{B}, P)$, on définit $E(X \mid \mathcal{B})$ comme un représentant $\mathcal{B}$-mesurable de la classe de QX. En prenant $U = \mathbb{1}_B$, $B \in \mathcal{B}$, on voit que

$$0 = \int_\Omega (X - QX)\mathbb{1}_B \, \mathrm{d}P = \int_B X \, \mathrm{d}P - \int_B E(X \mid \mathcal{B}) \, \mathrm{d}P \,,$$

ce qui fournit (ii) et prouve l'existence dans ce cas.

Pour étendre l'existence au cas des variables uniquement intégrables, notons que si $X \in \mathrm{L}^2(\Omega, \mathcal{A}, P)$ et $X \geq 0$ p.s., alors $E(X \mid \mathcal{B}) \geq 0$ p.s. (prendre $B = \{ E(X \mid \mathcal{B}) < 0 \} \in \mathcal{B}$ dans (ii)).

Supposons maintenant X intégrable, positive p.s. Pour tout n, $X_n = X \wedge n$ est de carré intégrable. On peut ainsi définir $E(X_n \mid \mathcal{B})$ vérifiant (i)–(ii). De plus

$$\begin{aligned} E(X_{n+1} \mid \mathcal{B}) - E(X_n \mid \mathcal{B}) &= QX_{n+1} - QX_n = Q(X_{n+1} - X_n) \\ &= E(X_{n+1} - X_n \mid \mathcal{B}) \geq 0 \end{aligned}$$

d'après ce qui précède. Enfin, $E(E(X_n \mid \mathcal{B})) = E(X_n) \leq E(X) < \infty$. D'après le théorème de convergence monotone II.2.1, la suite $E(X_n \mid \mathcal{B})$ converge p.s. vers une variable aléatoire notée $E(X \mid \mathcal{B})$, $\mathcal{B}$-mesurable et intégrable. Il ne reste plus qu'à vérifier que $E(X \mid \mathcal{B})$ vérifie (ii), ce qui est encore une conséquence du théorème de convergence monotone. En effet, si $B \in \mathcal{B}$,

$$\begin{aligned} \int_B E(X \mid \mathcal{B}) \, \mathrm{d}P &= \lim_{n\to\infty} \int_B E(X_n \mid \mathcal{B}) \, \mathrm{d}P \\ &= \lim_{n\to\infty} \int_B X_n \, \mathrm{d}P = \int_B X \, \mathrm{d}P \,. \end{aligned}$$

Enfin, si X est intégrable, écrivons $X = X^+ - X^-$ et posons

$$E(X \mid \mathcal{B}) = E(X^+ \mid \mathcal{B}) - E(X^- \mid \mathcal{B}) \,.$$

Ceci termine la construction de l'espérance conditionnelle. $\square$

Une autre preuve de l'existence de l'espérance conditionnelle $E(X \mid \mathcal{B})$ peut être fournie à l'aide du théorème de Radon-Nikodym II.3.3. En effet, la mesure $\mu(B) = \int_B X \, \mathrm{d}P$, $B \in \mathcal{B}$, est absolument continue par rapport à P restreinte à $\mathcal{B}$. Il existe donc un élément $Z \in \mathcal{L}^1(\Omega, \mathcal{B}, P)$ tel que $\mu(B) = \int_B Z \, \mathrm{d}P$. Il est immédiat de vérifier que Z est alors une version de l'espérance conditionnelle $E(X \mid \mathcal{B})$.

Un certain nombre de propriétés découlent immédiatement de la démonstration de la définition VI.2.1.

Proposition VI.2.2. *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, et soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Soient de plus X, Y des variables aléatoires réelles intégrables sur $(\Omega, \mathcal{A}, P)$; alors :*

(i) $E(aX + bY + c \mid \mathcal{B}) = aE(X \mid \mathcal{B}) + bE(Y \mid \mathcal{B}) + c$ p.s.

(ii) Si $X \leq Y$, alors $E(X \mid \mathcal{B}) \leq E(Y \mid \mathcal{B})$ p.s.

(iii) Si X_n converge p.s. vers X en croissant, alors $E(X_n \mid \mathcal{B})$ converge p.s. et en croissant vers $E(X \mid \mathcal{B})$.

(iv) Si $\phi : \mathbb{R} \to \mathbb{R}$ est convexe et $\phi(X)$ est intégrable, on a l'inégalité de Jensen : $\phi(E(X \mid \mathcal{B})) \leq E(\phi(X) \mid \mathcal{B})$ p.s. En particulier, $|E(X \mid \mathcal{B})| \leq E(|X| \mid \mathcal{B})$ et $(E(X \mid \mathcal{B}))^2 \leq E(X^2 \mid \mathcal{B})$ p.s.

(v) Si $\mathcal{B} = \{\, \Omega, \emptyset \,\}$, $E(X \mid \mathcal{B}) = E(X)$ p.s.

(vi) Si $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$, $E(E(X \mid \mathcal{B}) \mid \mathcal{C}) = E(X \mid \mathcal{C})$. (Le conditionnement successif $E(E(X \mid \mathcal{B}) \mid \mathcal{C})$ sera noté par la suite $E(X \mid \mathcal{B} \mid \mathcal{C})$.)

(vii) $E(E(X \mid \mathcal{B})) = E(X)$.

(viii) Si $\mathcal{B}$ est indépendante de $\sigma(X)$, $E(X \mid \mathcal{B}) = E(X)$ p.s.

(ix) Si Y est $\mathcal{B}$-mesurable et XY est intégrable, $E(XY \mid \mathcal{B}) = YE(X \mid \mathcal{B})$.

(x) Si X est de carré intégrable, $E(X \mid \mathcal{B})$ est la projection orthogonale de X sur le sous espace $\mathrm{L}^2(\Omega, \mathcal{B}, P)$ dans l'espace de Hilbert $\mathrm{L}^2(\Omega, \mathcal{A}, P)$.

Démonstration. (i) vient essentiellement de la linéarité de la projection Q dans la démonstration de VI.2.1.
(ii) a été démontré dans la démonstration de VI.2.1 : si $X \geq 0$, $E(X \mid \mathcal{B}) \geq 0$ p.s. en prenant $B = \{\, E(X \mid \mathcal{B}) < 0 \,\}$ dans VI.2.1.ii.
(iii) vient de la construction dans la démonstration de VI.2.1.
(iv) se démontre comme l'inégalité de Jensen II.2.10 en utilisant (ii).
(v) vient de VI.2.1.ii.
(vi) vient de ce que $\mathrm{L}^2(\Omega, \mathcal{C}, P) \subset \mathrm{L}^2(\Omega, \mathcal{B}, P) \subset \mathrm{L}^2(\Omega, \mathcal{A}, P)$, et que pour projeter sur $\mathrm{L}^2(\Omega, \mathcal{C}, P)$, on peut commencer par projeter sur $\mathrm{L}^2(\Omega, \mathcal{B}, P)$.
(vii) Prendre $B = \Omega$ dans VI.2.1.ii.
(viii) Si $B \in \mathcal{B}$, $\mathbb{1}_B$ et X sont indépendantes et donc pour tout $B \in \mathcal{B}$,

$$\int_B E(X \mid \mathcal{B}) \, \mathrm{d}P = \int \mathbb{1}_B X \, \mathrm{d}P = E(X) \, P(B) \, .$$

Puisque $E(X \mid \mathcal{B})$ est $\mathcal{B}$-mesurable, $E(X \mid \mathcal{B}) = E(X)$ p.s.
(ix) Le résultat est clair si $Y = \mathbb{1}_B$, $B \in \mathcal{B}$, et donc pour les variables aléatoires étagées. Suivant le schéma général de l'intégration, on approxime ensuite les variables positives par des variables étagées, puis on décompose parties positive et négative.

(x) Soit Z une variable $\mathcal{B}$-mesurable. Introduisons les variables aléatoires $U = X - E(X \mid \mathcal{B})$ et $V = E(X \mid \mathcal{B}) - Z$. Alors V est $\mathcal{B}$-mesurable et $E(U \mid \mathcal{B}) = 0$ d'après (i) et (vi) (avec $\mathcal{B} = \mathcal{C}$ pour montrer que $E(E(X \mid \mathcal{B}) \mid \mathcal{B}) = E(X \mid \mathcal{B})$). Donc, en utilisant (vii)

$$\begin{aligned} E\big((X-Z)^2\big) &= E\Big(E\big((U+V)^2 \mid \mathcal{B}\big)\Big) \\ &= E\big(E(U^2 \mid \mathcal{B}) + 2E(U \mid \mathcal{B})V + V^2\big) \\ &= E\big(E(U^2 \mid \mathcal{B})\big) + E(V^2)\,. \end{aligned}$$

Ainsi, $E((X-Z)^2)$ est minimal lorsque $E(V^2) = 0$, c'est-à-dire $V = 0$ p.s. et donc $E(X \mid \mathcal{B}) = Z$ p.s. Autrement dit, $\|X - Z\|_2$ est minimal pour $Z = E(X \mid \mathcal{B})$, ce qui est la définition de la projection orthogonale. □

***Notation* VI.2.3.** *Si $\mathcal{B} = \sigma(Y)$ est la tribu engendrée par Y, on note $E(X \mid Y)$ pour $E(X \mid \mathcal{B})$.*

Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire intégrable, $E(X \mid \mathcal{B})$ est le vecteur $(E(X_1 \mid \mathcal{B}), \dots, E(X_d \mid \mathcal{B}))$.

Si $X = \mathbb{1}_A$, on note $P(A \mid \mathcal{B}) = E(\mathbb{1}_A \mid \mathcal{B})$.

Par construction, la notation $P(A \mid \mathcal{B}) = E(\mathbb{1}_A \mid \mathcal{B})$ est compatible avec la définition $P(A \mid \mathcal{B})$ que nous avons donnée dans le cas d'un conditionnement discret.

VI.3. Lois conditionnelles

Le principe de conditionnement s'étend des espérances aux lois. Cette extension s'appuie sur le résultat suivant connu sous le nom de lemme de Doob.

***Lemme* VI.3.1 (de Doob).** *Soit Y une variable aléatoire réelle sur $(\Omega, \mathcal{A})$ et soit $X : \Omega \to \mathbb{R}$. Pour que X soit mesurable par rapport à $\sigma(Y)$ (et la tribu borélienne), il faut et il suffit qu'il existe une application borélienne $h : \mathbb{R} \to \mathbb{R}$, telle que $X = h(Y)$.*

Démonstration. Si $X = h(Y)$ avec h borélienne, alors X est $\sigma(Y)$-mesurable.

Réciproquement, il suffit de démontrer le résultat pour une variable aléatoire X positive ou nulle (écrire $X = X^+ - X^-$). D'après la proposition I.2.7, X est limite croissante d'une suite de variables aléatoires étagées

$\sigma(Y)$-mesurables. D'après la définition de la tribu $\sigma(Y)$, une variable aléatoire étagée $\sigma(Y)$-mesurable est de la forme

$$\sum_i a_i \mathbb{1}_{Y^{-1}(B_i)} = \sum_i a_i \mathbb{1}_{B_i} \circ Y ,$$

où la somme est finie, les B_i sont des boréliens et les $a_i \geq 0$. Elle s'écrit donc $h(Y)$ où $h : \mathbb{R} \to \mathbb{R}$ est la fonction borélienne $\sum_i a_i \mathbb{1}_{B_i}$. Il existe donc une suite $(h_n)_{n\in\mathbb{N}}$ de fonctions boréliennes (étagées, positives) telle que $X = \lim_{n\to\infty} h_n(Y)$. En particulier, la suite $(h_n)_{n\in\mathbb{N}}$ converge en tout point de $Y(\Omega)$, l'image de Y. Poser alors (par exemple) $h = \limsup_{n\to\infty} h_n$. La fonction $h : \mathbb{R} \to \mathbb{R}$ est borélienne et $X = h(Y)$. ☐

Soit à présent un couple (X, Y) de variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$ tel que X soit intégrable. L'espérance conditionnelle $E(X \mid Y)$ est $\sigma(Y)$-mesurable. Ainsi, par le lemme VI.3.1, il existe une fonction borélienne h telle que $E(X \mid Y) = h(Y)$. On conviendra d'appeler $h(y)$, $y \in \mathbb{R}$, l'espérance conditionnelle de X sachant $Y = y$, notée $h(y) = E(X \mid Y = y)$. On notera le caractère abusif de cette notation puisque $P\{ Y = y \}$ peut être nul.

***Exemples* VI.3.2.** (i) Si Y prend un nombre fini ou dénombrable de valeurs y_i, $i \in I$, d'après VI.1.9,

$$E(X \mid Y) = \sum_{i\in I^*} \frac{1}{P\{ Y = y_i \}} \int_{\{Y=y_i\}} X \, \mathrm{d}P$$

où $I^* = \big\{ i \in I : P\{Y = y_i\} > 0 \big\}$. Ainsi, si $i \in I^*$,

$$E(X \mid Y = y_i) = \frac{1}{P\{ Y = y_i \}} \int_{\{Y=y_i\}} X \, \mathrm{d}P = \int_\Omega X \, \mathrm{d}P(\,\cdot \mid Y = y_i)$$

où $P(\,\cdot \mid Y = y_i)$ est la probabilité conditionnelle sachant $\{ Y = y_i \}$.

(ii) Supposons que la loi du couple $(X, Y) \in \mathbb{R}^2$ ait une densité $f(x, y)$ par rapport à la mesure de Lebesgue. Vérifions que l'on peut choisir

$$h(y) = E(X \mid Y = y) = \frac{\int_\mathbb{R} x f(x, y) \, \mathrm{d}x}{\int_\mathbb{R} f(x, y) \, \mathrm{d}x} ,$$

lorsque $\int f(x,y)\,\mathrm{d}y > 0$. Soit, à cet effet, C un borélien et $B = Y^{-1}(C)$. Alors, puisque la loi de Y a pour densité $\int_{\mathbb{R}} f(x,y)\,\mathrm{d}x$,

$$\begin{aligned}
\int_{Y^{-1}(C)} h(Y)\ \mathrm{d}P &= \int_{\{y\in C\}} h(y) \left(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}x\right) \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\{y\in C\}} x f(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{\Omega} \mathbb{1}_C(Y) X\,\mathrm{d}P \\
&= \int_{Y^{-1}(C)} X\,\mathrm{d}P\,.
\end{aligned}$$

Donc $h(Y)$ vérifie VI.2.1.i–ii, et par unicité, $h(Y) = E(X \mid Y)$ p.s.

(iii) Les deux exemples précédents peuvent être en fait approfondis quant à l'expression des lois. En pratique, ceci permet de ramener le calcul des lois conditionnelles à un calcul d'intégrales. En remplaçant, dans (i), X par $\phi(X)$, où ϕ est borélienne bornée, l'on voit que si $P\{\, Y = y_i \,\} > 0$,

$$E\big(\phi(X) \bigm| Y = y_i\big) = \int_{\Omega} \phi(X)\,\mathrm{d}P(\,\cdot \mid Y = y_i)\,.$$

Ainsi, d'après la formule du transport, la mesure image $P(\,\cdot \mid Y = y_i)$ peut s'interpréter comme la loi de X « conditionnellement à $Y = y_i$ ». Pour tout borélien B, on a

$$P(\,\cdot \mid Y = y_i)^X(B) = P\{\, X \in B \mid Y = y_i \,\}\,.$$

De la même façon, pour l'exemple VI.3.2.ii, si ϕ est borélienne bornée,

$$E(\phi(X) \mid Y) = \frac{\int \phi(x) f(x,Y)\,\mathrm{d}x}{\int f(x,Y)\,\mathrm{d}x} = \int \phi(x) K^Y(\,\mathrm{d}x)$$

où $K^y(dx) = \dfrac{f(x,y)\,\mathrm{d}x}{\int f(x,y)\,\mathrm{d}x}$ s'interprète comme la loi conditionnelle de X sachant $Y = y$. Il s'ensuit que la densité conditionnelle de X sachant $Y = y$ est $f(x,y)/f^Y(y)$ où f^Y est la densité de Y. Cette formule permet le calcul pratique des lois conditionnelles.

Ces exemples conduisent à la définition suivante.

Définition VI.3.3. On appelle transition, ou noyau de transition, toute fonction $K : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ telle que

(i) pour tout $B \in \mathcal{B}(\mathbb{R})$, $y \mapsto K(y,B) = K^y(B)$ est mesurable ;

(ii) pour tout $y \in \mathbb{R}$, $B \mapsto K(y,B) = K^y(B)$ est une mesure de probabilité.

Le théorème suivant fournit l'existence d'un noyau de transition d'un couple de variables aléatoires réelles. Il généralise les exemples précédents.

Théorème VI.3.4. *Soit (X,Y) un vecteur aléatoire dans $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)_P)$, de loi P, où $\mathcal{B}(\mathbb{R}^2)_P$ est la tribu borélienne de $\mathbb{R}^2$, P-complétée (cf. exercice I.8). Il existe un noyau de transition K tel que pour toute fonction borélienne bornée ϕ,*

$$E\big(\phi(X) \mid Y\big) = \int \phi \, \mathrm{d}K^Y \quad p.s.$$

La mesure $K^y(dx)$ est appelée la loi conditionnelle de X sachant Y, ou sachant $Y = y$. On note aussi $\mathcal{L}(X \mid Y)$ ou $\mathcal{L}(X \mid Y = y)$ cette loi conditionnelle.

Démonstration. (Esquissée) La démonstration est dans le même esprit que celle du théorème V.4.4. Observons que pour toute fonction ϕ continue bornée, la variable aléatoire $E(\phi(X) \mid Y)$ est définie p.s. par VI.2.1 et VI.3.1, c'est-à-dire sauf sur un ensemble de mesure nulle dépendant a priori de ϕ, noté $N(\phi)$. Soit$(\phi_i)_{i\in\mathbb{N}}$ une famille dense dans $(\mathrm{C}_0(\mathbb{R}), \|.\|_\infty)$. Alors $N = \bigcup_{i\in\mathbb{N}} N(\phi_i)$ est de mesure nulle, et $K^{Y(\omega)}(\phi_i)(\omega) = E(\phi_i(X) \mid Y)(\omega)$ est défini sur $\Omega \setminus N$ pour tout $i \in \mathbb{N}$.

Soit maintenant $\phi \in \mathrm{C}_0(\mathbb{R})$. Pour définir $K^{Y(\omega)}(\phi)$, on considère une sous-suite (dépendant de ϕ) $(i_k)_{k\in\mathbb{N}}$ telle que $\lim_{k\to\infty} \|\phi - \phi_{i_k}\|_\infty = 0$. On pose alors $K^{Y(\omega)}(\phi) = \lim_{k\to\infty} K^{Y(\omega)}(\phi_{i_k})$. On vérifie que la limite des $K^{Y(\omega)}(\phi_{i_k})$ ne dépend pas de la sous-suite i_k choisie, mais seulement de ϕ, puisque

$$\Big| E\big((\phi_i - \phi_j)(X) \bigm| Y\big)\Big|(\omega) \leq E\big(|\phi_i - \phi_j|(X) \bigm| Y\big)(\omega) \leq \|\phi_i - \phi_j\|_\infty \text{ p.s.}$$

On définit ainsi pour tout $\omega \in \Omega \setminus N$ une forme linéaire continue sur $\mathrm{C}_0(\mathbb{R})$, laquelle peut être identifiée à une mesure de probabilité. On pourra se référer à Dudley (1989) pour une démonstration complète. □

Le noyau K dépend évidemment de la loi du couple (X,Y). Réciproquement, la loi du couple peut être obtenue à partir de K et de la loi de Y : si ϕ et ψ sont deux fonctions boréliennes bornées, par les points (vii) et (ix) de la proposition VI.2.2,

$$\begin{aligned} E\big(\psi(Y)\phi(X)\big) &= E\big(E(\psi(Y)\phi(X) \mid Y)\big) \\ &= E\big(\psi(Y)E(\phi(X) \mid Y)\big) \\ &= E\bigg(\psi(Y)\int \phi(x)K(Y, \mathrm{d}x)\bigg) . \end{aligned}$$

Testons à présent notre compréhension des lois conditionnelles sur quelques situations simples.

Exemples VI.3.5. (i) Soit X une variable aléatoire et $h : \mathbb{R} \to \mathbb{R}$ mesurable ; quelle est la loi conditionnelle de $h(X)$ sachant $X = x$? Pour toute fonction ϕ borélienne bornée, $E(\phi(h(X)) \mid X) = \phi(h(X))$, et $\phi(h(X))$ est aussi l'intégrale de ϕ contre la masse de Dirac en $h(X)$. Il s'ensuit que $\mathcal{L}(h(X) \mid X = x) = \delta_{h(x)}$.
(ii) Soit un couple (X, Y) de variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$, indépendantes ; soit également h une fonction mesurable de $\mathbb{R}^2$ dans $\mathbb{R}$. D'après le théorème de Fubini et l'indépendance de X et Y, pour toute fonction ϕ borélienne bornée,

$$E\big(\phi \circ h(X, Y) \bigm| Y\big) = \int_{\mathbb{R}} \phi \circ h(x, Y)\,\mathrm{d}P^X(x) .$$

Si $K^y(\cdot)$ désigne la loi de $h(X, y)$,

$$\int_{\mathbb{R}} \phi(u)K^y(\,\mathrm{d}u) = E\big(\phi \circ h(X, y)\big) = \int_{\mathbb{R}} \phi \circ h(x, y)\,\mathrm{d}P^X(x) .$$

Ainsi, si X et Y sont indépendantes, la loi conditionnelle de $h(X, Y)$ sachant $Y = y$ est la loi de $h(X, y)$. Il est aisé de constater sur un exemple que tel n'est plus le cas sans l'hypothèse d'indépendance.
(iii) Soit un couple (X, Y) de variables aléatoires réelles défini sur $(\Omega, \mathcal{A}, P)$, de loi admettant une densité $f(x, y)$ sur $\mathbb{R}^2$. L'exemple VI.3.2.iii et le théorème VI.3.4 montre que la loi de X sachant Y admet une densité donnée par

$$f^{X|Y}(x) = \frac{f(x, Y)}{f^Y(Y)} = \frac{f(x, Y)}{\int_{\mathbb{R}} f(u, Y)\,\mathrm{d}u} .$$

À l'image de la théorie usuelle de l'intégration et des lois, la classe des fonctions boréliennes bornées ϕ qui déterminent une loi conditionnelle dans le théorème VI.3.4 peut être considérablement restreinte. Il suffit par exemple de ne considérer que les exponentielles complexes (*cf.* théorème III.5.2) (fonctions caractéristiques).

Le paragraphe suivant décrit d'autres exemples de calculs d'espérances et de lois conditionnelles de variables gaussiennes. Il y est fait implicitement usage des conditions, des énoncés et des propriétés précédentes relatives à des vecteurs aléatoires de $\mathbb{R}^d$.

VI.4. Espérances conditionnelles dans les espaces gaussiens

Nous terminons ce chapitre par un exemple d'application de calcul d'espérance conditionnelle pour les vecteurs aléatoires gaussiens. Dans le cas gaussien, ces calculs sont relativement explicites. Il est commode de présenter cette application avec la notion d'espace gaussien.

***Définition* VI.4.1.** Un sous-espace vectoriel H de $\mathrm{L}^2(\Omega, \mathcal{A}, P)$ est dit gaussien si pour tous $X_1, \ldots, X_n \in H$, le vecteur $(X_1, \ldots, X_n)$ est gaussien (autrement dit, si pour tous $X_1, \ldots, X_n \in H$ et tous $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, la variable aléatoire réelle $\sum_{1 \leq k \leq n} \alpha_k X_k$ est gaussienne).

De plus, on dit que H est centré si toutes les variables de H sont centrées.

***Proposition* VI.4.2.** *Soit H un espace gaussien, et soit $\overline{H}$ sa fermeture dans $\mathrm{L}^2(\Omega, \mathcal{A}, P)$. Alors $\overline{H}$ est encore gaussien.*

Démonstration. (Démonstration dans le cas centré.) Il suffit de remarquer que si X_n est de loi $\mathcal{N}(0, \sigma_n^2)$ et converge dans L^2 vers X, alors σ_n^2 converge vers $E(X^2)$, et on voit sur les transformées de Fourier que X suit une loi $\mathcal{N}(0, E(X^2))$. □

Par convention, on ne s'intéresse plus désormais qu'aux espaces gaussiens fermés. Pour plus de simplicité, nous les supposerons aussi toujours centrés ; le cas général s'en déduit trivialement, puisque si H est gaussien, alors $H_0 = \{ X - EX : X \in H \}$ est un espace gaussien centré.

Si $X = (X_1, \ldots, X_n)$ est un vecteur aléatoire gaussien à valeurs dans $\mathbb{R}^n$ sur $(\Omega, \mathcal{A}, P)$, X engendre l'espace gaussien (fermé) $\{ \sum_{1 \leq k \leq n} \alpha_k X_k : \alpha_k \in \mathbb{R} \}$. C'est l'exemple canonique qu'il convient de garder à l'esprit.

Le théorème suivant décrit les propriétés d'indépendance dans les espaces gaussiens.

***Théorème* VI.4.3.** *Soient H un espace gaussien (fermé, centré) et H_1 un sous-espace (fermé) de H. Soit $X \in H$ fixé. Les propriétés suivantes sont équivalentes :*

(i) pour tout Y de H_1, $E(XY) = 0$;

(ii) pour tout Y de H_1, X est indépendante de Y ;

(iii) X est indépendante de la tribu $\sigma(H_1)$ engendrée par les variables de H_1.

Démonstration. Clairement (iii)$\Rightarrow$(ii)$\Rightarrow$(i). Pour montrer (i)$\Rightarrow$(ii), on note que (i) implique que le couple (X,Y) est gaussien, de matrice de covariance diagonale ; donc X et Y sont indépendantes par le théorème IV.4.3. De la même façon, pour (i)$\Rightarrow$(iii), il suffit de montrer que si $Y_1,\ldots,Y_n \in H_1$, le vecteur $(Y_1,\ldots,Y_n)$ est indépendant de X, ce qui se fait de façon identique. En effet, si tel est alors le cas, posons pour tout borélien B,

$$\mathcal{M} = \Big\{ E \in \mathcal{A} : P\big(\{ X \in B \} \cap E\big) = P\{ X \in B \} P(E) \Big\} .$$

L'ensemble $\mathcal{M}$ est une classe monotone qui contient la classe $\mathcal{E}$ des intersections finies de $Y^{-1}(C)$, $Y \in H$, C borélien. Donc $\mathcal{M} \supset \mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \sigma(H_1)$. Donc X est indépendante de $\sigma(H_1)$. □

Le résultat précédent autorise des calculs d'espérances conditionnelles.

Proposition VI.4.4. *Soit H_1 comme précédemment, et soit X un élément de H. On désigne par $\sigma(H_1)$ la tribu engendrée par H_1 (c'est-à-dire la plus petite tribu qui rend tous les éléments de H_1 mesurables). Alors, l'espérance conditionnelle $E(X \mid \sigma(H_1))$ est simplement la projection orthogonale (dans L^2) de X sur H_1. En particulier, c'est une variable gaussienne.*

Démonstration. Soit Y la projection de X sur H_1. Cette projection est $\sigma(H_1)$-mesurable et $X = Y + Z$ où Z est orthogonale à H_1, donc indépendante de $\sigma(H_1)$ (théorème VI.4.3). On écrit alors, par la propriété des espérances conditionnelles VI.2.2.i et VI.2.2.viii,

$$E(X \mid \sigma(H_1)) = E(Y \mid \sigma(H_1)) + E(Z \mid \sigma(H_1)) = Y + E(Z) = Y ,$$

d'où le résultat. □

Comment utiliser ce résultat dans un calcul pratique ? Supposons par exemple que $(X_1,\ldots,X_n)$ soit un vecteur gaussien centré, et soient $i_1,\ldots,i_p < n$. On voudrait calculer $E(X_n \mid X_{i_1},\ldots,X_{i_p})$, c'est-à-dire

$$E(X_n \mid \sigma(X_{i_1},\ldots,X_{i_p})) = E(X_n \mid \sigma(H_1))$$

où H_1 est engendré par $(X_{i_1},\ldots,X_{i_p})$, H étant engendré par $(X_1,\ldots,X_n)$. D'après la proposition VI.4.4, cette espérance conditionnelle est un élément de H_1 et donc

$$E(X_n \mid X_{i_1},\ldots,X_{i_p}) = \sum_{1\leq j\leq p} \alpha_j X_{i_j}$$

pour des coefficients réels $\alpha_1, \ldots, \alpha_p$ qu'il convient de calculer. À cet effet, on peut par exemple commencer par multiplier cette identité par X_{i_1}, et intégrer, pour obtenir

$$E(X_n X_{i_1}) = \sum_{1 \leq j \leq p} \alpha_j E(X_{i_j} X_{i_1})$$

et ainsi de suite avec $X_{i_2}, \ldots, X_{i_p}$. La donnée de la covariance du vecteur $(X_1, \ldots, X_n)$ permet ensuite de résoudre le système linéaire de p équations à p inconnues $\alpha_1, \ldots, \alpha_p$.

***Exemples* VI.4.5.** (i) Soit (X, Y, Z) un vecteur gaussien centré de matrice de covariance

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 5 & 3 \\ -1 & 3 & 4 \end{pmatrix}.$$

Calculons $E(Y \mid X, Z)$. D'après ce qui précède, cette espérance conditionnelle est de la forme $\alpha X + \beta Z$. Les égalités

$$\begin{aligned} E(XY) &= \alpha E(X^2) + \beta E(XZ) \\ E(YZ) &= \alpha E(XZ) + \beta E(Z^2), \end{aligned}$$

conduisent au système

$$\begin{cases} 0 = \alpha - \beta \\ 3 = -\alpha + 4\beta. \end{cases}$$

Il vient $\alpha = 1$, $\beta = 1$ et donc $E(Y \mid X, Z) = X + Z$.

(ii) Soit (X, Y) un couple gaussien centré de matrice de covariance $\begin{pmatrix} 4/3 & -1 \\ -1 & 1 \end{pmatrix}$. Calculons $E(X \mid Y - X)$. On a $E(X \mid Y - X) = \alpha(Y - X)$ et

$$E\big(X(Y - X)\big) = E\Big(E\big(X(Y - X) \mid Y - X\big)\Big) = \alpha E\big((Y - X)^2\big),$$

d'où $-7/3 = \alpha 11/3$ et $E(X \mid Y - X) = -\frac{7}{11}(Y - X)$, qui est une variable gaussienne centrée de variance $49/33$.

Le calcul des espérances conditionnelles gaussiennes est en un certain sens suffisant pour la connaissance plus précise des lois conditionnelles. Soit $(Z_1, \ldots, Z_n)$ un vecteur aléatoire gaussien, centré, et soient pour $1 \leq k \leq n$, $X = (Z_1, \ldots, Z_k)$ et $Y = (Z_{k+1}, \ldots, Z_n)$. On s'intéresse à la loi conditionnelle de X sachant que $Y = y$ $(\in \mathbb{R}^{n-k})$. Pour la déterminer, il nous suffit de connaître sa transformée de

Fourier conditionnelle $E(e^{i\langle t,X\rangle} \mid Y)$, $t \in \mathbb{R}^k$. Or, puisque $\langle t,X\rangle - E(\langle t,X\rangle \mid Y)$ et Y sont orthogonales et donc indépendantes,

$$\begin{aligned}
E(e^{i\langle t,X\rangle} \mid Y) &= e^{iE(\langle t,X\rangle|Y)} E\Big(e^{i(\langle t,X\rangle - E(\langle t,X\rangle|Y))} \,\Big|\, Y\Big) \\
&= e^{iE(\langle t,X\rangle|Y)} E\Big(e^{i(\langle t,X\rangle - E(\langle t,X\rangle|Y))}\Big) \\
&= \exp\Big(iE\big(\langle t,X\rangle \bigm| Y\big) - \frac{1}{2}E\big((\langle t,X\rangle - E(\langle t,X\rangle \mid Y))^2\big)\Big) .
\end{aligned}$$

Ainsi la loi conditionnelle de X sachant $Y = y$ est une loi gaussienne de moyenne $E(X \mid Y = y)$ (vecteur dans $\mathbb{R}^k$) et de matrice de covariance $(k \times k)$

$$E\big((X_i - E(X_i \mid Y))(X_j - E(X_j \mid Y))\big) , \qquad 1 \leq i,j \leq k .$$

Ceci explique qu'il suffit de calculer des espérances conditionnelles pour connaître les lois conditionnelles gaussiennes.

On peut également travailler directement sur les densités. Soit par exemple (X,Y) un couple gaussien centré sur $\mathbb{R}^2$, de matrice de covariance

$$\Gamma = \begin{pmatrix} a & c \\ c & b \end{pmatrix} .$$

La loi conditionnelle de X sachant $Y = y$ est donnée par la densité (*cf.* exemple VI.3.2.iii) $f(x,y)/\int f(x,y)\,\mathrm{d}x$, où

$$\begin{aligned}
f(x,y) &= \frac{1}{2\pi}\frac{1}{\sqrt{\gamma}} \exp\Big(-\frac{1}{2}\Big\langle \Gamma^{-1}\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix}\Big\rangle\Big) \\
&= \frac{1}{2\pi}\frac{1}{\sqrt{\gamma}} \exp\Big(-\frac{1}{2\gamma}(bx^2 - 2cxy + ay^2)\Big)
\end{aligned}$$

avec $\gamma = \mathrm{dét}\Gamma = ab - c^2 > 0$. On a

$$\begin{aligned}
\int f(x,y)\,\mathrm{d}x &= \frac{1}{2\pi}\frac{1}{\sqrt{\gamma}} \exp\Big(-\frac{1}{2\gamma}\Big(a - \frac{c^2}{b}\Big)y^2\Big) \int \exp\Big(-\frac{b}{2\gamma}\Big(x - \frac{c}{b}\Big)^2\Big)\,\mathrm{d}x \\
&= \frac{1}{2\pi}\frac{1}{\sqrt{\gamma}} \exp\Big(-\frac{1}{2\gamma}\Big(a - \frac{c^2}{b}\Big)y^2\Big) .
\end{aligned}$$

Ainsi

$$\frac{f(x,y)}{\int f(x,y)\,\mathrm{d}x} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{b}{\gamma}} \exp\Big(-\frac{b}{2\gamma}\Big(x - \frac{c}{b}y\Big)^2\Big) ,$$

de sorte que la loi de X conditionnelle à $Y = y$ est $\mathcal{N}(m,\sigma^2)$ avec $m = (c/b)y$ et $\sigma^2 = \gamma/b$. On notera que, comme précédemment, σ ne dépend pas de y. En fait, on retrouve le résultat obtenu précédemment sous une autre méthode. En effet, $E(X \mid Y) = \alpha Y$ où α est tel que $E(XY) = \alpha E(Y^2)$, et donc $\alpha = c/b$; ainsi, $\sigma^2 = E((X - E(X \mid Y))^2) = E((X - (c/b)Y)^2) = \gamma/b$. À noter que si $c = 0$, X et Y sont indépendantes et $E(X \mid Y) = E(X) = 0$ (puisque X est centrée). Le cas $b = 0$ (et donc $c = 0$) est trivial.

Exercices

Exercice VI.1. Soient X et Y des variables aléatoires indépendantes, de même loi, intégrables. Comparer les lois des couples $(X, X+Y)$ et $(Y, X+Y)$. En déduire que $E(X \mid X+Y) = E(Y \mid X+Y) = (X+Y)/2$.

Exercice VI.2. X_1 et X_2 étant les résultats indépendants de deux jets de dés, et S étant leur somme, quelle est la loi de X_1 sachant que S est paire ?

Exercice VI.3. Soit X une variable aléatoire réelle quelconque, et soit a une constante réelle. Déterminer la loi de X conditionnée par $X \wedge a$.

Exercice VI.4. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, telle que pour tous $m, n \in \mathbb{N}$,

$$P\{X \geq m+n \mid X \geq m\} = P\{X \geq n\}$$

(on dit que X est sans mémoire).
a) On pose $P\{X = 0\} = a$. Déterminer la loi de X.
b) Soit Y une copie indépendante de X. Quelle est la loi de $S = X + Y$? Déterminer la loi conditionnelle de X sachant $S = p$, $p \in \mathbb{N}$. Interpréter le résultat.

Exercice VI.5. Soit $X = (X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Soit N une variable aléatoire à valeurs dans $\mathbb{N}$, indépendante de la suite X. Montrer que X_N est une variable aléatoire. Montrer que pour tout $k \in \mathbb{N}$, la loi de X_N sachant $N = k$ est la loi de X_k.

Exercice VI.6. Soient $X_1, \ldots, X_p$ des variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs $\lambda_1, \ldots, \lambda_p$. Déterminer la loi conditionnelle du vecteur aléatoire $(X_1, \ldots, X_p)$ sachant que $\sum_{1 \leq i \leq p} X_i = n$.

Exercice VI.7. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes suivant chacune la loi $\mathcal{N}(0,1)$. Démontrer que la loi de X_1 sachant $S_n = \sum_{1 \leq i \leq n} X_i$ est la loi $\mathcal{N}(S_n/n, 1 - 1/n)$.

Exercice VI.8. Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\theta > 0$. Établir que

$$P\{X \geq t+s \mid X > t\} = P\{X > s\}, \quad s, t \geq 0.$$

Montrer que cette propriété caractérise la loi exponentielle parmi les lois à densité. Prouver que $\lim_{h\to 0} h^{-1}P\{\, t < X < t+h \mid X > t\,\} = \theta$ pour tout t.

Exercice VI.9. Soient X et Y deux variables aléatoires réelles indépendantes de loi $\mathcal{N}(0,1)$. On pose $X = R\cos\theta$ et $Y = R\sin\theta$.

a) Montrer que $X+Y$ et $X-Y$ sont indépendantes et en déduire la loi de R^2 sachant que $Y = X$.

Indication : on pourra écrire $R^2 = \frac{1}{2}((X+Y)^2 + (X-Y)^2)$.

b) Montrer que R et θ sont indépendantes et en déduire la loi de R^2 sachant que $\theta = \pi/4$ ou $5\pi/4$ (c'est-à-dire sachant que $Y = X$).

c) Pour montrer que les résultats ne sont pas contradictoires, préciser les sous-tribus de conditionnement dans les deux questions.

Exercice VI.10. On se donne une matrice carrée $\mathbb{P} = (P_{i,j})_{1\le i,j\le n}$. Déterminer à quelle condition sur $\mathbb{P}$ il existe des variables aléatoires X et Y à valeurs dans $\{\,1,\dots,n\,\}$ telles que

$$P_{i,j} = P\{\, Y = j \mid X = i\,\}\,, \qquad i,j = 1,\dots,n\,.$$

On appellera une telle matrice, matrice de transition (voir chapitre VIII).

$\mathbb{P}$ étant une matrice de transition (loi conditionnelle de Y sachant X), on désigne par M le vecteur de $\mathbb{R}^n$ représentant la loi de X : $M_i = P\{\, X = i\,\}$, $i = 1,\dots,n$. Démontrer que la loi de Y se représente par le vecteur ${}^t\mathbb{P}M$.

Exercice VI.11. Nous avons vu à l'exercice V.14 comment construire une suite infinie de variables aléatoires indépendantes sur l'espace probabilisé $([\,0,1\,], \mathcal{B}([\,0,1\,]), \lambda)$. À l'aide de l'exercice V.14, construire sur cet espace une suite de vecteurs aléatoires indépendants de loi P_i, $i \in \mathbb{N}$, données sur $\mathbb{R}^2$.

Exercice VI.12. Soit P une loi sur $\mathbb{R}^2$, de marges P^X et P^Y, et (X,Y) de loi P. Soit $F^{X|y}(x)$ la fonction de répartition de la loi conditionnelle $\mathcal{L}(X \mid Y = y)$. Soient U, V deux variables aléatoires indépendantes et de loi uniforme sur $[\,0,1\,]$. Montrer que le couple $(F^{Y\leftarrow}(U), F^{X|F^{Y\leftarrow}(U)}(V))$ est de loi P. Ceci donne un procédé de simulation d'un vecteur aléatoire.

Exercice VI.13. On reprend les notations de l'exercice IV.13. Montrer que

$$P\{\, S_{i+1,n} \ge s \mid X_{i,n} = x\,\} = \Big(\frac{1 - F(x+s)}{1 - F(x)}\Big)^{n-i}\,, \quad x \in \mathbb{R}\,,\ s \ge 0\,,$$

et que

$$P\{\, S_{i+1,n} \ge s \mid X_{i+1,n} = x\,\} = \Big(\frac{F(x-s)}{F(x)}\Big)^{i}\,, \quad x \in \mathbb{R}\,,\ s \ge 0\,.$$

***Exercice* VI.14.** Soient $X_1, \ldots, X_n$ des variables aléatoires réelles, indépendantes et de même loi admettant une densité f. Soit $X_{1,n} \leq \cdots \leq X_{n,n}$ ces variables aléatoires ordonnées, et définissons les espacements $S_{i,n} = X_{i,n} - X_{i-1,n}$, $2 \leq i \leq n$, qui mesurent les distances entre les variables adjacentes (faire un dessin). Soit

$$L_n(x) = \frac{1}{n-1} \sum_{2 \leq i \leq n} \mathbb{1}_{[0,x]}(nS_{i,n})$$

la fonction de répartition empirique des espacements, laquelle compte la proportion d'espacements plus petits que x/n. Notons

$$L(x) = 1 - \int_{\mathbb{R}} f(z) e^{-xf(z)} \, \mathrm{d}z \, .$$

Soit enfin $I_{i,n} = 1$, si aucune des variables $X_1, \ldots, X_n$ ne tombe dans l'intervalle $]\, X_i, X_i + x/n\,]$ et $I_{i,n} = 0$ sinon.

a) Montrer que le vecteur $(I_{1,n}, \ldots, I_{n,n})$ est échangeable, c'est-à-dire que sa loi est invariante par permutation des coordonnées (voir aussi exercice III.8).

b) Montrer que $1 - L_n(x) = (n-1)^{-1} \sum_{1 \leq i \leq n} I_{i,n}$.

c) Montrer que $I_{i,n}$ suit une loi de Bernoulli de paramètre

$$E\Big(\big(1 - F(X_1 + x/n) + F(X_1)\big)^{n-1} \Big) \, .$$

d) Évaluer $P\{\, I_{i,n} = 1 \, ; \, I_{j,n} = 1 \,\}$.

e) Montrer que $\lim_{n\to\infty} E(L_n(x)) = L(x)$ et que $\lim_{n\to\infty} E(L_n(x)^2) = L(x)^2$.
Indication : Penser au théorème de convergence dominée !

En déduire que $L_n(x)$ converge vers $L(x)$ en probabilité

f) En utilisant la continuité, la bornitude et la monotonie de L, montrer que

$$\lim_{n\to\infty} \sup_{x \in \mathbb{R}} |L_n(x) - L(x)| = 0 \qquad \text{en probabilité.}$$

Pour n assez grand, ce résultat donne une idée sur la taille des écarts entre les points aléatoires adjacents $X_{1,n}, \ldots, X_{n,n}$.

g) Soit maintenant h une fonction continue bornée sur $\mathbb{R}$. Observons que L est la fonction de répartition d'une loi Q. Montrer que e) implique

$$\lim_{n\to\infty} \frac{1}{n-1} \sum_{2 \leq i \leq n} h(nS_{i,n}) = \int h \, \mathrm{d}Q \qquad \text{en probabilité.}$$

Indication : Soit Q_n la loi de probabilité de fonction de répartition L_n. Remarquer que $(n-1)^{-1} \sum_{2 \leq i \leq n} h(nS_{i,n}) = \int h \, \mathrm{d}Q_n$, puis utiliser la définition-théorème V.4.1.

Exercice VI.15. La proposition III.2.7 nous donne une façon d'engendrer des variables aléatoires réelles, pourvu que la fonction de quantile soit facile à calculer. Ce n'est pas toujours le cas en pratique. Une méthode assez efficace est la méthode dite du rejet qui fonctionne comme suit.

Soient f, g, deux densités sur $\mathbb{R}$. On souhaite simuler une variable de densité g, en supposant qu'on sache facilement simuler une variable de densité f, et qu'il existe une constante c telle que $g \leq cf$. Soit (X, U) un couple de variables aléatoires indépendantes, respectivement de lois de densité f et uniforme sur $[\,0,1\,]$.

a) Montrer que le couple $(X, cUf(X))$ est uniformément distribué sous le graphe de f

$$\underline{f} = \{\,(x,y) \in \mathbb{R}^2 \,:\, 0 \leq y \leq f(x)\,\}\,;$$

c'est-à-dire qu'en notant λ la mesure de Lebesgue sur $\mathbb{R}^2$,

$$\forall A \in \mathcal{B}(\mathbb{R}^2)\,, \quad P\{\,(X, cUf(X)) \in A\,\} = \lambda(A \cap \underline{f}).$$

Indication : Remarquer que

$$\int \mathbb{1}_A(x, cuf(x)) f(x)\,\mathrm{d}u\,\mathrm{d}x = \int \mathbb{1}_{A_x}\big(cuf(x)\big)\,\mathrm{d}u f(x)\,\mathrm{d}x$$

où A_x est la section de A selon x.

En déduire que $\mathcal{L}(X \mid cUf(X) \leq g(X))$ a pour densité g.

b) Soient (U_i, X_i) des couples indépendants, de même loi que (X, U). Soit $N_0 = 0$ et,

$$N_i = \min\{\,i \geq N_{i-1} \,:\, cU_i f(X_i) \leq g(X_i)\,\}\,, \qquad i \geq 1\,.$$

Montrer que $P\{\,N_1 = k\,\} = (1 - c^{-1})^{k-1} c^{-1}$ et que $E(N_1) = c$. Montrer que X_{N_i}, $i \geq 1$, est une suite de variables aléatoires indépendantes, de lois de densité g. Expliquer pourquoi en pratique il faut prendre c le plus petit possible.

c) Soit maintenant

$$\gamma_p(x) = \Gamma(p)^{-1} x^{p-1} e^{-x}\,, \qquad x \geq 0\,, \quad p \geq 1\,,$$

la densité de la loi Γ_p. Soit $f(x) = e^{-x}$, $x \geq 0$. Expliquer comment simuler des variables aléatoires indépendantes de loi Γ_p à partir d'une suite de variables aléatoires indépendantes et uniformes sur $[\,0,1\,]$.

Exercice VI.16. (Processus de Poisson)

a) On considère une famille de variables aléatoires $(X_1, \ldots, X_n)$, indépendantes et uniformément distribuées sur $[\,0,t\,]$. On note $X_{1,n} \leq \cdots \leq X_{n,n}$ la famille réarrangée dans l'ordre croissant. On dit alors que $(X_{1,n} \leq \cdots \leq X_{n,n})$ est une n-statistique d'ordre sur $[\,0,t\,]$. Donner la loi de $(X_{1,n} \leq \cdots \leq X_{n,n})$.

Indication : on pourra introduire les ensembles

$$A_\sigma = \{\, (X_{1,n} \leq \cdots \leq X_{n,n}) = (X_{\sigma(1)} \leq \cdots \leq X_{\sigma(n)}) \,\}$$

pour toute permutation σ à n éléments.
b) Montrer que si $(X_{1,n} \leq \cdots \leq X_{n,n})$ est une n-statistique d'ordre sur $[\,0,t\,]$, alors la loi conditionnelle de $(X_{1,n} \leq \cdots \leq X_{n,n})$ sachant $\{\, X_{n,n} = x \,\}$ a la loi d'une $(n-1)$-statistique d'ordre sur $[\,0,x\,]$.
c) Supposons que $(X_{1,n} \leq \cdots \leq X_{n,n})$ est une n-statistique d'ordre sur $[\,0,t\,]$. Considérons des réels $0 = t_0 \leq t_1 \leq \cdots \leq t_p \leq t$ et des entiers $0 = k_0 \leq k_1 \leq \cdots \leq k_p = n$. Montrer que

$$P\{\, \forall j = 0,\ldots,p-1, \forall i = k_j+1,\ldots,k_{j+1}, x_{i,n} \in]\, t_j, t_{j+1}\,] \,\} = \frac{n!}{t^n} \prod_{0\leq j\leq p-1} \frac{(t_{j+1}-t_j)^{k_{j+1}-k_j}}{(k_{j+1}-k_j)!}\,.$$

Indication : On pourra utiliser a) et comparer le résultat cherché à une loi multinomiale.
d) On considère une suite de variables exponentielles de paramètre λ, indépendantes, $(T_k)_{k\geq 1}$, et on note $S_n = T_1 + \cdots + T_n$, $n \geq 1$. Calculer la loi de $(S_1,\ldots,S_n)$, puis la loi de S_n. Montrer que la loi conditionnelle de $(S_1,\ldots,S_n)$ sachant $S_{n+1} = s$ est la loi d'une n-statistique d'ordre sur $[\,0,s\,]$.
e) On pose $N_t = \sum_n \mathbb{1}_{[0,t]}(S_n)$. Montrer que la variable N_t est finie presque sûrement. En utilisant c) et d), montrer que, pour tous $0 \leq t_1 \leq \cdots \leq t_n$, pour tous entiers $k_1,\ldots,k_n$, on a

$$P\{\, N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2 \ldots, N_{t_n} - N_{t_{n-1}} = k_n \,\} = \prod_{1\leq i\leq n} \frac{\big(\lambda(t_i - t_{i-1})\big)^{k_i}}{k_i!} \exp\big(-\lambda(t_{i+1}-t_i)\big)\,.$$

En déduire que les variables $N_{t_{i+1}} - N_{t_i}$ sont indépendantes et suivent des lois de Poisson de paramètre $\lambda(t_{i+1} - t_i)$.

VII

MARTINGALES (À TEMPS DISCRET)

La notion de martingale est une notion fondamentale du calcul des probabilités. Elle a son origine en théorie des jeux et introduit le temps dans l'analyse probabiliste. Sa donnée fondamentale est celle d'une famille croissante $(\mathcal{F}_t)_{t\geq 0}$ de tribus représentant l'évolution de l'information avec le temps. Dans cette étude, nous nous contenterons de l'examen de modèles à temps discret.

VII.1. Généralités

Définition VII.1.1. Sur un espace probabilisé $(\Omega, \mathcal{F}, P)$, on appelle filtration toute suite croissante $(\mathcal{F}_n)_{n\in\mathbb{N}}$ de sous-tribus de $\mathcal{F}$ (on pourra prendre pour $\mathcal{F}$ la tribu, notée $\mathcal{F}_\infty$, engendrée par les tribus $\mathcal{F}_n$, $n \in \mathbb{N}$).

Intuitivement, la tribu $\mathcal{F}_n$ contient tous les événements qui peuvent survenir avant l'instant n.

Définition VII.1.2. Une suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ définies sur $(\Omega, \mathcal{A}, P)$ est appelée un processus. De plus, on dit que le processus est adapté à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si pour tout $n \in \mathbb{N}$ la variable aléatoire X_n est $\mathcal{F}_n$-mesurable.

Définition VII.1.3. Soit un processus adapté $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ tel que X_n est intégrable pour tout n. On dit que le processus est

(i) une martingale, si pour tous $0 \le m \le n$,

$$E(X_n \mid \mathcal{F}_m) = X_m \quad p.s.;$$

(ii) une sur-martingale, si pour tous $0 \le m \le n$,

$$E(X_n \mid \mathcal{F}_m) \le X_m \quad p.s.;$$

(iii) une sous-martingale, si pour tous $0 \le m \le n$,

$$E(X_n \mid \mathcal{F}_m) \ge X_m \quad p.s.$$

En particulier, un processus adapté $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale si et seulement si c'est à la fois une sur- et une sous-martingale. C'est une sous-martingale si et seulement si le processus adapté $(-X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une sur-martingale. Par conséquent, nous nous contenterons parfois d'énoncer des résultats pour des sur- ou des sous-martingales.

On voit que $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale (resp. une sur-martingale, sous-martingale) si et seulement si $E(X_n - X_m \mid \mathcal{F}_n) = 0$ (resp. ≤ 0, ≥ 0) pour tous $m \le n$, ce qui équivaut à ce que pour tout $A \in \mathcal{F}_n$, $\int_A (X_n - X_m)\,\mathrm{d}P = 0$ (resp. ≤ 0, ≥ 0.)

Il suffit de vérifier la définition VII.1.3 pour tous n et $m = n-1$. En effet, d'après les propriétés de conditionnements successifs des espérances conditionnelles, si $m < n$,

$$\begin{aligned} E(X_n - X_m \mid \mathcal{F}_m) &= \sum_{m+1\le k\le n} E(X_k - X_{k-1} \mid \mathcal{F}_m) \\ &= \sum_{m+1\le k\le n} E(X_k - X_{k-1} \mid \mathcal{F}_{k-1} \mid \mathcal{F}_m) = 0 \end{aligned}$$

(resp. ≤ 0, ≥ 0).

Observons aussi que si $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale (resp. sur-martingale, resp. sous-martingale), la suite $(E(X_n))_{n\in\mathbb{N}}$ est constante (resp. décroissante, resp. croissante) car $E(X_n) = E(E(X_n \mid \mathcal{F}_{n-1})) = E(X_{n-1})$ (resp. $\le E(X_{n-1})$, resp. $\ge E(X_{n-1})$).

Parfois, nous ne considèrerons que des martingales, des sur-martingales ou des sous-martingales $(X_n, \mathcal{F}_n)_{0\le n\le k}$ indexées sur un nombre fini d'instants. On peut aussi démarrer ces processus à $n = 1$ au lieu de $n = 0$.

Si X_n représente la fortune d'un joueur à l'instant n, dire que (X_n) est une martingale signifie que le jeu est équilibré, au sens où la connaissance des parties passées ne donne pas, en moyenne, d'avantage pour la partie à venir.

Exemples VII.1.4. (i) Soit Z une variable aléatoire intégrable sur $(\Omega, \mathcal{F}, P)$ et $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une filtration de $\mathcal{F}$. Posons $X_n = E(Z \mid \mathcal{F}_n)$, $n \in \mathbb{N}$. Alors $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale.

(ii) Soient Z_n, $n \geq 1$, des variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, P)$, intégrables et de moyenne M (*i.e.* $E(Z_n) = M$). On désigne par $\mathcal{F}_n$ la tribu engendrée par $Z_1, \ldots, Z_n$. La suite $(\mathcal{F}_n)_{n\geq 1}$ est une filtration. Considérons les sommes $X_n = Z_1 + \cdots + Z_n$, $n \geq 1$. Alors $(X_n, \mathcal{F}_n)_{n\geq 1}$ est une martingale (resp. surmartingale, resp. sous-martingale) si $M = 0$ (resp. $M < 0$, resp. $M > 0$). En effet, si $n \geq 2$, par les propriétés des espérances conditionnelles (IV.2.2),

$$\begin{aligned} E(X_n \mid \mathcal{F}_{n-1}) &= E(Z_1 + \cdots + Z_{n-1} + Z_n) \mid \mathcal{F}_{n-1}) \\ &= E(X_{n-1} \mid \mathcal{F}_{n-1}) + E(Z_n \mid \mathcal{F}_{n-1}) \\ &= X_{n-1} + M\,. \end{aligned}$$

(iii) Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une martingale ; soit ϕ une fonction convexe sur $\mathbb{R}$ telle que $\phi(X_n)$ soit intégrable pour tout $n \in \mathbb{N}$. Alors $(\phi(X_n), \mathcal{F}_n)_{n\in\mathbb{N}}$ est une sous-martingale ; en effet, l'inégalité de Jensen VI.2.2.iv fournit, pour $n \geq m$,

$$E\big(\phi(X_n) \bigm| \mathcal{F}_m\big) \geq \phi\big(E(X_n \mid \mathcal{F}_m)\big) = \phi(X_m).$$

Noter en particulier le choix de $\phi(x) = |x|$ ou $\phi(x) = x^2$. Le résultat est bien sûr encore vrai si $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une sous-martingale et si ϕ est en outre croissante.

Par définition, une sous-martingale (resp. sur-martingale) est un processus croissant (resp. décroissant) en moyenne conditionnelle, et donc en moyenne. Le résultat suivant, la décomposition de Doob, nous dit qu'une sous-martingale (resp. sur-martingale) peut toujours être vue comme une martingale à laquelle est ajoutée un processus croissant (resp. décroissant). De plus, ce processus monotone $(Z_n)_{n\in\mathbb{N}}$ peut être pris non seulement adapté à $(\mathcal{F}_n)_{n\in\mathbb{N}}$ mais à $(\mathcal{F}_{n-1})_{n\in\mathbb{N}}$, où l'on convient que $\mathcal{F}_{-1}$ est la tribu triviale $\{\emptyset, \Omega\}$. Autrement dit, idéalement, la valeur de Z_n peut être parfaitement prédite à l'instant $n - 1$.

Théorème VII.1.5 **(décomposition de Doob).** *Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une sous-martingale. Il existe des processus $(Y_n)_{n\in\mathbb{N}}$ et $(Z_n)_{n\in\mathbb{N}}$ uniques presque sûrement, tels que*

(i) $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale ;

(ii) $Z_0 = 0$ *et* Z_n *est* $\mathcal{F}_{n-1}$*-mesurable pour tout* $n \in \mathbb{N}$ *(avec la convention* $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$*) et presque sûrement croissant, i.e.* $Z_n \leq Z_{n+1}$ *p.s. pour tout* $n \in \mathbb{N}$ *;*

(iii) $X_n = Y_n + Z_n$ *pour tout* $n \in \mathbb{N}$.

Démonstration. Une martingale est en moyenne constante. Donc le processus Z_n doit cumuler les sauts de la sous-martingale X_n. Ceci conduit à considérer les différences $\Delta_n = X_n - X_{n-1}$, $n \geq 1$. Soit $Z_0 = 0$, $Y_0 = X_0$ et pour tout $n \geq 1$,

$$Z_n = \sum_{1 \leq i \leq n} E(\Delta_i \mid \mathcal{F}_{i-1}) \quad \text{et} \quad Y_n = X_n - Z_n \,.$$

Le processus Z_n est croissant (car $E(\Delta_n \mid \mathcal{F}_{n-1}) \geq 0$) et Y_n est une $\mathcal{F}_n$-martingale puisque

$$\begin{aligned} E(Y_n \mid \mathcal{F}_{n-1}) &= E(Y_n - Y_{n-1} \mid \mathcal{F}_{n-1}) + Y_{n-1} \\ &= E\big(\Delta_n - (Z_n - Z_{n-1}) \mid \mathcal{F}_{n-1}\big) + Y_{n-1} \\ &= E(\Delta_n \mid \mathcal{F}_{n-1}) - (Z_n - Z_{n-1}) + Y_{n-1} \\ &= Y_{n-1} \end{aligned}$$

(nous avons utilisé la $\mathcal{F}_{n-1}$-mesurabilité de Z_n, et le fait que Z_{n-1} est $\mathcal{F}_{n-2}$-mesurable, donc aussi $\mathcal{F}_{n-1}$-mesurable).

Pour démontrer l'unicité de la décomposition, soit (Y'_n, Z'_n) une autre décomposition vérifiant (i)–(iii). Alors $Z_0 = Z'_0 = 0$ et donc $Y'_0 = Y_0$. Par récurrence, supposons $Z'_j = Z_j$ et $Y'_j = Y_j$ pour tout $0 \leq j \leq n$. Alors

$$\begin{aligned} Z'_{n+1} &= E(Z'_{n+1} \mid \mathcal{F}_n) = E(X_{n+1} - Y'_{n+1} \mid \mathcal{F}_n) = E(X_{n+1} \mid \mathcal{F}_n) - Y'_n \\ &= E(Y_{n+1} + Z_{n+1} \mid \mathcal{F}_n) - Y'_n = Y_n - Y'_n + Z_{n+1} \,. \end{aligned}$$

En utilisant l'hypothèse de récurence, il vient $Z'_{n+1} = Z_{n+1}$ p.s., et donc $Y'_{n+1} = Y_{n+1}$ p.s., ce qui prouve l'unicité de la décomposition. □

Intimement liée à la notion de martingale se trouve être celle de temps d'arrêt.

***Définition* VII.1.6.** Sur $(\Omega, \mathcal{F}, P)$ muni d'une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, une variable aléatoire $T : \Omega \to \mathbb{N} \cup \{\infty\}$ est appelée un temps d'arrêt si l'on a $\{T \leq n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$.

Il est immédiat que l'on pourrait définir un temps d'arrêt T comme étant une variable aléatoire à valeurs dans $\mathbb{N}$ telle que $\{T = n\} \in \mathcal{F}_n$ (puisque $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^{\mathrm{c}}$ et $\{T \leq n-1\} = \bigcup_{1 \leq i \leq n-1} \{T = i\}$).

Néanmoins cette seconde définition ne se généralise pas convenablement au cas des martingales à temps continu (*i.e.* on ne dispose plus d'une suite (X_n) indexée par les entiers mais d'une fonction X_t indexée par $\mathbb{R}$).

Si T est un temps d'arrêt, on définit la tribu des événements antérieurs à T en posant

$$\mathcal{F}_T = \{\, A \in \mathcal{F} \,:\, A \cap \{\, T \leq n \,\} \in \mathcal{F}_n \text{ pour tout } n \in \mathbb{N} \,\}\,.$$

On obtient bien sûr une définition équivalente en remplaçant l'événement $\{\, T \leq n \,\}$ par l'événement $\{\, T = n \,\}$. On vérifie immédiatement que $\mathcal{F}_T$ est effectivement une tribu et que T est $\mathcal{F}_T$-mesurable.

***Exemple* VII.1.7.** Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ un processus adapté. Soit t un nombre réel et $T = \min\{\, n \in \mathbb{N} \,:\, X_n \geq t \,\}$ avec $T = \infty$ s'il n'existe pas de tel n. Alors T est un temps d'arrêt puisque

$$\{\, T \leq n \,\} = \{\, \exists\, m \leq n \,:\, X_m \geq t \,\} \in \mathcal{F}_n$$

et $\{\, T \leq \infty \,\} = \Omega \in \mathcal{F}_\infty$.

Dans la suite, lorsque nous parlerons de temps d'arrêt, il sera toujours sous-entendu par rapport à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Notons à présent quelques propriétés des temps d'arrêt : si S et T sont deux temps d'arrêt, alors $S \vee T$ et $S \wedge T$ sont aussi des temps d'arrêt. En particulier, une variable S constante étant un temps d'arrêt, pour tout $m \in \mathbb{N}$, $T \wedge m$ est un temps d'arrêt. On a aussi la proposition suivante.

***Proposition* VII.1.8.** *Soient S et T deux temps d'arrêt, tels que $S \leq T$ (partout) ; alors $\mathcal{F}_S \subset \mathcal{F}_T$.*

Démonstration. Soit $A \in \mathcal{F}_S$. Puisque $S \leq T$, pour tout $n \in \mathbb{N}$,

$$A \cap \{\, T \leq n \,\} = \big(A \cap \{\, S \leq n \,\}\big) \cap \{\, T \leq n \,\}$$

est bien élément de $\mathcal{F}_n$ car intersection de deux éléments de $\mathcal{F}_n$. □

Montrons maintenant que d'un point de vue probabiliste il est raisonnable de s'intéresser à l'objet X_T, c'est-à-dire au processus $(X_n)_{n\in\mathbb{N}}$ vu à l'instant aléatoire T.

***Lemme* VII.1.9.** *Si $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est un processus adapté et si T est un temps d'arrêt de la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, on définit une variable aléatoire X_T en posant $X_T(\omega) = X_{T(\omega)}(\omega)$ si $T(\omega) < \infty$ (la valeur $\mathcal{F}$-mesurable de $X_T(\omega)$ quand $T(\omega) = +\infty$ est indifférente). Alors X_T est $\mathcal{F}_T$-mesurable.*

Démonstration. Si B est un borélien de $\mathbb{R}$ et n un entier,

$$\{\, X_T \in B \,\} \cap \{\, T \le n \,\} = \bigcup_{0 \le k \le n} \{\, X_k \in B\,;\, T = k \,\}$$

qui est clairement mesurable par rapport à $\mathcal{F}_n$ puisque X_k est $\mathcal{F}_k$-mesurable pour tout k. □

Nous avons commencé par définir les martingales, puis la notion de temps d'arrêt, et venons de montrer que si T est un temps d'arrêt de la martingale, alors X_T est une variable aléatoire. Remarquons que la définition d'une martingale (X_n) suppose que chaque X_n est intégrable. Il est naturel d'étudier l'intégrabilité de X_T. En géneral, cette variable aléatoire n'a aucune raison d'être intégrable. Une classe naturelle de martingales à considérer pour conserver la propriété d'intégrabilité par arrêt est la classe des martingales dites L^1.

Définition VII.1.10. Une martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est dite L^1 si

$$\sup_{n\in\mathbb{N}} \|X_n\|_1 = \sup_{n\in\mathbb{N}} E\big(|X_n|\big) < \infty\,.$$

Proposition VII.1.11. *Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une martingale L^1 et T un temps d'arrêt fini p.s. (i.e. $P\{T < \infty\} = 1$). Alors X_T est intégrable et de plus $E(|X_T|) \le \sup_{n\in\mathbb{N}} E(|X_n|)$.*

Démonstration. Soit la fonction

$$\psi(a,b) = |a| - |b| - (a-b)\,\mathrm{signe}(b) = |a| - a\,\mathrm{signe}(b) \ge 0\,, \quad a,b \in \mathbb{R}\,.$$

Soit $m \in \mathbb{N}$. Pour le temps d'arrêt $T \wedge m$, observons que

$$|X_{T\wedge m}| = \sum_{0\le i\le m-1} |X_i|\,\mathbb{1}_{\{i\}}(T) + |X_m|\,\mathbb{1}_{[m,\infty)}(T)\,,$$

et donc

$$|X_{T\wedge m}| - |X_0| = \sum_{0\le i\le m-1} \big(|X_{i+1}| - |X_i|\big)\,\mathbb{1}_{]i,\infty[}(T)\,.$$

Puisque $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale et $\mathrm{signe}(X_i)$ et $\mathbb{1}_{]i,\infty[}(T)$ sont $\mathcal{F}_i$-mesurables,

$$\begin{aligned} &E\big((X_{i+1} - X_i)\mathrm{signe}(X_i)\mathbb{1}_{]i,\infty[}(T)\big) \\ &\qquad = E\Big(\big(E(X_{i+1} \mid \mathcal{F}_i) - X_i\big)\mathrm{signe}(X_i)\mathbb{1}_{]i,\infty[}(T)\Big) = 0\,. \end{aligned}$$

Ainsi, en utilisant la positivité de ψ,

$$\begin{aligned}E\big(|X_{T\wedge m}| - |X_0|\big) &= \sum_{0\leq i\leq m-1} E\big(\psi(X_{i+1}, X_i)\mathbb{1}_{]i,\infty[}(T)\big)\\ &\leq \sum_{0\leq i\leq m-1} E\big(\psi(X_{i+1}, X_i)\big)\\ &= \sum_{0\leq i\leq m-1} E\big(|X_{i+1}| - |X_i|\big)\\ &= E\big(|X_m|\big) - E\big(|X_0|\big)\,.\end{aligned}$$

Donc pour tout $m \in \mathbb{N}$,

$$E\big(|X_{T\wedge m}|\big) \leq E\big(|X_m|\big) \leq \sup_{n\in\mathbb{N}} E\big(|X_n|\big)\,.$$

Puisque $T < \infty$ p.s., $\lim_{m\to\infty} |X_{T\wedge m}| = |X_T|$ p.s. et en utilisant le lemme de Fatou II.2.3,

$$E\big(|X_T|\big) \leq \liminf_{m\to\infty} E\big(|X_{T\wedge m}|\big) \leq \sup_{n\in\mathbb{N}} E\big(|X_n|\big) < \infty\,,$$

ce qui démontre l'intégrabilité de X_T. □

Pour une martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ et une suite de temps d'arrêt $(T_n)_{n\in\mathbb{N}}$, la proposition VII.1.11 donne une condition suffisante pour vérifier la condition d'intégrabilité $E(|X_{T_m}|) < \infty$; il suffit que la martingale soit L^1. Pour les sur- ou les sous-martingales, une condition suffisante (et plus restrictive) est d'avoir $T_m < t_m < \infty$ p.s. où t_m est une suite déterministe. En effet, dans ce cas,

$$E\big(|X_{T_m}|\big) \leq \sum_{1\leq n\leq t_m} \int_{\{T=t_n\}} |X_n|\,\mathrm{d}P \leq \sum_{1\leq n\leq t_m} E\big(|X_n|\big) < \infty\,.$$

Le théorème suivant, le théorème d'arrêt de Doob, est fondamental. Il exprime qu'un jeu reste équilibré à tout temps (d'arrêt) aléatoire.

***Théorème VII.1.12* (d'arrêt de Doob).** *Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une sous-martingale (resp. une sur-martingale, resp. une martingale), et soit $(T_m)_{m\in\mathbb{N}}$ une suite de temps d'arrêt bornés de la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, et tels que $T_n \leq T_m$ pour tous $n \leq m$. Alors, le processus $(X_{T_m}, \mathcal{F}_{T_m})_{m\in\mathbb{N}}$ est une sous-martingale (resp. une sur-martingale, resp. une martingale).*

Démonstration. D'après le lemme VII.1.9, les X_{T_m} sont $\mathcal{F}_{T_m}$-mesurables. Il ne reste plus qu'à vérifier l'inégalité des sous-martingales (resp. des sur-martingales, resp. des martingales). Nous nous contentons du cas des sous-martingales, les autres cas se traitant de façon tout à fait identique. Comme les temps d'arrêt T_m sont bornés, il suffit de considérer une sous-martingale $(X_n, \mathcal{F}_n)_{1\le n\le k}$ et deux temps d'arrêt S et T de la filtration $(\mathcal{F}_n)_{1\le n\le k}$ tels que $S \le T$ et de montrer que

$$E(X_T \mid \mathcal{F}_S) \ge X_S\,.$$

Nous montrons à cet effet que pour tout $A \in \mathcal{F}_S$,

$$\int_A (X_T - X_S)\,\mathrm{d}P \ge 0$$

(prendre $A = \{\, E(X_T \mid \mathcal{F}_S) < X_S \,\}$ pour conclure).

Nous examinons d'abord le cas où la différence $T - S \in \{\,0,1\,\}$. Dans ce cas, on écrit pour tout $A \in \mathcal{F}_S$,

$$\begin{aligned}\int_A (X_T - X_S)\,\mathrm{d}P &= \sum_{1\le n\le k} \int_{A\cap\{S=n\}} (X_T - X_n)\,\mathrm{d}P\\ &= \sum_{1\le n\le k} \int_{A\cap\{S=n\}\cap\{T\neq n\}} (X_{n+1} - X_n)\,\mathrm{d}P\\ &\ge 0\end{aligned}$$

puisque $(X_n, \mathcal{F}_n)_{1\le n\le k}$ est une sous-martingale, $A \cap \{\, S = n \,\} \in \mathcal{F}_n$ et

$$\{\, T \neq n \,\} = \{\, T = n \,\}^{\mathrm{c}} \in \mathcal{F}_n\,.$$

Pour en déduire le cas général, on pose $R_l = \min(T, S + l)$, $1 \le l \le k$. Les R_l sont des temps d'arrêt de $(\mathcal{F}_n)_{1\le n\le k}$. Observons en outre que $R_0 = S$ et $R_k = T$. De plus $R_{l+1} \ge R_l$ et $R_{l+1} - R_l \in \{\,0,1\,\}$. Maintenant, si $A \in \mathcal{F}_S$, alors $A \in \mathcal{F}_{R_l}$ (proposition VII.1.8) et, d'après le premier cas,

$$\int_A (X_T - X_S)\,\mathrm{d}P = \sum_{1\le l\le k} \int_A (X_{R_{l+1}} - X_{R_l})\,\mathrm{d}P \ge 0\,.$$

Le théorème est établi. □

Si $(X_n, \mathcal{F}_n)_{1\le n\le k}$ est une sous-martingale, et si T est un temps d'arrêt de la filtration $(\mathcal{F}_n)_{1\le n\le k}$, le théorème d'arrêt VII.1.12 implique $E(X_1) \le E(X_T) \le E(X_k)$.

Le théorème suivant est une conséquence du théorème d'arrêt par l'intermédiaire de ce corollaire. C'est un énoncé faisant partie des inégalités dites maximales.

Théorème VII.1.13. *Soit* $(X_n, \mathcal{F}_n)_{1 \le n \le k}$ *une sous-martingale ; pour tout* $t > 0$,

$$P\Big\{ \max_{1 \le n \le k} X_n \ge t \Big\} \le \frac{E(X_k^+)}{t}.$$

Démonstration. On considère le temps d'arrêt

$$T = \min\{ 1 \le n \le k : X_n \ge t \}$$

ou $T = k$ si cet ensemble est vide. On notera que si $\max_{1 \le n \le k} X_n \ge t$, alors $X_T \ge t$, et si $\max_{1 \le n \le k} X_n < t$, alors $X_T = X_k$. Ainsi, par le théorème d'arrêt VII.1.12, plus précisément sa conséquence ci-dessus,

$$\begin{aligned} E(X_k) \ge E(X_T) &= \int_{\{\max_{1 \le n \le k} X_n < t\}} X_T \,\mathrm{d}P + \int_{\{\max_{1 \le n \le k} X_n \ge t\}} X_T \,\mathrm{d}P \\ &\ge \int_{\{\max_{1 \le n \le k} X_n < t\}} X_k \,\mathrm{d}P + tP\Big\{ \max_{1 \le n \le k} X_n \ge t \Big\}. \end{aligned}$$

Par conséquent,

$$tP\Big\{ \max_{1 \le n \le k} X_n \ge t \Big\} \le \int_{\{\max_{1 \le n \le k} X_n \ge t\}} X_k \,\mathrm{d}P \le E(X_k^+)$$

puisque $Z \mathbb{1}_A \le Z^+$ pour toute variable Z et tout événement A. □

Si $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ est une martingale, alors $(|X_n|, \mathcal{F}_n)_{n \in \mathbb{N}}$ est une sous-martingale, et donc, pour tout $k \in \mathbb{N}$ et tout $t > 0$,

$$P\Big\{ \max_{0 \le n \le k} |X_n| \ge t \Big\} \le \frac{E\big(|X_k|\big)}{t}.$$

Observons que la suite $\big(\mathbb{1}_{[t,\infty[}(\max_{0 \le n \le k} |X_n|)\big)_{k \in \mathbb{N}}$ est croissante et majorée par 1, et converge presque sûrement vers $\mathbb{1}_{[t,\infty[}(\sup_{n \in \mathbb{N}} (|X_n|))$. On déduit donc de l'inégalité précédente et du théorème de convergence dominée (II.2.8) que pour une martingale L^1,

$$P\Big\{ \sup_{n \in \mathbb{N}} |X_n| \ge t \Big\} \le \frac{1}{t} \sup_{n \in \mathbb{N}} E\big(|X_n|\big).$$

En particulier, $\sup_{n \in \mathbb{N}} |X_n| < \infty$ p.s. Lorsque de plus X_n est de carré intégrable, $(X_n^2, \mathcal{F}_n)_{n \in \mathbb{N}}$ est une sous-martingale (exemple VII.1.4.iii). Le théorème VII.1.12 fournit dans ce cas,

$$P\Big\{ \max_{1 \le n \le k} |X_n| \ge t \Big\} \le P\Big\{ \max_{1 \le n \le k} X_n^2 \ge t^2 \Big\} \le \frac{1}{t^2} E(X_k^2).$$

Par exemple, si $X_n = Z_1 + \cdots + Z_n$ où les Z_i sont indépendantes centrées et de carré intégrable, on retrouve par ces méthodes l'inégalité de Kolmogorov (exercice IV.16).

VII.2. Théorèmes de convergence

Il convient de remarquer que la définition d'une sur-martingale est à peu près celle d'une suite qui en tendance, conditionnellement au passé, décroît. Il est bien connu en analyse qu'une suite décroissante minorée converge. L'un des buts de cette partie est de démontrer le résultat analogue pour les sur-martingales. La condition de minoration des suites réelles ($\inf_n x_n > -\infty$) implique la bornitude et deviendra ici une condition de bornitude d'espérance, $\sup_n E(|X_n|) < \infty$. On peut imaginer alors l'importance d'un tel théorème de convergence.

Avant de montrer la convergence des sur-martingales, nous montrerons celle des martingales L^1, et conclurons grâce à la décomposition de Doob. L'étude de la convergence des martingales fournira dans certaines situations une alternative à l'utilisation du lemme de Borel-Cantelli V.1.2.

Toujours par analogie avec l'étude des suites réelles, observons que pour qu'une suite réelle $(x_n)_{n\in\mathbb{N}}$ converge, il suffit (mais ce n'est pas nécessaire) que pour toute suite strictement croissante d'entiers n_j, $j \in \mathbb{N}$, avec $n_0 = 0$, on ait $\sum_{j\in\mathbb{N}}(x_{n_{j+1}} - x_{n_j})^2 < \infty$. En effet, si $(x_n)_{n\in\mathbb{N}}$ ne converge pas, elle ne vérifie pas le critère de Cauchy ; alors il existe $\varepsilon > 0$ et une suite croissante n_j avec par exemple $n_0 = 0$ telle que $|x_{n_{j+1}} - x_{n_j}| \geq \varepsilon$, et donc $\sum_{j\in\mathbb{N}}(x_{n_{j+1}} - x_{n_j})^2 = \infty$. Si nous voulons qu'une martingale converge p.s., nous pouvons tenter d'utiliser ce critère de convergence pour presque tout aléa ω. L'analogue de la suite croissante n_j est naturellement une suite croissante de temps d'arrêt. Nous pouvons maintenant énoncer puis démontrer la convergence des martingales L^1.

***Théorème VII.2.1* (de convergence des martingales).** *Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une martingale L^1. Alors $\lim_{n\to\infty} X_n$ existe p.s.*

Démonstration. En suivant le commentaire précédent, montrons d'abord que pour toute suite presque sûrement croissante de temps d'arrêt bornés, $(T_n)_{n\in\mathbb{N}}$, avec $T_0 = 0$ p.s., la série $\sum_{n\in\mathbb{N}}(X_{T_{n+1}} - X_{T_n})^2$ converge p.s.

Puisque le développement $(X_{T_{n+1}} - X_{T_n})^2 = X_{T_{n+1}}^2 + X_{T_n}^2 - 2X_{T_{n+1}}X_{T_n}$ fait apparaître des carrés et que nous supposons seulement que les X_n sont intégrables, nous utilisons une troncature. La démonstration est alors dans le même esprit que celle de la proposition VII.1.11.

Pour tout $p > 0$, soit ϕ_p la fonction positive, convexe, dérivable, définie par

$$\phi_p(x) = \begin{cases} x^2 & \text{si } |x| \leq p, \\ 2p|x| - p^2 & \text{si } |x| \geq p. \end{cases}$$

Soit de plus la fonction positive

$$\psi_p(x,y) = \phi_p(y) - \phi_p(x) - (y-x)\phi'(x)\,, \qquad x,y \in \mathbb{R}\,.$$

Observons que $\psi_p(x,y) = (y-x)^2$ si $|x| \vee |y| \leq p$ et que de plus $\phi_p(x) \leq 2p|x|$ pour tout $x \in \mathbb{R}$. Considérons la variable aléatoire $X^* = \sup_{n\in\mathbb{N}} |X_n|$ qui est bien définie d'après la discussion suivant l'inégalité maximale VII.1.13. Pour tout k,

$$\begin{aligned}
&E\Big(\sum_{0\leq n\leq k} (X_{T_{n+1}} - X_{T_n})^2 \mathbb{1}_{[0,p]}(X^*)\Big) \\
&\qquad = E\Big(\sum_{0\leq n\leq k} \psi_p(X_{T_n}, X_{T_{n+1}}) \mathbb{1}_{[0,p]}(X^*)\Big) \\
&\qquad \leq E\Big(\sum_{0\leq n\leq k} \psi_p(X_{T_n}, X_{T_{n+1}})\Big) && \text{(puisque } \psi_p \geq 0\text{)} \\
&\qquad = E\Big(\sum_{0\leq n\leq k} \phi_p(X_{T_{n+1}}) - \phi_p(X_{T_n}) - (X_{T_{n+1}} - X_{T_n})\phi'_p(X_{T_n})\Big) \\
&\qquad = E\big(\phi_p(X_{T_{k+1}})\big) - E\big(\phi_p(X_{T_0})\big) - \sum_{0\leq n\leq k} E\big((X_{T_{n+1}} - X_{T_n})\phi'_p(X_{T_n})\big)\,.
\end{aligned}$$

D'après le théorème d'arrêt VII.1.12, $(X_{T_n}, \mathcal{F}_{T_n})_{n\in\mathbb{N}}$ est une martingale. Ainsi,

$$E\big((X_{T_{n+1}} - X_{T_n})\phi'_p(X_{T_n})\big) = E\Big(E\big(X_{T_{n+1}} - X_{T_n} \mid \mathcal{F}_{T_n}\big)\phi'_p(X_{T_n})\Big) = 0\,.$$

Il s'ensuit que

$$\begin{aligned}
E\Big(\sum_{0\leq n\leq k} (X_{T_{n+1}} - X_{T_n})^2 \mathbb{1}_{[0,p]}(X^*)\Big) &\leq E\big(\phi_p(X_{T_{k+1}})\big) - E\big(\phi_p(X_0)\big) \\
&\leq 2pE\big(|X_{T_{k+1}}|\big) \\
&\leq 2p \sup_{n\in\mathbb{N}} E\big(|X_n|\big) < \infty\,. \qquad (1)
\end{aligned}$$

Supposons alors que la martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ ne converge pas presque sûrement. Considérons l'événement

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} \{\, |X_m - X_n| > \varepsilon \,\}\,.$$

La discussion suivant la définition V.1.1 montre qu'il existe $\varepsilon > 0$ tel que $P(A) > 2\varepsilon$. Par convergence monotone, $P(A \cap \{\, X^* \leq p \,\}) > \varepsilon$ pour tout p assez grand.

Définissons alors la suite croissante de temps d'arrêt $(T_n)_{n\in\mathbb{N}}$ par $T_0 = 0$ et pour tout $n \geq 0$,

$$T_{n+1} = \min\{\, m > T_n : |X_m - X_{T_n}| > \varepsilon \,\}$$

si $T_n < \infty$ (et $T_{n+1} = \infty$ si $T_n = \infty$). Soit N un entier positif. L'égalité (1) appliquée aux temps d'arrêt $T_n \wedge N$ montre que

$$\begin{aligned} 2p \sup_{n\in\mathbb{N}} E\big(|X_n|\big) &\geq E\Big(\mathbb{1}_A \mathbb{1}_{[0,p]}(X^*) \sum_{0\leq n\leq k} (X_{T_{n+1}\wedge N} - X_{T_n\wedge N})^2\Big) \\ &\geq \varepsilon^2 E\big(\mathbb{1}_A \mathbb{1}_{[0,p]}(X^*) \mathrm{card}\{\, 0 \leq n \leq k : T_{n+1} \leq N \,\}\big)\,. \end{aligned}$$

Par convergence monotone (en k et N),

$$\varepsilon^2 E\big(\mathbb{1}_A \mathbb{1}_{[0,p]}(X^*) \mathrm{card}\{\, n \in \mathbb{N} : T_n < \infty \,\}\big) \leq 2p \sup_{n\in\mathbb{N}} E\big(|X_n|\big)\,. \tag{2}$$

Nous avons montré que l'événement $A \cap \{X^* \leq p\}$ a une probabilité positive ; or si l'événement A a lieu, alors l'ensemble $\{\, n \in \mathbb{N} : T_n < \infty \,\}$ est infini, ce qui contredit (2). □

Comme annoncé, nous déduisons de la convergence des martingales L^1 celle des sous-martingales.

Corollaire VII.2.2. *Soit $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ une sous-martingale (resp. sur-martingale), telle que $\sup_{n\in\mathbb{N}} E(|X_n|) < \infty$. Alors $\lim_{n\to\infty} X_n$ existe p.s.*

Démonstration. Soit $X_n = Y_n + Z_n$ la décomposition de Doob 1.5 de la sous-martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$. Comme $Z_n \geq 0$ p.s., $\sup_n E(|Y_n|) < \infty$ et, par convergence monotone,

$$E(\sup_n Z_n) = \sup_n E(Z_n) < \infty\,.$$

Ainsi, le processus (Z_n) est croissant et borné p.s., donc converge p.s. La martingale (Y_n) est quant à elle dans L^1, donc converge p.s. d'après le théorème VII.2.1. La convergence presque sûre de la sous-martingale $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ s'en déduit. □

L'énoncé suivant décrit les martingales uniformément intégrables.

Théorème VII.2.3. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires adaptées à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$; pour que (X_n) soit une martingale uniformément intégrable (relativement à $(\mathcal{F}_n)_{n\in\mathbb{N}}$), il faut et il suffit qu'il existe une variable aléatoire intégrable Y telle que $X_n = E(Y \mid \mathcal{F}_n)$ p.s. pour tout n.*

Démonstration. Si (X_n) est uniformément intégrable, par le corollaire VII.2.2 et le théorème V.VII.3.5, X_n converge p.s. vers X_∞ et aussi dans L^1. On choisit $Y = X_\infty$, pour lequel il faut vérifier que $E(Y \mid \mathcal{F}_n) = X_n$ p.s. pour tout n. Or pour tout $m \geq n$, $E(X_m \mid \mathcal{F}_n) = X_n$ et

$$E\Big(\big|E(Y|\mathcal{F}_n) - E(X_m|\mathcal{F}_n)\big|\Big) \leq E\big(|Y - X_m|\big)$$

qui tend vers 0 lorsque $m \to \infty$.

Réciproquement, il faut montrer l'uniforme intégrabilité de toute suite du type $(E(Y \mid \mathcal{F}_n))_{n\in\mathbb{N}}$. Cela se fait en revenant à la définition. Soit $X_n = E(Y \mid \mathcal{F}_n)$. Pour tout $n \in \mathbb{N}$ et tout $c > 0$,

$$\int_{\{|X_n|>c\}} |X_n| \,\mathrm{d}P \leq \int_{\{|X_n|>c\}} E\big(|Y| \bigm| \mathcal{F}_n\big) \,\mathrm{d}P \leq E\big(|Y| \mathbb{1}_{]c,\infty[}(|X_n|)\big)$$

puisque X_n est $\mathcal{F}_n$-mesurable. Comme Y est intégrable, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que $P(A) \leq \eta$ assure $\int_A |Y| \,\mathrm{d}P \leq \varepsilon$. Or, pour chaque n,

$$P\big\{\, |X_n| > c \,\big\} \leq \frac{1}{c} E\big(|X_n|\big) = \frac{1}{c} E\Big(\big|E(Y|\mathcal{F}_n)\big|\Big) \leq \frac{1}{c} E\big(|Y|\big) \,.$$

Donc, si $c_0 = E(|Y|)/\eta$, pour tout $c > c_0$,

$$\sup_n E\big(|Y| \mathbb{1}_{]c,\infty[}(|X_n|)\big) \leq \varepsilon \,,$$

et la conclusion s'ensuit. □

On peut aussi démontrer des théorèmes de convergence presque sûre pour des ensembles d'indices filtrant à gauche, et ceux-ci sont parfois bien utiles. Une telle situation est par exemple le cas des entiers négatifs, ou de façon équivalente, de l'ensemble des entiers naturels avec un ordre renversé.

***Définition** VII.2.4.* Sur $(\Omega, \mathcal{F}, P)$, soient une suite décroissante $(\mathcal{F}_n)_{n\in\mathbb{N}}$ de sous-tribus de $\mathcal{F}$, et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires intégrables adaptées à $(\mathcal{F}_n)_{n\in\mathbb{N}}$. La suite $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale (resp. sur-martingale, resp. sous-martingale) renversée, si, lorsque $m \leq n$,

$$E(X_m \mid \mathcal{F}_n) = X_n \quad p.s.$$

(resp. $E(X_m \mid \mathcal{F}_n) \leq X_n$, resp. $E(X_m \mid \mathcal{F}_n) \geq X_n$).

Le théorème suivant se démontre en utilisant le même schéma que pour l'ordre habituel. Les hypothèses sont quelques peu modifiées. C'est l'analogue du théorème d'analyse affirmant que toute suite de réels croissante et majorée converge.

Théorème VII.2.5. *Soit* $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ *une sur-martingale renversée telle que* $\sup_{n\in\mathbb{N}} E(X_n) < \infty$. *Alors les variables aléatoires* X_n *convergent p.s. vers une variable aléatoire intégrable* X_∞.

Notons en particulier qu'une martingale renversée est toujours p.s. convergente (puisque la suite $(E(X_n))_{n\in\mathbb{N}}$ est constante).

Nous omettons la démonstration du théorème VII.2.5. Elle est tout à fait semblable à celle du théorème VII.2.1 ; il suffit de noter que, puisque X_n^- est une sous-martingale renversée, pour tout n,

$$E\big(|X_n|\big) = E(X_n) + 2E(X_n^-) \leq \sup_{k\in\mathbb{N}} E(X_k) + 2E(X_0^-)\,,$$

et donc $\sup_{n\in\mathbb{N}} E(|X_n|) < \infty$. Cette hypothèse suffit alors pour appliquer le schéma usuel.

Remarque. Il est possible de démontrer que sous les hypothèses du théorème, la suite $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable. La convergence a donc aussi lieu dans L^1.

VII.3. Application à la loi des grands nombres

Soient Z_n, $n \geq 1$, des variables aléatoires réelles, indépendantes, de même loi, définies sur $(\Omega, \mathcal{F}, P)$, et telles que $E(|Z_1|) < \infty$. Pour tout $n \geq 1$, posons $S_n = Z_1 + \cdots + Z_n$.

Nous allons vérifier que $X_n = S_n/n$, $n \geq 1$, est une martingale renversée pour la filtration $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots)$, $n \geq 1$. À cet effet, il nous suffit de montrer que pour tout $n \geq 1$,

$$E(X_1 \mid \mathcal{F}_n) = X_n \quad \text{p.s.},$$

car si c'est le cas, pour tout $m \leq n$,

$$E(X_m \mid \mathcal{F}_n) = E(X_1 \mid \mathcal{F}_m \mid \mathcal{F}_n) = E(X_1 \mid \mathcal{F}_n) = X_n\,.$$

Notons en outre que $X_1 = Z_1$. Or, par linéarité, on peut écrire, pour tout $n \geq 1$,

$$S_n = E(S_n \mid \mathcal{F}_n) = \sum_{1\leq i\leq n} E(Z_i \mid \mathcal{F}_n)\,.$$

La tribu $\mathcal{F}_n$ est aussi engendrée par S_n, $Z_{n+1}, Z_{n+2}, \ldots$ Comme les Z_i sont indépendantes, la proposition VI.2.2.viii montre que

$$S_n = \sum_{1\leq i\leq n} E(Z_i \mid S_n, Z_{n+1}, Z_{n+2}, \ldots) = \sum_{1\leq i\leq n} E(Z_i \mid S_n)\,.$$

Les Z_i ayant même loi, il vient ensuite

$$\sum_{1\leq i\leq n} E(Z_i \mid S_n) = nE(Z_1 \mid S_n) = nE(Z_1 \mid \mathcal{F}_n) = nE(X_1 \mid \mathcal{F}_n)\,,$$

d'où le résultat.

En vertu du théorème VII.2.5, X_n converge p.s. Par la loi du 0–1 (IV.3.3), la limite est non aléatoire. Notons la a. Il ne reste plus qu'à montrer que $a = E(X_1)$. Ceci sera en particulier le cas si la suite $(X_n)_{n\geq 1}$ est uniformément intégrable puisqu'alors X_n convergera dans L^1 vers a, et donc $E(X_n) \to a$. Cela fournira le résultat puisque $E(X_n) = E(X_1)$ pour tout n. D'après la remarque suivant le théorème VII.2.5, une martingale renversée est toujours uniformément intégrable. La démonstration est identique à la deuxième partie du théorème VII.2.3. Pour tout $c > 0$ et tout $n \geq 1$,

$$\int_{\{|X_n|>c\}} |X_n|\,\mathrm{d}P \leq \frac{1}{n}\sum_{1\leq i\leq n}\int_{\{|S_n/n|>c\}} |Z_i|\,\mathrm{d}P = \int_{\{|S_n/n|>c\}} |Z_1|\,\mathrm{d}P\,.$$

Soit $\varepsilon > 0$ fixé, et soit $\eta > 0$ tel que si $P(A) \leq \eta$ alors $\int_A |Z_1|\,\mathrm{d}P \leq \varepsilon$. Pour tout $c > 0$ et tout n,

$$P\big\{\,|S_n/n| > c\,\big\} \leq \frac{1}{nc}E\big(|S_n|\big) \leq \frac{1}{c}E\big(|Z_1|\big)\,,$$

de sorte que si $c_0 = E(|Z_1|)/\eta$, pour tout $c \geq c_0$ et tout $n \geq 1$,

$$\int_{\{|X_n|>c\}} |X_n|\,\mathrm{d}P \leq \int_{\{|S_n/n|>c\}} |Z_1|\,\mathrm{d}P \leq \varepsilon\,.$$

La suite $(X_n)_{n\geq 1}$ est donc bien uniformément intégrable.

En résumé, nous venons de démontrer la loi forte des grands nombres :

Théorème VII.3.1. *Soit $(Z_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et de même loi et soit $S_n = Z_1 + \cdots + Z_n$, $n \geq 1$. Alors*

$$\lim_{n\to\infty} \frac{S_n}{n} = E(Z_1) \text{ p.s. si et seulement si } E\big(|Z_1|\big) < \infty\,.$$

En fait, nous n'avons fait que démontrer une partie de la loi des grands nombres, *i.e.* que $E(|Z_1|) < \infty$ implique la loi forte. La réciproque a été établie dans le théorème V.5.2.

Exercices

Exercice VII.1. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de même loi de Bernoulli $P\{\,X_n = 0\,\} = P\{\,X_n = 2\,\} = 1/2$. Pour tout $n \geq 1$, on désigne par $\mathcal{F}_n$ la tribu engendrée par $X_1, \ldots, X_n$, et l'on pose $Z_n = \prod_{1\leq k\leq n} X_k$. Démontrer que $(Z_n)_{n\geq 1}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$ qui n'est pas uniformément intégrable.

Exercice VII.2. Soient $c_1, \ldots, c_k$ des réels tels que $\sum_{1\leq i\leq k} c_i = 0$. Soit π une permutation aléatoire de $\{\,1, 2, \ldots, k\,\}$ uniformément répartie sur le groupe des permutations de k éléments, c'est-à-dire telle que pour toute permutation τ de k éléments, $P\{\,\pi = \tau\,\} = 1/k!$. Soit

$$X_n = \frac{k}{k-n} \sum_{1\leq i\leq n} c_{\pi(i)}$$

et soit la suite de tribus $\mathcal{F}_n = \sigma(\pi(1), \ldots, \pi(n))$, $1 \leq n \leq k$. Montrer que $(X_n, \mathcal{F}_n)_{1\leq n\leq k}$ est une martingale.
Indication : Montrer que

$$X_n - X_{n-1} = \frac{k}{k-n}\Big(c_{\pi(n)} - \frac{1}{k-n+1} \sum_{n\leq i\leq k} c_{\pi(i)}\Big),$$

puis montrer que pour tout $n \leq i \leq k$, $\mathcal{L}(\pi(i) \mid \pi(1), \ldots, \pi(n-1))$ est la loi uniforme sur $\{\,1, 2, \ldots, n\,\} \setminus \{\,\pi(1), \ldots, \pi(n-1)\,\}$.

Exercice VII.3. (Urne de Polya) Une urne contient n boules noires et b boules blanches. Une boule est tirée au hasard, selon une probabilité uniforme sur les boules dans l'urne. Elle est remise dans l'urne, et on ajoute aussi a boules de la couleur tirée. On itère cette procédure de tirage-ajout. Soit $X_0 = n/(n+b)$ la proportion de boules noires initialement dans l'urne, et soit X_k la proportion de boules noires à la k-ième étape du tirage-ajout. Montrer que X_k est une martingale, pour la suite de tribus $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Montrer que cette martingale converge, et donc que la proportion de boules noires converge vers une proportion a priori aléatoire Y.
Note : On peut montrer, mais cela demande un peu de calcul, que Y a pour loi une loi de densité

$$\frac{\Gamma\big(\frac{n+b}{a}\big)}{\Gamma\big(\frac{n}{a}\big)\Gamma\big(\frac{b}{a}\big)}(1-x)^{\frac{n}{a}-1}x^{\frac{b}{a}-1}\,, \qquad 0 < x < 1$$

(voir par exemple Feller (1971)).

Exercice* VII.*4. (Lemme de Wald.) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, de même loi, et soit, pour tout $n \geq 1$, $S_n = X_1 + \cdots + X_n$. Soit en outre T un temps d'arrêt intégrable relatif à la filtration engendrée par cette suite. Démontrer que $E(S_T) = E(X_1)E(T)$.

Exercice* VII.*5. Sur $(\Omega, \mathcal{F}, P)$, soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes, de même loi. Pour tout $n \geq 1$, soit $\mathcal{F}_n$ la tribu engendrée par $X_1, \ldots, X_n$. On note les sommes partielles $S_n = X_1 + \cdots + X_n$, $n \geq 1$. On convient que $S_0 = 0$ et, pour tout $x \in \mathbb{R}$, on désigne par E^x l'espérance définie par $E^x(\cdot) = E(\cdot + x)$. On parle alors de la marche aléatoire S_n partant de x au temps 0.
a) Soit $N \geq 1$ un entier fixé et soit T un temps d'arrêt à valeurs dans $\{1, \ldots, N\}$ de la filtration $(\mathcal{F}_n)_{1\leq n\leq N}$. Démontrer que, pour tout $n \geq 1$, $S_{n+T} - S_T$ est indépendant de $\mathcal{F}_T$ et de même loi que S_n.
b) Déduire de la question précédente que pour toute fonction borélienne bornée ϕ sur $\mathbb{R}$, et tout $n \geq 1$,

$$E\big(\phi(S_{n+T}) \,\big|\, \mathcal{F}_T\big) = E^{S_T}\big(\phi(S_n)\big) \quad \text{p.s.}$$

Exercice* VII.*6. Soit $(X_n, \mathcal{F}_n)_{1\leq n\leq k}$ une martingale de carré intégrable. On définit $X^* = \max_{1\leq n\leq k} |X_n|$. En utilisant l'inégalité maximale de Doob, démontrer que

$$E\big((X^*)^2\big) \leq 4E(X_k^2)\,.$$

Exercice* VII.*7. Sur un espace probabilisé $(\Omega, \mathcal{F}, P)$, soit $(M_n)_{1\leq n\leq k}$ une martingale par rapport à une filtration $(\mathcal{F}_n)_{1\leq n\leq k}$ et soit $(H_n)_{1\leq n\leq k}$ une famille de variables aléatoires sur $(\Omega, \mathcal{F}, P)$ telles que H_n soit mesurable par rapport à $\mathcal{F}_{n-1}$, pour tout $n = 1, \ldots, k$ (avec la convention $\mathcal{F}_0 = \{\emptyset, \Omega\}$).

Soit $a > 0$; on définit $T = \min\{1 \leq n \leq k-1 : |H_{n+1}| > a\}$ et $T = k$ si l'ensemble dont on prend le minimum est vide. Démontrer que T est un temps d'arrêt de la filtration $(\mathcal{F}_n)_{1\leq n\leq k}$. On pose, pour tout $n = 1, \ldots, k$,

$$X_n = \sum_{1\leq i\leq T\wedge n} H_i(M_i - M_{i-1})$$

$(M_{-1} = 0)$. Démontrer que $(X_n)_{1\leq n\leq k}$ est une martingale de $(\mathcal{F}_n)_{1\leq n\leq k}$.

Exercice* VII.*8. On considère une variable aléatoire T à valeurs dans $\mathbb{N}$, de loi géométrique

$$P\{T = n\} = a(1+a)^{-n-1}\,, \qquad n \in \mathbb{N}\,,$$

où a est un réel positif donné. On appelle $\mathcal{F}_n$ la plus petite tribu rendant mesurable la variable $T \wedge n$, $n \in \mathbb{N}$. Vérifier que la famille de tribus $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est une filtration. Démontrer que $\mathcal{F}_n$ est engendrée par une partition de $n+1$ atomes que l'on précisera.

a) Démontrer que, pour tout n,

$$E\big(\mathbb{1}_{\{T\geq n+1\}} \mid \mathcal{F}_n\big) = (1+a)^{-1}\mathbb{1}_{\{T\geq n\}}\,.$$

b) Déduire de la question précédente que

$$E\big(T \wedge (n+1) \mid \mathcal{F}_n\big) = T \wedge n + (1+a)^{-1}\mathbb{1}_{\{T\geq n\}}\,.$$

c) Pour quelle valeur du paramètre réel α le processus

$$X_n = \alpha(T \wedge n) + \mathbb{1}_{\{T\geq n\}}\,, \quad n \in \mathbb{N}\,,$$

est-il une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$?

d) En prenant pour α la valeur trouvée à la question c), calculer l'espérance conditionnelle $E((X_{n+1} - X_n)^2 \mid \mathcal{F}_n)$. En déduire que le processus

$$X_n^2 - a(T \wedge (n-1))\,, \qquad n \geq 1\,,$$

est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Exercice VII.9. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{R}^d$; on considère une norme quelconque $\|\cdot\|$ sur $\mathbb{R}^d$, et on suppose que $E(\|X_i\|^2) < \infty$ pour tout $i = 1, \ldots, n$. Posons $S_n = X_1 + \cdots + X_n$.

Désignons par $\mathcal{A}_i$, $1 \leq i \leq n$, la sous-tribu de $\mathcal{A}$ engendrée par les variables $X_1, \ldots, X_i$ et par $\mathcal{A}_0$ la tribu triviale composée de $\emptyset$ et Ω. Pour tout $i = 1, \ldots, n$, posons

$$d_i = E\big(\|S_n\| \mid \mathcal{A}_i\big) - E\big(\|S_n\| \mid \mathcal{A}_{i-1}\big)\,.$$

a) Établir que

$$\|S_n\| - E\big(\|S_n\|\big) = \sum_{1\leq i\leq n} d_i\,.$$

Démontrer que pour tous $i < j$, $E(d_j \mid \mathcal{A}_i) = 0$, et que, de plus, les variables d_i, $i = 1, \ldots, n$, sont orthogonales.

b) Démontrer que pour tout $i = 1, \ldots, n$,

$$E\big(\|S_n - X_i\| \mid \mathcal{A}_i\big) = E\big(\|S_n - X_i\| \mid \mathcal{A}_{i-1}\big)\,.$$

Indication : On pourra utiliser le fait que si X est une variable aléatoire intégrable sur $(\Omega, \mathcal{A}, P)$, et $\mathcal{T}_1$, $\mathcal{T}_2$ sont deux sous-tribus de $\mathcal{A}$ telles que $\mathcal{T}_2$ est indépendante de la tribu engendrée par $\mathcal{T}_1$ et X, alors $E(X \mid \mathcal{T}_1) = E(X \mid \mathcal{T})$ où $\mathcal{T}$ est la tribu engendrée par $\mathcal{T}_1$ et $\mathcal{T}_2$. En déduire que

$$d_i = E\big(\|S_n\| - \|S_n - X_i\| \bigm| \mathcal{A}_i\big) - E\big(\|S_n\| - \|S_n - X_i\| \bigm| \mathcal{A}_{i-1}\big) .$$

c) Par l'inégalité du triangle et la question précédente, établir que

$$E(d_i^2 \mid \mathcal{A}_{i-1}) \le E\big(\|X_i\|^2\big) , \qquad i = 1, \dots, n .$$

En conclure, à l'aide de la première question, que

$$\mathrm{Var}\big(\|S_n\|\big) \le \sum_{1 \le i \le n} E\big(\|X_i\|^2\big) .$$

Exercice VII.10. Soit A_k^n, $k = 1, \dots, 2^{n-1}$, $n \ge 1$, la famille des intervalles dyadiques de l'intervalle $[\,0,1\,]$ muni de la mesure de Lebesgue λ. Si P est une mesure de probabilité sur $[\,0,1\,]$ absolument continue par rapport à λ, poser

$$X_n = \sum_{1 \le k \le 2^{n-1}} \frac{P(A_k^n)}{\lambda(A_k^n)} \, \mathbb{1}_{A_k^n} , \quad n \ge 1 .$$

Démontrer que, sur $([\,0,1\,], \lambda)$, $(X_n)_{n \ge 1}$ est une martingale par rapport à la suite de tribus $\mathcal{F}_n = \sigma(A_k^n, 1 \le k \le 2^{n-1})$, $n \ge 1$. Démontrer par l'absurde qu'elle est uniformément intégrable et en conclure l'existence de la densité de Radon-Nikodym de P par rapport à λ.

VIII

CHAÎNES DE MARKOV (À ESPACE D'ÉTATS DÉNOMBRABLE)

Le but de ce chapitre est de définir et de construire dans un cadre simple (ensemble d'indices et espace d'états discrets) des évolutions markoviennes et d'étudier leur comportement asymptotique. La propriété de Markov, à la base de ce chapitre, décrit un processus stochastique qui évolue avec le temps. La dépendance simple vis-à-vis du passé autorise de nombreux développements qui font la richesse de ces modèles.

VIII.1. La propriété de Markov

Dans tout ce chapitre, $\mathbb{E}$ est un ensemble dénombrable ; $\mathcal{P}(\mathbb{E})$ est l'ensemble de ses parties. La définition suivante présente l'objet fondamental de cette étude.

Définition VIII.1.1. On dit qu'une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$, à valeurs dans $(\mathbb{E},\mathcal{P}(\mathbb{E}))$ et définies sur un espace probabilisé $(\Omega,\mathcal{A},P)$, est une chaîne de Markov si, pour tout $(n+1)$-uplet $(i_0,\ldots,i_n)$ de points de $\mathbb{E}$ tel que $P\big(\bigcap_{0\leq j\leq n-1}\{\,X_j=i_j\,\}\big)>0$,

$$P\bigg\{\,X_n=i_n\;\bigg|\;\bigcap_{0\leq j\leq n-1}\{X_j=i_j\}\bigg\}=P\{\,X_n=i_n\mid X_{n-1}=i_{n-1}\,\}\,. \tag{1}$$

Autrement dit, la loi de X_n conditionnellement à $(X_0,\ldots,X_{n-1})$ et la loi de X_n conditionnellement à X_{n-1} sont identiques.

On appelle $\mathbb{E}$ l'espace des états. La loi de X_0 est appelée la loi ou la mesure initale.

L'égalité (1) s'appelle propriété de Markov. Nous verrons (proposition VIII.1.3) qu'elle traduit le fait que le futur du processus ne dépend du passé qu'à travers le présent.

Les exemples suivants montrent qu'il existe des chaînes de Markov.

***Exemples* VIII.1.2.** (i) Marche aléatoire sur $\mathbb{Z}^d$. Soit $(Y_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes sur $(\mathbb{Z}^d, \mathcal{P}(\mathbb{Z}^d))$. Soit $X_n = Y_0 + \cdots + Y_n$, $n \in \mathbb{N}$. Pour tous $i_0, \ldots, i_n \in \mathbb{Z}^d$ tels que l'on ait $P\big(\bigcap_{0\leq j\leq n-1}\{ X_j = i_j \}\big) > 0$,

$$\begin{aligned} P\bigg\{ X_n = i_n \,\bigg|\, \bigcap_{0\leq j\leq n-1} \{ X_j = i_j \} \bigg\} & \\ &= \frac{P\big(\bigcap_{0\leq j\leq n}\{ X_j = i_j \}\big)}{P\big(\bigcap_{0\leq j\leq n-1}\{ X_j = i_j \}\big)} \\ &= \frac{P\big(\{ Y_n = i_n - i_{n-1} \} \cap \bigcap_{0\leq j\leq n-1}\{ X_j = i_j \}\big)}{P\big(\bigcap_{0\leq j\leq n-1}\{ X_j = i_j \}\big)} \\ &= P\{ Y_n = i_n - i_{n-1} \} \end{aligned}$$

puisque Y_n est indépendante de $(X_0, \ldots, X_{n-1}) \in \sigma(Y_0, \ldots, Y_{n-1})$. De façon analogue,

$$P\{ X_n = i_n \mid X_{n-1} = i_{n-1} \} = P\{ Y_n = i_n - i_{n-1} \},$$

et donc $(X_n)_{n\in\mathbb{N}}$ est bien une chaîne de Markov à valeurs dans $(\mathbb{Z}^d, \mathcal{P}(\mathbb{Z}^d))$.

Lorsque $d = 1$ et Y_n suit une loi de Bernoulli sur $\{ -1, 1 \}$ de paramètre p, on appelle $(X_n)_{n\in\mathbb{N}}$ la marche aléatoire sur $\mathbb{Z}$. Lorsque de plus $p = 1/2$, on parle de la marche aléatoire symétrique.

(ii) Marche aléatoire sur $\mathbb{Z}$ avec barrières absorbantes. Soit $N \geq 1$ et considérons $\mathbb{E} = [\,-N, N\,] \cap \mathbb{Z}$. Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli symétriques sur $\{ -1, 1 \}$, et Y_0 une variable aléatoire indépendante de cette suite, à valeurs dans $\mathbb{E}$. On définit

$$\tau = \min\bigg\{ n \geq 0 \,:\, \bigg| \sum_{0\leq k\leq n} Y_k \bigg| = N \bigg\}.$$

La variable aléatoire τ est un temps d'arrêt relativement à la filtration naturelle $\mathcal{F}_n = \sigma(Y_0, \ldots, Y_n)$, $n \in \mathbb{N}$. Ce temps d'arrêt est fini presque sûrement puisque

$$\begin{aligned}
P\{\, \tau = \infty \,\} &= P\Big\{\, \forall n \in \mathbb{N} : \Big| \sum_{0 \leq k \leq n} Y_k \Big| < N \,\Big\} \\
&\leq P\Big\{\, \forall n \in \mathbb{N} : \Big| \sum_{1 \leq k \leq n} Y_k \Big| < 2N \,\Big\} \\
&\leq \inf_{n \in \mathbb{N}} P\Big\{\, \Big| \frac{1}{\sqrt{n}} \sum_{1 \leq k \leq n} Y_k \Big| < \frac{2N}{\sqrt{n}} \,\Big\} .
\end{aligned}$$

Cette dernière quantité est nulle d'après le théorème limite central (V.5.4).

Pour tout $n \geq 0$, définissons $X_n = S_{n \wedge \tau}$, où $S_n = \sum_{0 \leq k \leq n} Y_k$ et montrons que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov à valeurs dans $\mathbb{E}$.

Soient $i_0, \ldots, i_{n-2}$, i et j des éléments de $\mathbb{E}$. On veut calculer

$$P\{\, X_n = j \mid X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0 \,\} .$$

Distinguons deux cas. Si $|i| \neq N$,

$$\begin{aligned}
&P\Big\{\, X_n = j \,\Big|\, \{\, X_{n-1} = i \,\} \cap \bigcap_{0 \leq j \leq n-2} \{\, X_j = i_j \,\} \Big\} \\
&\qquad = P\Big\{\, Y_n = j - i \,\Big|\, \{X_{n-1} = i \,\} \cap \bigcap_{0 \leq j \leq n-2} \{\, X_j = i_j \,\} \Big\} \\
&\qquad = P\{\, Y_n = j - i \,\} \\
&\qquad = P\{\, X_n = j \mid X_{n-1} = i \,\} ,
\end{aligned}$$

et si $|i| = N$,

$$\begin{aligned}
&P\Big\{\, X_n = j \,\Big|\, \{\, X_{n-1} = i \,\} \cap \bigcap_{0 \leq j \leq n-2} \{\, X_j = i_j \,\} \Big\} \\
&\qquad\qquad = \delta_{ij} = P\{\, X_n = j \mid X_{n-1} = i \,\} .
\end{aligned}$$

(iii) Marche aléatoire sur $\mathbb{Z}/m\mathbb{Z}$. Dans cet exemple, $\mathbb{Z}/m\mathbb{Z}$ est identifié à l'ensemble $\{\, e^{2ik\pi/m} : k \in \{\, 0, 1, \ldots, m-1 \,\} \,\}$.

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi à valeurs dans $\{\, 0, 1, \ldots, p-1 \,\}$. On définit $X_0 = Y_0$, et pour tout $n \geq 1$,

$$X_n = Y_0 \exp\Big(\frac{2i\pi}{p} \sum_{1 \leq k \leq n} Y_k \Big) .$$

Puisque X_n est une fonction de $X_0, Y_1, \ldots, Y_n$, c'est une variable aléatoire $\sigma(Y_0, \ldots, Y_n)$-mesurable, et donc indépendante de Y_{n+1}. De plus,

$$X_{n+1} = X_n \exp(2i\pi Y_{n+1}/m)\,, \qquad n \in \mathbb{N}\,.$$

D'où pour tout $i_0, \ldots, i_n \in \{e^{2ik\pi/m} : k \in \{0, 1, \ldots, m-1\}\}$,

$$\begin{aligned} P\{\, X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0 \,\} \\ &= P\{\, i_{n-1} \exp\bigl(2i\pi Y_{n+1}/m\bigr) = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0 \,\} \\ &= P\Bigl\{ \exp\bigl(2i\pi Y_{n+1}/m\bigr) = \frac{i_n}{i_{n-1}} \Bigr\} \\ &= P\{\, X_n = i_n \mid X_{n-1} = i_{n-1} \,\}\,. \end{aligned}$$

La proposition suivante réexprime la propriété de Markov en montrant qu'elle équivaut d'une part à l'indépendance du présent au passé non immédiatement antérieur, d'autre part à l'indépendance du présent et futur au passé non immédiatement antérieur, et enfin à l'indépendance du futur et du passé du processus, conditionnellement à son présent.

***Proposition* VIII.1.3.** *Une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ définies sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$, est une chaîne de Markov si et seulement si l'une des trois propriétés équivalentes suivantes est vérifiée :*

(i) Pour tout $1 \leq k \leq n$ et tous $i_k, \ldots, i_n \in \mathbb{E}$ tels que

$$P\{\, X_{n-1} = i_{n-1}, \ldots, X_k = i_k \,\} > 0\,,$$

l'égalité suivante à lieu :

$$P\Bigl\{ X_n = i_n \Bigm| \bigcap_{k\leq j\leq n-1} \{X_j = i_j\} \Bigr\} = P\{\, X_n = i_n \mid X_{n-1} = i_{n-1} \,\}\,.$$

(ii) Pour tout $1 \leq k \leq n$, tout $m \geq 0$ et tous points $i_k, \ldots, i_{n+m} \in \mathbb{E}$ tels que $P\{\, X_{n-1} = i_{n-1}, \ldots, X_k = i_k \,\} > 0$,

$$\begin{aligned} P\Bigl(\bigcap_{0\leq j\leq m} X_{n+j} = i_{n+j} \Bigm| \bigcap_{k\leq j\leq n-1} \{X_j = i_j\} \Bigr) \\ = P\Bigl(\bigcap_{0\leq j\leq m} \{X_{n+j} = i_{n+j}\} \Bigm| X_{n-1} = i_{n-1} \Bigr)\,. \end{aligned}$$

(iii) Pour tous points $i_0, \ldots, i_{n+m} \in \mathbb{E}$ tels que $P\{ X_n = i_0 \} > 0$,

$$\begin{aligned} &P\Bigl(\bigcap_{1\le j\le m} \{X_{n+j} = i_{n+j}\} \cap \bigcap_{0\le j\le n-1} \{X_j = i_j\} \Bigm| X_n = i_n \Bigr) \\ &\qquad = P\Bigl(\bigcap_{1\le j\le m} \{X_j = i_j\} \Bigm| \{X_n = i_n\} \Bigr) P\Bigl(\bigcap_{0\le j\le n-1} \{X_j = i_j\} \Bigm| X_n = i_n \Bigr). \end{aligned}$$

Démonstration. (i) En prenant $k = 0$, on voit que la condition est clairement suffisante. Pour montrer sa nécessité, par définition d'une chaîne de Markov,

$$\begin{aligned} P\{ X_n = i_n \mid X_{n-1} = i_{n-1} \} &= P\Bigl\{ X_n = i_n \mid \bigcap_{0\le j\le n-1} \{ X_j = i_j \} \Bigr\} \\ &= \frac{P\bigl(\bigcap_{0\le j\le n}\{ X_j = i_j \}\bigr)}{P\bigl(\bigcap_{0\le j\le n-1}\{ X_j = i_j \}\bigr)}. \end{aligned} \tag{2}$$

Par récurrence, on montre que si $A_0, \ldots, A_n$ sont des événements tels que $P\bigl(\bigcap_{0\le j\le n-1} A_j\bigr) > 0$, alors

$$P\Bigl(\bigcap_{0\le j\le n} A_j \Bigr) = \prod_{1\le l\le n} P\Bigl(A_j \Bigm| \bigcap_{0\le j\le l-1} A_j \Bigr) P(A_0). \tag{3}$$

En utilisant cette formule avec $A_j = \{ X_j = i_j \}$, en vertu de (2),

$$\begin{aligned} &P\{ X_n = i_n \mid X_{n-1} = i_{n-1} \} \\ &\quad = \frac{\prod_{1\le l\le n} P\Bigl\{ X_l = i_l \Bigm| \bigcap_{0\le j\le l-1}\{ X_j = i_j \} \Bigr\}}{\prod_{1\le l\le n-1} P\Bigl\{ X_l = i_l \Bigm| \bigcap_{0\le j\le l-1}\{ X_j = i_j \} \Bigr\}} \\ &\quad = \frac{\prod_{k+1\le l\le n} P\Bigl\{ X_k = i_k \Bigm| \bigcap_{0\le j\le l-1}\{ X_j = i_j \} \Bigr\} P\{ X_k = i_k \}}{\prod_{k+1\le l\le n-1} P\Bigl\{ X_k = i_k \Bigm| \bigcap_{0\le j\le l-1}\{ X_j = i_j \} \Bigr\} P\{ X_k = i_k \}}. \end{aligned}$$

En utilisant encore (3), le membre de droite de l'égalité précédente est

$$\frac{P\{ X_n = i_n, \ldots, X_k = i_k \}}{P\{ X_{n-1} = i_{n-1}, \ldots, X_k = i_k \}} = P\Bigl\{ X_n = i_n \Bigm| \bigcap_{k\le j\le n-1} X_j = i_j \Bigr\}.$$

(ii) Là encore, la condition est clairement suffisante en prenant $m = 0$ et $k = 0$. Pour montrer qu'elle est nécessaire, observons d'abord que, par σ-additivité de la mesure de probabilité, la positivité de $P\{ X_{n-1} = i_{n-1}, \ldots, X_k = i_k \}$ entraîne qu'il existe $i_0, \ldots, i_{k-1}$ tels que $P\{\, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0 \,\} > 0$. Pour un tel $(i_0, \ldots, i_{n-1})$,

$$\begin{aligned}
&P\Big(\bigcap_{n \leq j \leq n+m} \{X_j = i_j\} \,\Big|\, X_{n-1} = i_{n-1} \Big) \\
&\qquad\qquad = \frac{P\Big(\bigcap_{n-1 \leq j \leq n+m} \{X_j = i_j\} \Big)}{P\{\, X_{n-1} = i_{n-1} \,\}} \\
&\qquad\qquad = \prod_{n \leq l \leq n+m} P\Big\{\, X_l = i_l \,\Big|\, \bigcap_{n-1 \leq j \leq l-1} \{X_j = i_j\} \Big\}.
\end{aligned}$$

D'après le premier point, ce produit est égal à (pour tout $k \leq n$)

$$\prod_{n \leq l \leq n+m} P\Big\{\, X_l = i_l \,\Big|\, \bigcap_{k \leq j \leq l-1} \{X_j = i_j\} \Big\},$$

c'est-à-dire

$$P\Big(\bigcap_{n \leq l \leq n+m} \{X_l = i_l\} \,\Big|\, \bigcap_{k \leq j \leq n-1} \{X_j = i_j\} \Big).$$

(iii) La condition est nécessaire puisque

$$\begin{aligned}
&P\Big(\bigcap_{n+1 \leq j \leq n+m} \{X_j = i_j\} \cap \bigcap_{0 \leq j \leq n-1} \{X_j = i_j\} \,\Big|\, X_n = i_n \Big) \\
&\quad = \frac{P\Big(\bigcap_{0 \leq j \leq n+m} \{X_j = i_j\} \Big)}{P\{\, X_n = i_n \,\}} \\
&\quad = \frac{P\Big(\bigcap_{n+1 \leq j \leq n+m} \{X_j = i_j\} \,\Big|\, \bigcap_{0 \leq j \leq n} \{X_j = i_j\} \Big) P\Big(\bigcap_{0 \leq j \leq n} \{X_j = i_j\} \Big)}{P\{X_n = i_n\}}.
\end{aligned}$$

Or, d'après le point (ii), ce rapport est égal à

$$\begin{aligned}
&\frac{P\Big(\bigcap_{n+1 \leq j \leq n+m} \{\, X_j = i_j \,\} \,\Big|\, X_n = i_n \Big) P\Big(\bigcap_{0 \leq j \leq n} \{\, X_j = i_j \,\} \Big)}{P\{\, X_n = i_n \,\}} \\
&\qquad = P\Big(\bigcap_{n+1 \leq j \leq n+m} \{X_j = i_j\} \,\Big|\, X_n = i_n \Big) P\Big(\bigcap_{0 \leq j \leq n-1} \{\, X_j = i_j \,\} \,\Big|\, X_n = i_n \Big).
\end{aligned}$$

Si maintenant la condition est vérifiée,

$$
\begin{aligned}
&P\Big\{\, X_{n+1} = i_{n+1} \;\Big|\; \bigcap_{0\leq j\leq n} \{\, X_j = i_j \,\}\Big\}\\
&\quad = \frac{P\Big(\bigcap_{0\leq j\leq n+1}\{\, X_j = i_j \,\}\Big)}{P\Big(\bigcap_{0\leq j\leq n}\{\, X_j = i_j \,\}\Big)}\\
&\quad = \frac{P\Big(\{\, X_{n+1} = i_{n+1} \,\} \cap \bigcap_{0\leq j\leq n-1}\{\, X_j = i_j \,\} \;\Big|\; X_n = i_n\Big) P\{\, X_n = i_n \,\}}{P\Big(\bigcap_{0\leq j\leq n-1}\{\, X_j = i_j \,\} \;\Big|\; X_n = i_n\Big) P\{\, X_n = i_n \,\}}\\
&\quad = P\{\, X_{n+1} = i_{n+1} \mid X_n = i_n \,\}\,. \qquad \square
\end{aligned}
$$

***Définition* VIII.1.4.** On dit qu'une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ est homogène si, pour tout couple (i,j) de points de $\mathbb{E}$, $P\{\, X_{n+1} = j \mid X_n = i \,\}$ est indépendant de n, n décrivant l'ensemble des entiers pour lesquels $P\{\, X_n = i \,\} > 0$.

Observons que pour un état i donné, si l'ensemble des entiers n pour lesquels $P\{\, X_n = i \,\} \neq 0$ est vide, la chaîne est à valeurs dans $\mathbb{E} \setminus \{\, i \,\}$ avec probabilité 1. On peut donc, en remplaçant au besoin $\mathbb{E}$ par $\mathbb{E} \setminus \{\, i \,\}$, supposer que ce cas ne se produit pas. On note alors P_{ij} la valeur commune des $P\{\, X_{n+1} = j \mid X_n = i \,\}$ et $\mathbb{P} = (P_{ij})_{i,j\in\mathbb{E}}$. La matrice $\mathbb{P}$ est appelée matrice de transition de la chaîne (nous utilisons encore le terme de matrice lorsque $\mathbb{E}$ est infini).

***Définition* VIII.1.5.** Une matrice $M = (M_{i,j})_{i,j\in\mathbb{E}}$ (éventuellement de taille infinie) est une matrice stochastique si elle vérifie

(i) $M_{ij} \geq 0$ pour tous $i, j \in \mathbb{E}$,

(ii) $\sum_{j\in\mathbb{E}} M_{ij} = 1$ pour tout $i \in \mathbb{E}$.

Ainsi, la matrice de transition d'une chaîne de Markov est une matrice stochastique.

***Exemples* VIII.1.6.** (i) La marche aléatoire sur $\mathbb{Z}/m\mathbb{Z}$ est homogène, et sa matrice de transition est

$$
\mathbb{P} = \begin{pmatrix}
q_0 & q_1 & \cdots & q_{p-1}\\
q_{p-1} & q_0 & & \vdots\\
\vdots & & \ddots & \\
q_1 & \cdots & & q_0
\end{pmatrix}.
$$

(ii) Chaîne d'Ehrenfest. Soit d un entier supérieur ou égal à 1. On répartit d boules numérotées dans deux urnes A et B. On tire un nombre i au hasard (c'est-à-dire suivant la probabilité uniforme) entre 1 et d et on change la boule numérotée i d'urne. Soit X_n^d le nombre de boules dans l'urne A après n tirages indépendants. La suite $(X_n^d)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène à valeurs dans l'ensemble $\mathbb{E} = \{0, \ldots, d\}$ et sa matrice de transition, $\mathbb{P}$, est donnée par la formule

$$P_{d,i,i+1} = \frac{d-i}{d} \quad ; \quad P_{d,i+1,i} = \frac{i+1}{d}, \quad i \in \{1, \ldots, d-1\}.$$

Proposition VIII.1.7. *Le produit de deux matrices stochastiques est encore une matrice stochastique.*

Démonstration. Soient $\mathbb{P}$ et $\mathbb{Q}$ deux matrices stochastiques. Leur produit est bien défini. En effet, la série $(\mathbb{PQ})_{ij} = \sum_{l\in E} P_{il}Q_{lj}$ converge puisque la série $\sum_{l\in E} P_{il}$ converge et que les $(Q_{lj})_{l,j\in\mathbb{E}}$ sont bornés par 1. Clairement, $\mathbb{PQ}$ est une matrice à coefficients positifs, et de plus,

$$\sum_{j\in\mathbb{E}}(PQ)_{ij} = \sum_{j\in\mathbb{E}}\sum_{l\in\mathbb{E}} P_{il}Q_{lj} = \sum_{l\in\mathbb{E}}\sum_{j\in\mathbb{E}} P_{il}Q_{lj} = \sum_{l\in\mathbb{E}} P_{il}\sum_{j\in\mathbb{E}} Q_{lj} = \sum_{l\in\mathbb{E}} P_{il} = 1\,. \quad \square$$

Remarquons alors que la suite $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène de matrice de transition $\mathbb{P}$ si et seulement si, pour tous $i_0, \ldots, i_n \in \mathbb{E}$,

$$P\Big(\bigcap_{0\leq k\leq n}\{X_k = i_k\}\Big) = P_{i_{n-1}i_n} P\Big(\bigcap_{0\leq k\leq n-1}\{X_k = i_k\}\Big).$$

VIII.2. Calcul des lois marginales

La proposition suivante prolonge la dernière remarque de la section précédente, en montrant que la donnée de la matrice de transition et de la loi initiale suffit à caractériser la loi de la chaîne jusqu'à tout instant fixé.

Proposition VIII.2.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène, définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$, de matrice de transition $\mathbb{P}$ et de loi initiale μ_0. Alors, pour tout $n \geq 1$ et tous $i_0, \ldots, i_n \in \mathbb{E}$,*

$$P\{X_0 = i_0, \ldots, X_n = i_n\} = \mu_0\big(\{i_0\}\big)P_{i_0i_1}\cdots P_{i_{n-1}i_n}\,.$$

Démonstration. Elle se fait par récurrence sur n. La propriété est vraie pour $n = 0$ par définition de μ_0. Supposons la vraie au rang $n-1$. Distinguons deux cas :

(i) Si $P\{ X_0 = i_0, \ldots, X_{n-1} = i_{n-1} \} = 0$, il résulte de l'hypothèse de récurrence que $\mu_0(\{ i_0 \})P_{i_0 i_1} \cdots P_{i_{n-2} i_{n-1}} = 0$, et donc

$$\mu_0\big(\{ i_0 \}\big)P_{i_0,i_1} \cdots P_{i_{n-1} i_n} = 0 .$$

Or $P\{ X_0 = i_0, \ldots, X_n = i_n \} = 0$ et la propriété est vraie dans ce cas.
(ii) Si maintenant $P\{ X_0 = i_0, \ldots, X_{n-1} = i_{n-1} \} > 0$, il vient

$$\begin{aligned}
&P\{ X_0 = i_0, \ldots, X_{n-1} = i_n \} \\
&\quad = P\{ X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0 \} P\{ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0 \} \\
&\quad = P\Big(X_n = i_n \,\Big|\, \bigcap_{0 \leq j \leq n-1} \{ X_j = i_j \} \Big) P\Big(\bigcap_{0 \leq j \leq n-1} \{ X_j = i_j \} \Big) \\
&\quad = P\Big\{ X_n = i_n \mid X_{n-1} = i_{n-1} \Big\} \mu_0\big(\{ i_0 \}\big) P_{i_0 i_1} \cdots P_{i_{n-2} i_{n-1}} \\
&\quad = \mu_0\big(\{ i_0 \}\big) P_{i_0 i_1} \cdots P_{i_{n-1} i_n} ,
\end{aligned}$$

ce qui démontre la proposition. □

D'après la proposition VIII.1.7, la matrice $\mathbb{P}^n = \mathbb{P} \times \cdots \times \mathbb{P}$ (n fois) est une matrice stochastique. On notera $P^n_{i,j}$ ses éléments.

Corollaire VIII.2.2. *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. Pour tous entiers n, m et tous états $i, j \in \mathbb{E}$,*

(i) $P\{ X_n = j \} = \sum_{k \in E} \mu_0\big(\{ k \}\big) P^n_{kj}$;

(ii) $P\{ X_{m+n} = j \mid X_m = i \} = P^n_{ij}$;

(iii) $P\{ X_{m+n} = j \mid X_0 = i \} = \sum_{k \in E} P\{ X_m = j \mid X_0 = k \} P\{ X_n = k \mid X_0 = i \}$.

L'égalité (iii) est appelée équation de Chapman-Kolmogorov.

VIII.3. Généralisation de la propriété de Markov

Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. On peut voir X comme un élément de l'espace des suites sur $\mathbb{E}$, $\mathbb{E}^{\mathbb{N}} = \{ x = (x_n)_{n \in \mathbb{N}} : x_n \in \mathbb{E} \}$. Sur $\mathbb{E}^{\mathbb{N}}$, considérons la tribu cylindrique $\mathcal{B}$, c'est-à-dire la tribu engendrée par les parties (cylindres) de la forme

$$B_0 \times \cdots \times B_n \times \mathbb{E} \times \mathbb{E} \times \cdots , \quad B_1, \ldots, B_n \in \mathcal{P}(\mathbb{E}) , \, n \in \mathbb{N} .$$

La fonction X est mesurable de $(\Omega, \mathcal{A}, P)$ dans $(\mathbb{E}^{\mathbb{N}}, \mathcal{B})$ d'après la proposition I.1.14, puisque si $C = B_0 \times \cdots \times B_n \times \mathbb{E} \times \mathbb{E} \times \cdots$ est un cylindre,

$$X^{-1}(C) = \bigcap_{0 \leq i \leq n} X^{-1}(B_i) \in \mathcal{A}\,.$$

On peut donc parler de la loi de X, c'est-à-dire de la mesure image P^X de P par X (*cf.* III.1.7).

Sur $\mathbb{E}^{\mathbb{N}}$, on définit l'opérateur de translation (ou décalage),

$$\theta \;:\; x = (x_n)_{n \in \mathbb{N}} \in \mathbb{E}^{\mathbb{N}} \mapsto \theta(x) = (x_{n+1})_{n \in \mathbb{N}} \in \mathbb{E}^{\mathbb{N}}\,.$$

Autrement dit, $(\theta(x))_n = x_{n+1}$, $n \in \mathbb{N}$. On peut définir les itérés, θ^k, par $\theta^k = \theta \circ \theta^{k-1}$, $k \geq 1$, ce qui donne $\theta^k(x) = (x_{n+k})_{n \in \mathbb{N}}$.

Le théorème suivant montre qu'une chaîne de Markov homogène considérée à partir de l'instant n et conditionellement à X_n, a même loi que la chaîne initialisée à $X_0 = 0$.

Théorème VIII.3.1. *Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène, définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. Alors, pour tous les états $i_0, \ldots, i_k \in \mathbb{E}$,*

$$\mathcal{L}\Big(\theta^k(X) \;\Big|\; \bigcap_{0 \leq j \leq k} \{\, X_j = i_j \,\}\Big) = \mathcal{L}\big(X \mid X_0 = i_k\big)\,.$$

Démonstration. Si $C = B_0 \times \cdots \times B_n \times \mathbb{E} \times \mathbb{E} \times \cdots$ est un cylindre,

$$\begin{aligned}
&P\Big(\{\, \theta^k(X) \in C \,\} \;\Big|\; \bigcap_{0 \leq j \leq k} \{\, X_j = i_j \,\}\Big) \\
&\qquad = P\big\{\, X_k \in B_0, \ldots, X_{k+n} \in B_n \;\big|\; X_0 = i_0, \ldots, X_k = i_k \,\big\} \\
&\qquad = \sum_{(j_0, \ldots, j_n) \in B_0 \times \cdots \times B_n} P\big\{\, X_k = j_0, \ldots, X_{k+n} = j_n \;\big|\; X_k = i_k \,\big\} \\
&\qquad = P\{\, \theta_k(X) \in C \mid X_k = i_k \,\} \\
&\qquad = \sum_{(j_0, \ldots, j_n) \in B_0 \times \cdots \times B_n} \mathbb{1}_{\{i_k\}}(j_0) P_{j_0, j_1} P_{j_1, j_2} \ldots P_{j_{n-1}, j_n} \\
&\qquad = \sum_{(j_0, \ldots, j_n) \in B_0 \times \cdots \times B_n} \mathbb{1}_{\{i_k\}}(j_0) P\big\{\, X_1 = j_1, \ldots, X_n = j_n \;\big|\; X_0 = j_0 \,\big\}\,.
\end{aligned}$$

Donc les lois considérées coïncident sur les cylindres. En observant qu'une union de cylindres se décompose en une union disjointe de cylindres (puisque l'intersection de deux cylindres est un cylindre), on voit que les lois considérées coïncident sur l'algèbre de Boole engendrée par les cylindres. Donc elles sont égales d'après la proposition I.4.7. □

Le contenu du théorème précédent est essentiellement que la loi de la trajectoire d'un processus de Markov homogène après le temps n est donnée par la loi de la chaîne au temps n, pour peu bien sûr que la matrice de transition soit fixée. Il est remarquable, et cela sera utile pour la suite, que ce résultat reste vrai si l'on considère un temps non plus fixe mais aléatoire, pour peu que celui-ci soit un temps d'arrêt. C'est ce qu'exprime le théorème suivant.

Théorème VIII.3.2 **(Propriété de Markov forte).** *Soit* $X = (X_n)_{n\in\mathbb{N}}$ *une chaîne de Markov définie sur* $(\Omega, \mathcal{A}, P)$, *à valeurs dans* $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. *Soit* T *un temps d'arrêt pour la filtration* $\big(\mathcal{F}_n = \sigma(X_0, \ldots, X_n)\big)_{n\in\mathbb{N}}$. *Sur l'événement* $\{\, X_T = i \,\} \cap \{\, T < \infty \,\}$,

$$\mathcal{L}\big(\theta^T(X) \mid \mathcal{F}_T\big) = \mathcal{L}\big(X \mid X_0 = i\big)\,.$$

Démonstration. Comme pour démontrer le théorème VIII.3.1, soit un cylindre

$$C = B_0 \times \cdots \times B_n \times \mathbb{E} \times \mathbb{E} \times \cdots$$

Sur $\{\, X_T = i \,\} \cap \{\, T < \infty \,\}$, il vient

$$\begin{aligned} P\big\{\, \theta^T(X) \in C \;\big|\; X_T \,\big\} &= \sum_{n\in\mathbb{N}} P\big\{\, \theta^T(X) \in C \;\big|\; X_n = i \,\big\} \mathbb{1}_{\{n\}}(T) \\ &= \sum_{n\in\mathbb{N}} P\big\{\, \theta^n(X) \in C \mid X_n = i \,\big\} \mathbb{1}_{\{n\}}(T). \end{aligned}$$

En utilisant le théorème VIII.3.1, on en déduit

$$\begin{aligned} P\big\{\, \theta^T(X) \in C \;\big|\; X_T \,\big\} &= \sum_{n\in\mathbb{N}} P\big\{\, X \in C \;\big|\; X_0 = i \,\big\} \mathbb{1}_{\{n\}}(T) \\ &= P\big\{\, X \in C \mid X_0 = i \,\big\}\,. \end{aligned}$$
□

On remarquera que l'argument final dans la démonstration du théorème VIII.3.1 permet d'étendre la proposition VIII.2.1. On voit en effet que la loi d'une chaîne de Markov homogène (en tant qu'élément de $\mathbb{E}^{\mathbb{N}}$) est entièrement déterminée par la donnée de la mesure initiale et de la matrice de transition.

VIII.4. Comportement asymptotique. Mesures invariantes

L'objet des paragraphes suivants est l'étude de la convergence en loi d'une chaîne de Markov : le système qui évolue selon cette chaîne converge-t-il vers un « état d'équilibre » ?

Il y a plusieurs méthodes pour résoudre ce problème, en particulier une méthode de type totalement algébrique lorsque $\mathbb{E}$ est fini et une méthode totalement probabiliste dans le cas général ; nous avons choisi d'exposer en partie ces deux méthodes.

Dans toute la suite du chapitre, $\mathbb{P}$ est une matrice stochastique. On note $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène, définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$, de matrice de transition $\mathbb{P}$ et de mesure initiale μ. L'ensemble $\mathbb{E}^{\mathbb{N}}$ est muni de sa tribu cylindrique et de la probabilité image $P_{\mu,\mathbb{P}}$ de P par la chaîne X. Dans ce qui suit, la matrice $\mathbb{P}$ est fixe et on se contentera de noter P_μ, μ n'étant pas fixée.

***Notation* VIII.4.1.** *Si μ est une probabilité sur $\mathbb{E}$, pour tout $i \in \mathbb{E}$, on note $\mu_i = \mu(\{\, i \,\})$. On désignera encore par μ le vecteur de composantes $(\mu_i)_{i\in\mathbb{E}}$.*

En particulier, puisque μ est vue comme un vecteur, ${}^{\mathrm{t}}\mathbb{P}\mu$ est un vecteur ; il est associé à une mesure aussi notée ${}^{\mathrm{t}}\mathbb{P}\mu$.

***Définition* VIII.4.2.** On dit que μ, probabilité sur $\mathbb{E}$, est une mesure asymptotique de la chaîne $(X_n)_{n\in\mathbb{N}}$ s'il existe une probabilité μ_0 sur $\mathbb{E}$ telle que si μ_0 est la loi de X_0, alors $(X_n)_{n\in\mathbb{N}}$ converge en loi vers μ.

Une mesure asymptotique est donc une probabilité.

Le but de ce chapitre est de déterminer les mesures asymptotiques d'une chaîne de Markov et leur dépendance par rapport à la loi initiale. La notion déterminante dans cette recherche est celle de mesure invariante.

***Définition* VIII.4.3.** On dit que μ, mesure positive sur $\mathbb{E}$, est une mesure invariante de la chaîne si ${}^{\mathrm{t}}\mathbb{P}\mu = \mu$.

On prendra garde au fait que μ n'est pas nécessairement une probabilité, puisque l'on permet $\mu(\mathbb{E}) \neq 1$. Observons que si μ est une mesure invariante de la chaîne et que si X_0 est de loi μ, alors $P_\mu\{\, X_1 = j \,\} = \sum_{i\in\mathbb{E}} P_{i,j}\mu_i = \mu_j$ pour tout $j \in \mathbb{E}$. Donc X_1 est aussi de loi μ, et par récurrence, X_n est de loi μ pour tout $n \in \mathbb{N}$. Ceci justifie la terminologie.

***Proposition* VIII.4.4.** *Soit μ une probabilité sur $\mathbb{E}$. Les assertions suivantes sont équivalentes.*

(i) μ est une mesure asymptotique de la chaîne ;

(ii) μ est une mesure invariante de la chaîne ;

(iii) $\mathcal{L}(X_0) = \mu \Rightarrow \mathcal{L}(X_n) = \mu$ pour tout $n \in \mathbb{N}$.

Démonstration. Il est clair que (iii) et (ii) sont équivalentes et que (ii) implique (i). Montrons que (i) implique (ii). Supposons μ asymptotique ; il existe donc une probabilité μ_0 telle que

$$\lim_{n\to\infty} P_{\mu_0}\{\, X_n = j \,\} = \mu_j \qquad \text{pour tout } j \in \mathbb{E}\,,$$

ce qui s'écrit encore

$$\lim_{n\to\infty} \sum_{i\in\mathbb{E}} \mu_{0,i} P^n_{ij} = \mu_j \qquad \text{pour tout } j \in \mathbb{E}\,.$$

Soit $i \in \mathbb{E}$. Il résulte du lemme de Fatou (II.2.3) que

$$\begin{aligned}
\sum_{j\in\mathbb{E}} P_{ji}\mu_j &= \sum_{j\in\mathbb{E}} P_{ji} \lim_{n\to\infty} \sum_{k\in\mathbb{E}} \mu_{0,k} P^n_{kj} \\
&\le \liminf_n \sum_{j\in\mathbb{E}} \sum_{k\in\mathbb{E}} P_{ji}\mu_{0,k} P^n_{kj} \\
&= \liminf_n \sum_{k\in\mathbb{E}} P^{n+1}_{ki} \mu_{0,k} \\
&= \mu_i\,.
\end{aligned}$$

Puisque $\sum_{i\in\mathbb{E}} (\sum_{j\in\mathbb{E}} P_{ji}\mu_j) = \sum_{i\in\mathbb{E}} \mu_i = 1$, on en déduit que $\sum_{j\in\mathbb{E}} P_{ji}\mu_j = \mu_i$ pour tout élément i de $\mathbb{E}$. □

***Exemples* VIII.4.5.** Nous reprenons les exemples donnés en VIII.1.2.

(i) Marches aléatoires sur $\mathbb{Z}$. Une mesure μ est une mesure invariante de la marche aléatoire de paramètre p sur $\mathbb{Z}$ si et seulement si pour tout $i \in \mathbb{Z}$,

$$p\mu_{i+1} + (1-p)\mu_{i-1} = \mu_i\,,$$

ce qui est équivalent à

$$p(\mu_{i+1} - \mu_i) = (1-p)(\mu_i - \mu_{i-1})\,.$$

On en déduit que

$$\mu_{i+1} - \mu_i = \Big(\frac{1-p}{p}\Big)^i (\mu_1 - \mu_0)$$

pour tout $i \in \mathbb{Z}$, et par suite, pour $i \geq 1$,

$$\mu_i - \mu_0 = \sum_{k=0}^{i-1} \Big(\frac{1-p}{p}\Big)^k (\mu_1 - \mu_0)$$

et

$$\mu_{-i} - \mu_0 = -\sum_{k=1}^{i} \Big(\frac{p}{1-p}\Big)^k (\mu_1 - \mu_0)\,.$$

Donc, si $p = 1/2$, on a $\mu_i = \mu_0 + i(\mu_1 - \mu_0)$ pour tout $i \in \mathbb{Z}$. Une telle mesure μ n'est positive que si $\mu_1 = \mu_0$. Les mesures invariantes, dans ce cas, sont donc les multiples de la mesure de comptage sur $\mathbb{Z}$.

Si $p \neq 1/2$,

$$\mu_i = \mu_0 + \frac{p}{1-2p}\bigg(\Big(\frac{1-p}{p}\Big)^i - 1\bigg)(\mu_1 - \mu_0)$$

et

$$\mu_{-i} = \mu_0 + \frac{p}{2p-1}\bigg(1 - \Big(\frac{p}{1-p}\Big)^{i-1}\bigg)(\mu_1 - \mu_0)$$

pour tout $i \geq 1$. Donc, la mesure μ est positive si et seulement si

$$\begin{cases} \mu_0 \leq \mu_1 \leq \dfrac{1-p}{p}\mu_0 & \text{si } p < 1/2 \\ \dfrac{1-p}{p}\mu_0 \leq \mu_1 \leq \mu_0 & \text{si } p > 1/2. \end{cases}$$

La chaîne n'admet donc aucune probabilité invariante.

(ii) Marche aléatoire sur $\mathbb{Z}$ avec barrières absorbantes. Une mesure μ est invariante pour cette marche si et seulement si

$$\begin{aligned} \tfrac{1}{2}\mu_{i+1} + \tfrac{1}{2}\mu_{i-1} = \mu_i \quad & \text{si} \quad i \in [\,-N+2, N-2\,] \cap \mathbb{Z}\,, \\ \mu_{N-1} = \tfrac{1}{2}\mu_{N-2} \quad & \text{et} \quad \mu_{-N+1} = \tfrac{1}{2}\mu_{-N+2}\,, \\ \mu_N = \tfrac{1}{2}\mu_{N-1} + \mu_N \quad & \text{et} \quad \mu_{-N} = \tfrac{1}{2}\mu_{-N+1} + \mu_{-N}\,, \end{aligned}$$

c'est-à-dire si et seulement si $\mu_i = 0$ pour tout $i \in [\,-N+1, N-1\,] \cap \mathbb{Z}$. Les mesures invariantes sont donc les mesures positives portées par $\{\,-N, N\,\}$.

(iii) Marches aléatoires sur $\mathbb{Z}/m\mathbb{Z}$. Une mesure μ est invariante si et seulement si pour tout $i \in \{0, \ldots, p-1\}$, $\mu_i = \sum_{0 \le k \le p-1} q_{p-k}\mu_k$, c'est-à-dire si μ est la mesure uniforme.

(iv) Chaîne d'Ehrenfest. Une mesure μ est invariante pour la chaîne d'Ehrenfest si et seulement si

$$\begin{cases} \mu_i = \dfrac{d-i+1}{d}\mu_{i-1} + \dfrac{i+1}{d}\mu_{k+1} \quad \text{si } 0 < k < d, \\ \mu_0 = \dfrac{1}{d} \quad \text{et} \quad \mu_d = \dfrac{1}{d}\,\mu_{d-1}\,. \end{cases}$$

On en déduit aisément par récurrence que $\mu_i = \mathrm{C}_d^i\,\mu_0$, pour tout $i \le d$. Donc, à une constante de proportionnalité près, μ est la loi binomiale $\mathcal{B}(d, \frac{1}{2})$.

Dans la suite, nous allons montrer comme le suggèrent ces exemples que, lorsque $\mathbb{E}$ est fini, il y a toujours au moins une mesure (et par suite une probabilité) invariante. Le cas où $\mathbb{E}$ est infini est plus complexe ; tous les cas sont possibles : aucune mesure invariante, une ou des mesures invariantes mais pas de probabilité invariante, une ou des probabilités invariantes.

Théorème VIII.4.6. *Toute chaîne de Markov homogène à valeurs dans un ensemble fini admet au moins une mesure invariante.*

Démonstration. Nous donnons deux méthodes de démonstration, l'une topologique, l'autre algébrique.
Méthode topologique. On note $\mathcal{M}_1(\mathbb{E})$ l'ensemble des probabilités sur $\mathbb{E}$, *i.e.*

$$\mathcal{M}_1(\mathbb{E}) = \left\{ \mu = (\mu_i)_{i\in\mathbb{E}} \ : \ \forall i \in \mathbb{E}\,,\ \mu_i \ge 0\,,\ \text{ et } \sum_{i\in\mathbb{E}} \mu_i = 1 \right\}.$$

Puisque $\mathbb{E}$ est fini, $\mathcal{M}_1(\mathbb{E})$ est un compact de $\mathbb{R}^{\mathrm{card}\mathbb{E}}$. Soit $\mu_0 \in \mathcal{M}_1(E)$; pour tout $n \ge 1$, on définit

$$\mu_n = \frac{\mu_0 + {}^{\mathrm{t}}\mathbb{P}\mu_0 + \ldots + {}^{\mathrm{t}}\mathbb{P}^n\mu_0}{n+1}\,.$$

La suite $(\mu_n)_{n\in\mathbb{N}}$ d'éléments du compact $\mathcal{M}_1(\mathbb{E})$ admet une sous-suite convergente $(\mu_{n_k})_{k\in\mathbb{N}}$. Soit μ la limite de cette sous-suite ; c'est une mesure invariante puisque

$${}^{\mathrm{t}}\mathbb{P}\mu - \mu = \lim_{k\to\infty} ({}^{\mathrm{t}}\mathbb{P}\mu_{n_k} - \mu_{n_k}) = \lim_{k\to\infty} \frac{{}^{\mathrm{t}}\mathbb{P}^{n_k+1}\mu_0 - \mu_0}{n_k+1} = 0\,.$$

Méthode algébrique. La somme des colonnes de la matrice ${}^{\mathrm{t}}\mathbb{P} - I$ est nulle. La matrice ${}^{\mathrm{t}}\mathbb{P}$ admet donc 1 comme valeur propre. Le résultat cherché résulte du lemme suivant.

***Lemme VIII.4.7* (de Perron-Froebenius).** *Soit $\mathbb{P}$ une matrice stochastique de dimension n et v un vecteur propre complexe de la matrice ${}^{\mathrm{t}}\mathbb{P}$ associé à la valeur propre λ de module 1 ; on note w le vecteur de composantes $(|v_1|, \ldots, |v_N|)$. Alors, w est un vecteur propre de la matrice ${}^{\mathrm{t}}\mathbb{P}$ associé à la valeur propre 1.*

Démonstration. Pour tout $i \in \mathbb{E}$, soit $\alpha_i = ({}^{\mathrm{t}}\mathbb{P}w - w)_i$. On a

$$\alpha_i = \sum_{j \in \mathbb{E}} P_{ji} w_j - w_i = \sum_{j \in \mathbb{E}} P_{ji} |v_j| - |v_i| \geq \left| \sum_{j \in \mathbb{E}} P_{ij} v_j \right| - |v_i| = 0$$

car ${}^{\mathrm{t}}\mathbb{P}v = \lambda v$. D'autre part,

$$\sum_{i \in \mathbb{E}} \alpha_i = \sum_i \Big(\sum_{j \in \mathbb{E}} P_{ji} w_j - w_i \Big) = 0\,.$$

On en déduit que $\alpha_i = 0$ pour tout $i \in \mathbb{E}$. □

Remarque VIII.4.8. Nous pouvons reformuler ce lemme en disant qu'une matrice stochastique indexée par un ensemble fini admet toujours un vecteur propre associé à la valeur propre 1 dont les composantes sont positives.

Le problème de l'unicité de la mesure invariante demande une étude plus fine des communications entre états.

Exemple VIII.4.9. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène à valeurs dans l'ensemble $\mathbb{E} = \{1, \ldots, 5\}$ de matrice de transition

$$\begin{pmatrix} 1/2 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

La matrice de transition étant une matrice diagonale par blocs, l'équation ${}^{\mathrm{t}}\mathbb{P}\mu = \mu$ est donc équivalente à un couple de systèmes autonomes, le premier ne faisant intervenir que les variables μ_1, μ_2, μ_3, le second μ_4, μ_5. Il n'y a donc pas unicité de la mesure invariante. Plus précisément, μ est une mesure invariante si et seulement si

$$\mu_1 = \mu_3\,, \qquad \mu_2 = 0\,, \qquad \mu_4 = \mu_5\,.$$

Les mesures invariantes de la chaîne sont donc les combinaisons linéaires à coefficients positifs des mesures uniformes sur les ensembles $\{1, 3\}$ et $\{4, 5\}$. Le point 2 n'est pas chargé par les mesures invariantes, ce qui provient du fait qu'il n'est pas accessible par la chaîne si $X_0 \in \{1, 3, 4, 5\}$. Intuitivement, même si la chaîne démarre dans l'état 2, elle finira par en partir et n'y reviendra plus.

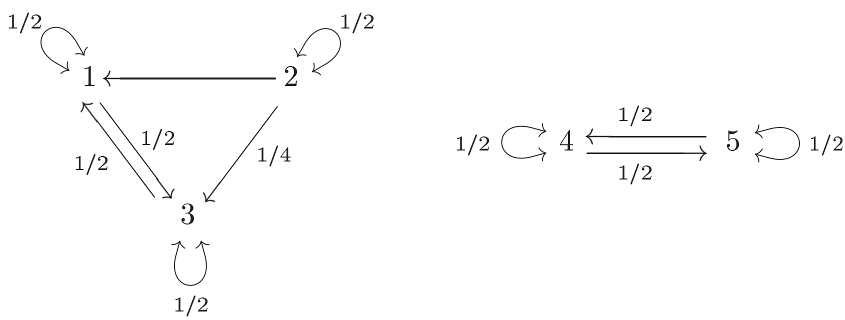
Définition VIII.4.10. Soient i et j deux éléments de $\mathbb{E}$. On dit que i conduit à j, noté $i \to j$, s'il existe $n > 0$ tel que $P^n_{ij} > 0$; on dit que i et j communiquent, noté $i \leftrightarrow j$, si i conduit à j et j conduit à i.

La relation $\leftrightarrow$ est symétrique et transitive ; elle est réflexive sur le sous-ensemble de $\mathbb{E}$, noté $\mathbb{E}'$, des éléments qui communiquent avec un autre état (qui peut être lui-même).

On appelle classe de la chaîne, soit un singleton de $\mathbb{E} \setminus \mathbb{E}'$, soit une classe d'équivalence de la relation $\leftrightarrow$ restreinte à $\mathbb{E}'$.

On peut représenter cette relation de communication entre états par un graphe, dit graphe de Markov, dont les sommets sont les points de $\mathbb{E}$, deux points de $\mathbb{E}$ étant joints par une arête si l'un d'eux conduit à l'autre. L'arête joignant i à j est orientée de i à j si i conduit à j.

Exemple VIII.4.11. Dessinons le graphe de Markov associé à la chaîne décrite dans l'exemple VIII.4.10.



Définition VIII.4.12. On dit qu'une chaîne de Markov est irréductible si elle n'admet qu'une classe. Dans ce cas, on dit aussi que la matrice de transition de la chaîne est irréductible.

Exemples VIII.4.13. (i) Dans l'exemple précédent, on vérifie que $\mathbb{E} \setminus \mathbb{E}' = \{2\}$ et que les classes d'équivalence de la relation $\leftrightarrow$ restreinte à $\{1, 3, 4, 5\}$ sont les ensembles $\{1, 3\}$ et $\{4, 5\}$.

(ii) Marches aléatoires sur $\mathbb{Z}/m\mathbb{Z}$. Nous utilisons les notations de l'exemple VIII.1.2.iii.

a) Considérons le cas où les variables $(Y_n)_{n\in\mathbb{N}}$ prennent leurs valeurs dans l'ensemble $\{-1, 1\}$. Si $P\{Y_0 = 1\}$ et $P\{Y_0 = -1\}$ sont non nuls, la chaîne est irréductible puisque chaque élément de $\mathbb{E}$ communique avec ses deux voisins.

b) Considérons le cas où les variables $(Y_n)_{n\in\mathbb{N}}$ prennent leurs valeurs dans l'ensemble $\{-2, 2\}$. Si $P\{Y_0 = 2\}$ et $P\{Y_0 = -2\}$ sont non nuls, la chaîne est irréductible si et seulement si m est impair ; elle admet deux classes si m est pair.

(iii) Les marches aléatoires sur $\mathbb{Z}$ de paramètres différents de 0 et 1 sont irréductibles puisque chaque élément de $\mathbb{E}$ communique avec ses deux voisins.

(iv) La marche aléatoire symétrique avec barrières absorbantes admet 3 classes : $\{N\}$, $\{-N\}$ et l'ensemble des entiers relatifs compris entre $-N+1$ et $N-1$.

(v) La chaîne d'Ehrenfest est irréductible puisque chaque élément de $\mathbb{E}$ communique avec ses deux voisins.

VIII.5. Récurrence et transience

***Notation* VIII.5.1.** *Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov à valeurs dans un espace dénombrable discret $\mathbb{E}$. Soit i un élément de $\mathbb{E}$. On note*

$$N_i = N_i(X) = \mathrm{card}\{\, n \geq 0 \,:\, X_n = i \,\}$$

le nombre de passages de la chaîne en i. On définit les instants successifs de passage en i,

$$\tau_i = \tau_i^1 = \tau_i^1(X) = \inf\{\, n > 0 \,:\, X_n = i \,\}$$

et, pour $n > 1$,

$$\tau_i^n = \tau_i^n(X) = \inf\{\, k > \tau_i^{n-1} \,:\, X_k = i \,\}\,.$$

De plus, on note P_i la loi de la chaîne conditionnée à débuter à l'état i. On note aussi E_i l'espérance sous P_i, c'est-à-dire l'espérance condionnelle à $X_0 = i$.

Les τ_i^n, $i \in \mathbb{E}$, $n \geq 1$, sont des temps d'arrêt relativement à toute filtration par rapport à laquelle la chaîne est adaptée. En effet,

$$\{\, \tau_i^n \leq m \,\} = \bigcup_{n\leq k\leq m} \; \bigcup_{1\leq j_1\leq\cdots\leq j_k\leq m} \Big(\bigcap_{1\leq l\leq k} \{\, X_{j_l} = i \,\} \bigcap_{j\notin\{j_1,\ldots,j_k\}} \{\, X_j \neq i \,\} \Big)$$

est $\sigma(X_1, \ldots, X_m)$-mesurable. Nous allons classifier les points de $\mathbb{E}$ suivant que ces temps sont finis ou non.

***Définition* VIII.5.2.** Un point i de $\mathbb{E}$ est dit récurrent pour la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ si $P_i\{\, \tau_i < \infty \,\} = 1$. Il est dit transient dans le cas contraire.

Autrement dit, le point i est récurrent si lorsque l'on en part, on est assuré d'y revenir en un temps fini. Pour une chaîne homogène, on est alors assuré d'y revenir infiniment souvent, comme le montre le lemme suivant.

Lemme VIII.5.3. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. Si i est un point récurrent, les $(\tau_i^n)_{n\geq 1}$ sont des temps d'arrêt P_i-p.s. finis.*

Démonstration. La démonstration se fait par récurrence en utilisant la propriété de Markov forte VIII.3.2. Par hypothèse, τ_i est presque sûrement fini. Supposons que τ_i^n soit presque sûrement fini. Alors,

$$P_i\{\,\tau_i^{n+1} = \infty\,\} = E_i\big(P_i\{\,\tau_i \circ \theta^{\tau_i^n} = \infty \mid \mathcal{F}_{\tau_i^n}\,\}\big) = E_i\big(P_i\{\,\tau_i = \infty\,\}\big) = 0\,. \qquad \square$$

Avant d'étudier la chaîne issue d'un de ses points récurrents, nous donnons quelques caractérisations de cette notion de récurrence fondées sur le nombre de visites de l'état i.

Théorème VIII.5.4. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{E}, \mathcal{P}(\mathbb{E}))$. Un point i de $\mathbb{E}$ est récurrent si et seulement si*

$$P_i\{\,N_i = \infty\,\} = 1\,.$$

Démonstration. C'est une conséquence du lemme suivant.

Lemme VIII.5.5. *Pour tout $i \in \mathbb{E}$ et tout entier $n \geq 1$,*

$$P_i\{\,N_i \geq n\,\} = \big(P_i\{\,\tau_i < \infty\,\}\big)^{n-1}\,.$$

Autrement dit, sous P_i, la variable aléatoire N_i suit une loi géométrique de paramètre $P_i\{\,\tau_i < \infty\,\}$.

Démonstration. Nous procédons par récurrence sur n. La propriété est de toute évidence vérifiée pour $n = 1$ puisque $P_i\{\,N_i \geq 1\,\} = 1$. Supposons la vérifiée au rang n. On peut écrire

$$\begin{aligned} P_i\{\,N_i \geq n+1\,\} &= P_i\big(\{\,N_i \geq n\,\} \cap \{\,\tau_i^n < \infty\,\}\big) \\ &= E_i\big(\mathbb{1}_{\{N_i\geq n\}} P\{\,\tau_i^n < \infty \mid \mathcal{F}_{\tau_i^{n-1}}\,\}\big)\,. \end{aligned}$$

Or $\tau_i^n(X) = \tau_i \circ \theta^{\tau_i^{n-1}}(X)$ sur $\{\,\tau_i^{n-1} < \infty\,\}$. Il résulte donc de la propriété de Markov forte (VIII.3.2) que

$$P\{\,\tau_i^n < \infty \mid \mathcal{F}_{\tau_i^{n-1}}\,\} = P_i\{\,\tau_i < \infty\,\}$$

puisque $X_{\tau_i^{n-1}} = i$ sur $\{\,\tau_i^{n-1} < \infty\,\}$. On en déduit que

$$P_i\{\,N_i \geq n+1\,\} = P_i\{\,N_i \geq n\,\} P_i\{\,\tau_i < \infty\,\} = \big(P_i\{\,\tau_i < \infty\,\}\big)^n\,,$$

d'après l'hypothèse de récurrence. $\square$

Corollaire VIII.5.6. *Soit i un point de $\mathbb{E}$. Alors*

$$P_i\{\, N_i = \infty \,\} = 1 \quad \Longleftrightarrow \quad P_i\{\, N_i = \infty \,\} > 0\,.$$

Corollaire VIII.5.7. *La variable aléatoire N_i est P_i-intégrable si et seulement si i est un point transient de $\mathbb{E}$.*

Corollaire VIII.5.8. *L'état i est récurrent si et seulement si la série $\sum_{n\geq 0} P_{ii}^n$ diverge.*

Démonstration. Il suffit de remarquer que $N_i = \sum_{n\geq 0} \mathbb{1}_{\{i\}}(X_n)$, de prendre l'espérance et d'utiliser le corollaire VIII.2.2.i. $\square$

***Exemple* VIII.5.9.** Marches aléatoires sur $\mathbb{Z}$. Étudions la récurrence du point 0 pour la marche aléatoire de paramètre p. Soit n un entier non nul ; la marche issue de 0 est de nouveau en 0 à l'instant n si elle a effectué autant de pas vers la gauche que vers la droite. Ainsi,

$$P_{00}^n = \begin{cases} \mathrm{C}_{2m}^m\, p^m(1-p)^m & \text{si } n = 2m, \\ 0 & \text{si } n \text{ est impair.} \end{cases}$$

Or $n! \sim (\frac{n}{e})^n\sqrt{2\pi n}$, d'après la formule de Stirling. On en déduit que

$$\begin{aligned} P_{00}^{2m} &\sim \Big(\frac{2m}{e}\Big)^{2m}\sqrt{4\pi m}\Big(\frac{m}{e}\Big)^{-2m}\frac{1}{2\pi m}p^m(1-p)^m \\ &= \frac{2^{2m}}{\sqrt{\pi m}}p^m(1-p)^m \\ &= \frac{\big(4p(1-p)\big)^m}{\sqrt{\pi m}}\,. \end{aligned}$$

Il s'ensuit que 0 est un état récurrent si $p = \frac{1}{2}$, transient sinon.

Nous allons en déduire qu'il en est de même de tous les états de la chaîne en montrant que la récurrence est une propriété de classe. Pour cela, nous utilisons la propriété dite de renouvellement d'une chaîne de Markov énoncée ci-dessous. Elle exprime que sur un point récurrent, une chaîne de Markov homogène se renouvelle, égale à elle-même en loi, indépendante de son passé.

Proposition VIII.5.10. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène et i un point récurrent de cette chaîne ; alors, pour tout entier n non nul, la loi de la chaîne $X\circ\theta^{\tau_i^n}$ est la même que la loi de la chaîne X sous P_i. De plus, la chaîne $X\circ\theta^{\tau_i^n}$ est indépendante de la tribu $\mathcal{F}_{\tau_i^n}$.*

Démonstration. C'est une conséquence directe de la propriété de Markov forte VIII.3.2 appliquée aux temps d'arrêt presque sûrement finis $(\tau_i^n)_{n\in\mathbb{N}}$. □

Théorème VIII.5.11. *La propriété de récurrence (ou de transience) d'un état est une propriété de classe.*

Démonstration. Soient i et j des états qui communiquent. Supposons que i est récurrent et démontrons que j l'est aussi. On a

$$\begin{aligned} P_j\{\, N_j = \infty \,\} &\geq E_j\big(P_j\{\, \mathbb{N}_j \circ \theta^{\tau_i} = \infty \mid \mathcal{F}_{\tau_i} \,\}\mathbb{1}_{\{\tau_i<\infty\}}\big) \\ &= E_j\big(P_i\{\, \mathbb{N}_j = \infty \,\}\mathbb{1}_{[0,\infty[}(\tau_i)\big) \\ &= P_i\{\, N_j = \infty \,\}P_j\{\, \tau_i < \infty \,\}\,. \end{aligned}$$

Nous sommes donc ramenés à montrer que $P_i\{\, N_j = \infty \,\} > 0$.

Lemme VIII.5.12. *Soit i un point récurrent et j un point de $\mathbb{E}$ distinct de i tel que i conduit à j. Alors,*

$$P_i\{\, N_j = \infty \,\} = P_i\{\, \tau_j < \infty \,\} = 1\,.$$

Démonstration. On définit le nombre de visites du point j avant la première visite en i, $N_j^i = \mathrm{card}\{\, n > 0 \,:\, X_n = j, n \leq \tau_i \,\}$. Sous P_i, $N_j = \sum_{n\geq 0} N_j^i \circ \theta^{\tau_n^i}$. Or, il résulte de la propriété de Markov forte que les variables aléatoires $N_j^i \circ \theta^{\tau_n^i}$ sont indépendantes et de même loi sous P_i. De plus,

$$E_i(N_j) \geq P_i\{\, N_j \geq 1 \,\} = P_i\{\, \tau_j < \infty \,\} > 0$$

puisque i conduit à j. On en déduit que $E(N_j^i) > 0$; il résulte alors de la loi des grands nombres V.5.2 que $N_j = \infty$ presque-sûrement. □

Définition VIII.5.13. Une chaîne de Markov est irréductible s'il n'existe qu'une seule classe de points récurrents.

Nous pouvons maintenant étudier les mesures invariantes pour la chaîne.

Théorème VIII.5.14. *Une mesure asymptotique ne charge pas les points transients, c'est-à-dire si μ est une mesure asymptotique et i un point transient, alors $\mu(\{\, i \,\}) = 0$.*

Démonstration. Nous allons montrer que, pour toute mesure initiale μ et tout point transient i de $\mathbb{E}$, $\lim_{n\to\infty} P_\mu\{\, X_n = i \,\} = 0$. Pour cela, il suffit de montrer que, pour tout point j de $\mathbb{E}$, $\lim_{n\to\infty} P_j\{\, X_n = i \,\} = 0$. Or

$$\begin{aligned} P_j\{\, X_n = i \,\} &= \sum_{1\le m\le n} P_j\{\, X_n = i, \tau_i = m \,\} \\ &= \sum_{1\le m\le n} P_j\{\, \tau_i = m \,\} P_i\{\, X_{n-m} = i \,\} . \end{aligned}$$

Il suffit donc de montrer que $\lim_{n\to\infty} P_i\{\, X_n = i \,\} = 0$. Soit

$$A_n = \{\, \forall p \ge n \,,\ X_p \ne i \,\} .$$

L'événement $\{\, X_n = i \}$ est inclus dans $\Omega \setminus A_n$. L'état i étant transient, il résulte du théorème VIII.5.4 que N_i est une variable aléatoire P_i-p.s. finie, d'où $P_i\big(\bigcup_{n\ge 1} A_n\big) = 1$. On en déduit que

$$\lim_{n\to\infty} P_i\{\, X_n = i \,\} \le \lim_{n\to\infty} P_i(\Omega \setminus A_n) = 0 . \qquad \square$$

Contrairement aux mesures asymptotiques, les mesures invariantes peuvent charger les classes transientes, comme nous l'avons vu pour les marches aléatoires sur $\mathbb{Z}$. Comme nous ne sommes intéressés que par les mesures asymptotiques, nous réduirons donc notre étude des mesures invariantes à celles qui ne chargent pas les classes transientes, c'est-à-dire à celles qui sont portées par les classes récurrentes. Rappelons que, lorsque la chaîne est irréductible, les mesures invariantes chargent tous les points de $\mathbb{E}$. La construction d'une mesure invariante est fondée sur l'idée suivante : un point de $\mathbb{E}$ est d'autant plus chargé par une mesure invariante qu'il est plus visité par la chaîne ; pour normaliser ce nombre de passages en un point, on se restreint à une excursion de la chaîne entre deux passages en un point i fixé de $\mathbb{E}$.

On travaille désormais sous l'hypothèse que la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ est irréductible et récurrente.

Soient i et j des éléments de $\mathbb{E}$. Rappelons que

$$N_j^i = \mathrm{card}\{\, n > 0 \,:\, X_n = j, n \le \tau_i \,\}$$

et définissons $\nu_j^i = E_i(N_j^i)$.

Nous allons montrer que le vecteur ν^i de composantes $(\nu_j^i)_{j\in\mathbb{E}}$ est un vecteur propre de la matrice ${}^{\mathrm{t}}\mathbb{P}$ associé à la valeur propre 1. Pour cela, étudions la loi de N_j^i sous P_i et P_j.

Lemme VIII.5.15. *Soient i et j deux points distincts de $\mathbb{E}$. Pour tout entier n,*

$$P_j\{ N_j^i = n \} = P_j\{ \tau_i < \tau_j \} P_j\{ \tau_j \leq \tau_i \}^n ,$$

et

$$P_i\{ N_j^i = n \} = \begin{cases} P_i\{ \tau_i < \tau_j \} & \text{si } n{=}0, \\ P_i\{ \tau_j \leq \tau_i \} P_j\{ \tau_i < \tau_j \} P_j\{ \tau_j \leq \tau_i \}^{n-1} & \text{si } n{>}0. \end{cases}$$

Démonstration. Remarquons tout d'abord que

$$\{ N_j^i = 0 \} = \{ \tau_i < \tau_j \} .$$

Soit à présent m un entier non nul. Observons l'appartenance de l'événement $\{ \tau_j \leq \tau_i \} = \Omega \setminus \{ \tau_i < \tau_j \}$ à $\mathcal{F}_{\tau_j}$. En conditionnant par la tribu $\mathcal{F}_{\tau_j}$ et en appliquant la propriété de Markov forte VIII.3.2, il vient

$$P_i\{ N_j^i = m \} = P_i\{ N_j^i \circ \theta_{\tau_j} = m-1, \tau_j \leq \tau_i \} = P_i\{ \tau_j \leq \tau_i \} P_j\{ N_j^i = m-1 \} .$$

Calculons $P_j\{ N_j^i = k \}$ par récurrence sur k en utilisant le même conditionnement. Pour tout $k \geq 1$

$$P_j\{ N_j^i = k \} = P_j\{ N_j^i = k, \tau_j \leq \tau_i \} = P_j\{ \tau_j \leq \tau_i \} P_j\{ N_j^i = k-1 \} .$$

On en déduit que pour tout entier k,

$$P_j\{ N_j^i = k \} = P_j\{ \tau_i < \tau_j \} P_j\{ \tau_j \leq \tau_i \}^k . \qquad \square$$

Montrons maintenant que si l'on part d'un état j, on ne peut pas être sûr d'atteindre l'état $i \neq j$ avant de revenir à l'état j.

Lemme VIII.5.16. *Soient i et j deux points de $\mathbb{E}$. Alors $P_j\{ \tau_j \leq \tau_i \} < 1$.*

Démonstration. Supposons que $P_j\{ \tau_j \leq \tau_i \} = 1$. Alors, pour tout entier $n \geq 2$, $P_j\{ \tau_j^n \leq \tau_i \} = 1$. En effet,

$$\begin{aligned} P_j\{ \tau_j^n \leq \tau_i \} &= P_j\{ \tau_j^{n-1} \leq \tau_i, \tau_j \circ \theta_{\tau_j^{n-1}} \leq \tau_i \circ \theta_{\tau_j^{n-1}} \} \\ &= P_j\{ \tau_j^{n-1} \leq \tau_i \} P_j\{ \tau_j \leq \tau_i \} \\ &= P_j\{ \tau_j^{n-1} \leq \tau_i \} . \end{aligned}$$

Les événements $\big(\{ \tau_j^n \leq \tau_i \}\big)_{n \geq 1}$ forment une suite décroissante d'événements de P_j-probabilité 1 ; leur intersection est donc de probabilité 1, c'est-à-dire P_j-presque sûrement, $\tau_i \geq \tau_j^n$ pour tout entier $n \geq 1$. Or, la suite des temps d'arrêt $(\tau_j^n)_{n \geq 1}$ est strictement croissante, donc elle tend vers ∞. On en déduit que $P_j\{ \tau_i = \infty \} = 1$, ce qui contredit le lemme VIII.5.12. $\square$

Corollaire VIII.5.17. *N_j^i est P_i-intégrable.*

Théorème VIII.5.18. *Pour tout élément i de $\mathbb{E}$, ${}^{\mathrm{t}}\mathbb{P}\nu^i = \nu^i$.*

Démonstration. Remarquons d'abord que pour tout $l \in \mathbb{E}$,

$$({}^{\mathrm{t}}\mathbb{P}\nu^i)_l = \sum_{j\in\mathbb{E}} P_{jl} \sum_{n\geq 1} E_i(N_j^i)\,.$$

De plus,

$$E_i(N_j^i) = \sum_{n\geq 1} P_i\{\,\tau_i \geq n, X_n = j\,\}$$

puisque

$$N_j^i = \sum_{1\leq n\leq \tau_i} \mathbb{1}_{\{\,j\,\}}(X_n) = \sum_{n\geq 1} \mathbb{1}_{[0,\tau_i]}(n)\mathbb{1}_j(X_n) = \sum_{n\geq 1} \mathbb{1}_{[n,\infty]\times\{\,j\,\}}(\tau_i, X_n)\,.$$

Donc

$$\begin{aligned} P_{jl}P_i\{\,\tau_i \geq n, X_n = j\,\} &= P\{\,X_{n+1} = l \;\big|\; X_n = j\,\}P_i\{\,\tau_i \geq n, X_n = j\,\} \\ &= P_i\{\,\tau_i \geq n, X_n = j, X_{n+1} = l\,\}\,. \end{aligned}$$

D'où

$$\begin{aligned} ({}^{\mathrm{t}}\mathbb{P}\nu^i)_l &= \sum_{n\geq 1} P_i\{\,\tau_i \geq n, X_{n+1} = l\,\} = E_i\biggl(\sum_{1\leq n\leq \tau_i} \mathbb{1}_{\{l\}}(X_{n+1})\biggr) \\ &= E_i\biggl(\sum_{0\leq n\leq \tau_i - 1} \mathbb{1}_{\{l\}}(X_{n+1})\biggr)\,. \end{aligned}$$

En effet $P_i\{\,X_{\tau_i+1} = l\,\} = P_i\{\,X_1 = l\,\}$. On en déduit, en posant $m = n+1$,

$$({}^{\mathrm{t}}\mathbb{P}\nu^i)_l = E_i\biggl(\sum_{1\leq m\leq \tau_i} \mathbb{1}_{\{l\}}(X_m)\biggr) = \nu_l^i\,.$$

C'est le résultat. □

On déduit du théorème précédent que si $\sum_{l\in\mathbb{E}} \nu_l^i$ est convergente, la chaîne admet une probabilité invariante. Nous allons montrer que cette condition est nécessaire et suffisante à l'existence et l'unicité d'une probabilité invariante. Remarquons auparavant que $\sum_{l\in\mathbb{E}} \nu_l^i = E_i(\sum_{l\in\mathbb{E}} N_l^i) = E_i(\tau_i)$ pour tout i.

Théorème VIII.5.19. *Soit μ une probabilité invariante de la chaîne. Alors, pour tout élément i et j de $\mathbb{E}$, $\mu_j = \nu_j^i \mu_i$.*

Démonstration. Nous allons utiliser, pour démontrer ce théorème, une technique de retournement du temps. Remarquons d'abord que pour tout n,

$$\begin{aligned} P_i\{\,\tau_i \geq n, X_n = j\,\} &= \frac{P_\mu\{\,\tau_i \geq n, X_n = j, X_0 = i\,\}}{P_\mu\{\,X_0 = i\,\}} \\ &= \frac{P_\mu\{\,\tau_i \geq n, X_n = j, X_0 = i\,\}}{\mu_i}\,. \end{aligned}$$

Pour tout entier $k \leq n$, introduisons $X_k^n = X_{n-k}$. Alors,

$$P_i\{\,\tau_i \geq n, X_n = j\,\} = \frac{P_\mu\{\,\tau_i \geq n, X_0^n = j, X_n^n = i\,\}}{\mu_i}\,.$$

Nous allons étudier la loi jointe des variables aléatoires $(X_k^n)_{0\leq k\leq n}$. Plus précisément, nous allons montrer qu'elles possèdent la propriété de Markov relativement à la famille de tribus $(\mathcal{G}_k^n)_{0\leq k\leq n}$, où $\mathcal{G}_k^n$ est la tribu engendrée par les variables aléatoires $X_n, X_{n-1}, \ldots, X_{n-k}$.

Lemme VIII.5.20. *Soient k et n des entiers tels que $k \leq n$, et soient i et j des éléments de $\mathbb{E}$. Alors,*

$$P_\mu\{\,X_{k+1}^n = j \mid \mathcal{G}_k^n\,\} = P_\mu\{\,X_{k+1}^n = j \mid X_k^n\,\} = P_{ji}\frac{\mu_j}{\mu_i}$$

sur $\{\,X_k^n = i\,\}$.

Démonstration. En utilisant la formule de Bayes (VI.1.4), il vient

$$\begin{aligned} P_\mu\{\,X_{k+1}^n = j \mid X_k^n = i\,\} &= P_\mu\{\,X_k^n = i \mid X_{k+1}^n = j\,\}\frac{P_\mu\{\,X_{k+1}^n = j\,\}}{P_\mu\{\,X_k^n = i\,\}} \\ &= P_\mu\{\,X_{n-k} = i \mid X_{n-k-1} = j\,\}\frac{P_\mu\{\,X_{n-k-1} = j\,\}}{P_\mu\{\,X_{n-k} = i\,\}} \\ &= P_{ji}\frac{\mu_j}{\mu_i}\,. \end{aligned}$$

Il suffit à présent de montrer que si $i_0, \ldots, i_k$ sont des éléments de $\mathbb{E}$,

$$\begin{aligned} &P_\mu\{\,X_{n-k-1} = j, X_n = i_0, \ldots, X_{n-k} = i_k\,\} \\ &\qquad\qquad = P_{ji_k}\frac{\mu_j}{\mu_{i_k}}P_\mu\{\,X_n = i_0, \ldots, X_{n-k} = i_k\,\}\,. \end{aligned}$$

Or, il résulte de la proposition VIII.2.1 et de l'invariance de μ que

$$\begin{aligned}
P_\mu\{\, X_{n-k-1} = j, X_n = i_0, \ldots, X_{n-k} = i_k \,\}& \\
&= P_{ji_k} P_{i_k i_{k-1}} \cdots P_{i_1 i_0} \mu_j \\
&= P_{ji_k} \frac{\mu_j}{\mu_{i_k}} P_{i_k i_{k-1}} \cdots P_{i_1 i_0} \mu_{i_k} \\
&= P_{ji_k} \frac{\mu_j}{\mu_{i_k}} P_\mu\{\, X_n = i_0, \ldots, X_{n-k} = i_k \,\}\,,
\end{aligned}$$

ce qui démontre le lemme. □

Proposition VIII.5.21. *Soit $\mathbb{P}$ une matrice stochastique indexée par un ensemble $\mathbb{E}$, irréductible, récurrente et admettant une probabilité invariante μ. On définit une matrice $\mathbb{Q}$ indexée par $\mathbb{E}$ en posant*

$$Q_{ij} = P_{ji} \frac{\mu_j}{\mu_i}\,, \qquad i, j \in \mathbb{E}\,.$$

La matrice $\mathbb{Q}$ est stochastique, irréductible, récurrente et admet μ comme probabilité invariante.

De plus, si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène, de matrice de transition $\mathbb{P}$ sur $\mathbb{E}^\mathbb{N}$ et n un entier non nul, la loi du vecteur aléatoire $(X_{n-k})_{0\le k\le n}$ sous $P_{\mu,\mathbb{P}}$ est identique à celle du vecteur aléatoire $(X_k)_{0\le k\le n}$ sous $P_{\mu,\mathbb{Q}}$.

Démonstration. On vérifie aisément que $\mathbb{Q}$ est une matrice stochastique qui admet μ comme mesure invariante. Pour montrer que $\mathbb{Q}$ est irréductible et récurrente, calculons les puissances successives de $\mathbb{Q}$; plus précisément, démontrons par récurrence sur n que

$$Q^n_{ij} = P^n_{ji} \frac{\mu_j}{\mu_i}\,.$$

C'est vrai pour $n = 1$. Supposons le vrai pour n. Pour $i, j \in \mathbb{E}$,

$$Q^{n+1}_{ij} = \sum_{l\in\mathbb{E}} Q^n_{il} Q_{lj} = \sum_{l\in\mathbb{E}} P^n_{li} \frac{\mu_l}{\mu_i} P_{jl} \frac{\mu_j}{\mu_l} = \frac{\mu_j}{\mu_i} \sum_{l\in\mathbb{E}} P^n_{li} P_{jl} = \frac{\mu_j}{\mu_i} P^{n+1}_{ji}\,.$$

Soient i et j deux éléments de $\mathbb{E}$; puisque $\mathbb{P}$ est irréductible il existe un entier N tel que $\mathbb{P}^N_{ij} > 0$; on en déduit que $Q^N_{ij} > 0$ et, par suite que $\mathbb{Q}$ est irréductible. D'autre part, si i est un élément de $\mathbb{E}$, la série $\sum_{n\in\mathbb{N}} Q^n_{ii}$ est de même nature que la série $\sum_{n\in\mathbb{N}} P^n_{ii}$, c'est-à-dire divergente. On en déduit que i est récurrent pour toute chaîne de matrice de transition $\mathbb{Q}$.

La fin de la proposition résulte du lemme VIII.5.20. □

Nous pouvons conclure à présent la démonstration du théorème VIII.5.19. Il résulte en effet de ce qui précède que si $i, j \in \mathbb{E}$, pour tout n,

$$\begin{aligned} P_i\{\,\tau_i \geq n, X_n = j\,\} &= \frac{P_\mu\{\,\tau_i \geq n, X_n = j, X_0 = i\,\}}{\mu_i} \\ &= \frac{P_{\mu,\mathbb{Q}}\{\,\tau_i \geq n, X_0 = j, X_n = i\,\}}{\mu_i} \\ &= P_{j,\mathbb{Q}}\{\,\tau_i \geq n, X_n = i\,\}\frac{\mu_j}{\mu_i}\,. \end{aligned}$$

En sommant par rapport à n, on obtient

$$\begin{aligned} E_i(N_j^i) &= \frac{\mu_j}{\mu_i}\sum_{n\geq 1} P_{j,\mathbb{Q}}\{\,\tau_i \geq n, X_n = i\,\} \\ &= \frac{\mu_j}{\mu_i}\sum_{n\geq 1} P_{j,\mathbb{Q}}(\tau_i = n) \\ &= \frac{\mu_j}{\mu_i} P_{j,\mathbb{Q}}\{\,\tau_i < \infty\,\}\,. \end{aligned}$$

L'état i est récurrent pour la chaîne de matrice de transition $\mathbb{Q}$. Le lemme VIII.5.12 montre que $P_{j,\mathbb{Q}}\{\,\tau_i < \infty\,\} = 1$, et ceci conclut la démonstration du théorème VIII.5.19. $\square$

Corollaire VIII.5.22. *Les assertions suivantes sont équivalentes :*

(i) la chaîne $(X_n)_{n\in\mathbb{N}}$ admet une probabilité invariante ;

(ii) il existe un élément i de $\mathbb{E}$ tel que $E_i(\tau_i) < \infty$;

(iii) $E_i(\tau_i) < \infty$ pour tout élément i de $\mathbb{E}$.

Lorsqu'elles sont vérifiées, la chaîne admet une unique probabilité invariante μ donnée pour tout i élément de $\mathbb{E}$ par :

$$\mu_j = \frac{E_i(N_j^i)}{E_i(\tau_i)} = \frac{P_i(\tau_j \leq \tau_i)}{E_i(\tau_i)} = \frac{1}{E_j(\tau_j)} \qquad \textit{pour tout } j \in \mathbb{E}\,.$$

Terminons ce paragraphe par une définition immédiatement issue de ce qui précède.

Définition VIII.5.23. Un point i de $\mathbb{E}$ est dit récurrent positif pour la chaîne $(X_n)_{n\in\mathbb{N}}$ si et seulement si $E_i(\tau_i) < \infty$. Le fait d'être récurrent positif est une propriété de classe. La classe d'un élément récurrent positif est dite récurrente positive.

Un point récurrent de $\mathbb{E}$ qui n'est pas récurrent positif est dit récurrent nul. La classe d'un élément récurrent nul est dite récurrente nulle.

***Exemple* VIII.5.24.** Puisque la marche aléatoire symétrique sur $\mathbb{Z}$ n'admet pas de probabilité invariante, les éléments de $\mathbb{Z}$ sont récurrents nuls pour cette chaîne en vertu du corollaire VIII.5.22.

VIII.6. Comportement asymptotique d'une chaîne de Markov

Pour les raisons exposées dans les paragraphes précédents, nous nous intéresserons, dans ce paragraphe, au comportement asymptotique d'une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ admettant une et une seule classe récurrente positive. Le résultat attendu est le suivant : pour toute loi initiale, la chaîne de Markov converge vers la mesure invariante. Nous allons voir sur un cas très simple que ceci n'est pas toujours vérifié.

***Exemple* VIII.6.1.** Reprenons l'exemple de la marche aléatoire sur $\mathbb{Z}/m\mathbb{Z}$ avec $m = 4$ et les $(Y_n)_{n\geq 1}$ suivant la loi uniforme sur $\{-1, 1\}$. Prenons $Y_0 = 0$. Alors la loi de X_n est portée par $\{0, 2\}$ pour n pair et par $\{1, 3\}$ pour n impair. Il ne peut donc y avoir convergence en loi de la chaîne. Cet exemple nous conduit à la définition de la période d'un point.

***Définition* VIII.6.2.** On dit qu'un point $i \in \mathbb{E}$ est de période d pour la chaîne $(X_n)_{n\in\mathbb{N}}$ si $d = \mathrm{pgcd}\{\, n \geq 1, P_{ii}^n > 0 \,\}$.

Dans l'exemple VIII.6.1, la période de chaque point est 2.

***Proposition* VIII.6.3.** *Le fait d'être de période d est une propriété de classe. Une classe de période 1 est dite apériodique.*

Démonstration. Soit $i \in \mathbb{E}$ un point de période d et $j \in \mathbb{E}$ tel que $i \leftrightarrow j$. Montrons que la période d' de j est égale à d. On a l'équivalence

$$i \leftrightarrow j \iff \exists\, k, l \geq 1\,,\ P_{ij}^k > 0 \text{ et } P_{ji}^l > 0\,.$$

On en déduit que $P_{ii}^{k+l} > 0$ et, par suite, que d divise $k + l$. Soit $n \geq 1$ tel que $P_{jj}^n > 0$. Alors $P_{ii}^{n+k+l} \geq P_{ij}^k P_{jj}^n P_{ji}^l > 0$. D'où d divise $n + k + l$ et par suite, d divise n. On en déduit que d divise d' et par symétrie $d = d'$. ☐

Comme il apparaît dans l'exemple ci-dessus, il est naturel, lorsqu'une chaîne est de période $d \neq 1$, d'étudier la chaîne de matrice de transition $\mathbb{P}^d$. Avant de décrire les propriétés de cette chaîne auxiliaire, nous énonçons un résultat technique très utile dans la suite.

Lemme VIII.6.4. *Pour $i \in \mathbb{E}$, notons $D_i = \{\, n \geq 1, P_{ii}^n > 0 \,\}$ et d_i la période de i. Alors, il existe un entier n_i non nul tel que D_i contienne tous les entiers multiples de d_i supérieurs à n_i.*

Démonstration. Remarquons tout d'abord que si $m, n \in D_i$, alors $m + n \in D_i$, c'est-à-dire que D_i est un semi-groupe pour l'addition dans $\mathbb{N}$. En effet $P_{ii}^{m+n} \geq P_{ii}^m P_{ii}^n > 0$. Par hypothèse, le pgcd des éléments de D_i est d_i. Notons n_k les éléments de D_i rangés par ordre croissant. Si D_i est infini, la suite $\mathrm{pgcd}(n_1, \ldots, n_k)$, $k \geq 1$, est une suite décroissante d'entiers. Son plus petit élément est donc atteint à partir d'un certain rang. Par hypothèse, ce plus petit élément est d_i. Il existe donc une famille finie d'éléments de D_i dont le pgcd est d_i ; notons les $n_1, \ldots, n_p$. Il existe alors des entiers relatifs $\alpha_1, \ldots, \alpha_p$ tels que $\sum_{i=1}^p \alpha_i n_i = d_i$. Notons $q = \sum_{i|\alpha_i > 0} \alpha_i n_i$ et $q' = \sum_{i|\alpha_i<0} \alpha_i n_i$. Alors q et q' sont éléments de D_i *et* $q - q' = d_i$. Soit n un entier non nul multiple de d_i : alors $n = aq' + rd_i$ avec $rd_i < q'$. D'où $n = (a-r)q' + rq \in D_i$ si $a \geq q'$. On pose $n_i = q'(q' + d_i)$; alors D_i contient tous les entiers multiples de d_i et supérieurs à n_i. □

Proposition VIII.6.5. *Supposons la matrice $\mathbb{P}$ irréductible et de période $d \neq 1$. Alors, la matrice $\mathbb{P}^d$ n'est pas irréductible ; elle possède exactement d classes que l'on peut désigner par $C_0, \ldots, C_{d-1}$ de telle façon que, si la loi de X_0 est portée par C_i, alors celle de X_1 est portée par C_{i+1} (avec $C_d = C_0$).*

Démonstration. Soit i un élément fixé de $\mathbb{E}$. Si j est un autre élément de $\mathbb{E}$, on note $D_{ij} = \{\, n \geq 1\,,\, P_{ij}^n > 0 \,\}$. Soit n_{ij} le plus petit élément de D_{ij} *et* r_{ij} le reste de sa division euclidienne par d. Démontrons que pour tout $n \in D_{ij}$, $n = r_{ij} \bmod d$.

La chaîne étant irréductible, il existe un entier $m > 0$ tel que $P_{ji}^m > 0$. Donc, si $P_{ii}^{m+n_{ij}} > 0$ et $P_{ii}^{m+n} > 0$, on a

$$m + n_{ij} \equiv 0 \mod d \qquad \text{et} \qquad m + n \equiv 0 \mod d,$$

d'où $n \equiv n_{ij} \mod d$. Pour $0 \leq r < d$, notons $C_r = \{\, j \in E, r_{ij} = r \,\}$. Démontrons que $C_0, \ldots, C_{d-1}$ sont les classes de la chaîne de matrice de transition $\mathbb{P}^d$. Pour cela, fixons r et démontrons que deux éléments j et k de C_r communiquent pour cette chaîne. Or $j \in C_r$ si et seulement si il existe $n \geq 1$, $n = r \bmod d$ tel que

$P^n_{ij} > 0$ et de la même façon $k \in C_r$ si et seulement si il existe $m \geq 1$, $m = r \mod d$ tel que $P^m_{ik} > 0$. D'autre part, k conduit à i ; donc il existe $l > 0$ tel que $P^l_{ki} > 0$. On en déduit que $m + l \in D_i$ et donc que d divise $m + l$. On a, de plus, $P^{l+n}_{kj} \geq P^l_{ki} P^n_{ij} > 0$ avec $l + n = l + m - m + n \equiv 0 \mod d$, d'où k conduit à j pour la chaîne de matrice de transition P^d.

Soient r *et* r' deux éléments distincts de $\{0, \ldots, d-1\}, j \in C_r$ et $k \in C_{r'}$. Supposons que j conduit à k pour la chaîne de matrice de transition $\mathbb{P}^d$. Il existe $n > 0$ tel que $P^{dn}_{jk} > 0$. Soit $m \in D_{ij}$;

$$
\begin{aligned}
P^{m+dn}_{ik} \geq P^m_{ij} P^{dn}_{jk} > 0 &\Rightarrow m + dn \in D_{ik} \\
&\Rightarrow m + dn \equiv r' \mod d \\
&\Rightarrow m = r' \mod d \,,
\end{aligned}
$$

ce qui est absurde puisque $j \in C_r$.

Enfin, soient $j \in C_r$ et $k \in E$ tels que $P_{jk} > 0$. Soit $n \in D_{ij}$; alors, $P^{n+1}_{ik} \geq P^n_{ij} P_{jk} > 0$. D'où $n + 1 \in D_{ik}$ et $n + 1 = r + 1 \mod d$, d'où l'on déduit que k est un élément de C_{r+1}. □

Ces deux lemmes vont nous permettre de donner une caractérisation algébrique des chaînes de Markov apériodiques à valeurs dans un espace fini.

Théorème VIII.6.6. *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur $(\Omega, \mathcal{A}, P)$, à valeurs dans un ensemble fini $\mathbb{E}$ et irréductible. Les propriétés suivantes sont équivalentes :*

(i) la chaîne est apériodique ;

(ii) pour tout n assez grand, pour tous $i, j \in \mathbb{E}$, $P^n_{ij} > 0$;

(iii) 1 est la seule valeur propre de module 1 de la matrice ${}^{\mathrm{t}}\mathbb{P}$.

Démonstration. Montrons tout d'abord que (i) implique (ii). Soit, pour tout élément i de $\mathbb{E}$, n_i l'entier construit dans le lemme VIII.6.4 et $N = \max_{i \in E} n_i$. Remarquons d'autre part que, pour tous $i, j \in \mathbb{E}$, il existe $N_{ij} > 0$ tel que $P^{N_{ij}}_{ij} > 0$. Soit $n' = \max_{i,j \in \mathbb{E}}(N_{ij})$ et $n = N + N'$. Alors, si $i, j \in \mathbb{E}$, $P^n_{ij} \geq P^{N_{ij}}_{ij} P^{n - N_{ij}}_{j,j} > 0$ puisque $n - N_{ij} \geq n - N' = N$.

Supposons à présent (ii) vérifié et démontrons (iii). Soit $\theta \in \mathbb{R}$ et $v \in \mathbb{R}^{\mathrm{card}\mathbb{E}}$ tels que ${}^{\mathrm{t}}\mathbb{P} v = e^{i\theta} v$. Alors ${}^{\mathrm{t}}\mathbb{P}^n v = e^{in\theta} v$ pour tout $n \in \mathbb{N}$. Il résulte alors du lemme de Perron-Froebenius (VIII.4.8) que ${}^{\mathrm{t}}\mathbb{P}^n |v| = |v|$. On en déduit que

$$
\Bigl| \sum_{j \in E} P^n_{ji} v_j \Bigr| = \sum_{j \in E} P^n_{ji} |v_j|
$$

pour tout $i \in \mathbb{E}$. Si $n \geq N$, les $(P_{ij}^n)_{j\in\mathbb{E}}$ sont non nuls ; il existe donc $\alpha \in \mathbb{R}$ tel que, pour tout $j \in \mathbb{E}$, $v_j = e^{i\alpha}|v_j|$. On a donc, d'une part ${}^{\mathrm{t}}\mathbb{P}^n v = e^{in\theta}v$, d'autre part

$$e^{in\theta}v = e^{in(\theta+\alpha)}|v| = e^{in(\theta+\alpha)}({}^{\mathrm{t}}\mathbb{P}^n)|v| = e^{in\theta}({}^{\mathrm{t}}\mathbb{P}^n)v\,.$$

On en déduit que $e^{in\theta} = 1$. Ceci étant vrai pour une infinité d'entiers n, il s'ensuit que $\theta = 0$.

Il reste à prouver que (iii) implique (i). Nous allons le faire par l'absurde. Supposons que la période d est strictement supérieure à 1. Nous allons utiliser la proposition VIII.6.5 pour construire un vecteur propre de la matrice ${}^{\mathrm{t}}\mathbb{P}$ associé à la valeur propre $e^{2i\pi/d}$. En effet, soit μ l'unique mesure invariante de la chaîne de matrice de transition $\mathbb{P}^d$ restreinte à C_0 (on identifie μ à une probabilité sur $\mathbb{E}$). Alors pour tout $0 \leq r \leq d-1$, ${}^{\mathrm{t}}\mathbb{P}^r\mu$ est portée par C_r. Donc

$$\nu = \sum_{0\leq r\leq d-1} e^{2i\pi\frac{r}{d}}({}^{\mathrm{t}}\mathbb{P}^r)\mu$$

est non nul et vérifie de plus ${}^{\mathrm{t}}\mathbb{P}\nu = e^{2i\pi/d}\nu$. □

Nous concluons ce chapitre par le théorème ergodique, lequel décrit la convergence des chaînes de Markov vers une mesure limite. Il énonce que les moyennes en temps (*i.e.* sur l'indice n) convergent vers les moyennes en espace (*i.e.* par rapport à la mesure invariante μ).

Définition VIII.6.7. On dit qu'une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ est ergodique s'il existe une probabilité μ telle que, pour toute condition initiale X_0, la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers μ.

Théorème VIII.6.8. *Une chaîne de Markov est ergodique si elle est irréductible, récurrente positive et apériodique.*

Démonstration. On pourra se reporter à Revuz (1975). □

Nous étudions à présent le comportement asymptotique de moyennes temporelles, c'est-à-dire du type

$$\frac{1}{n}\big(f(X_1) + \cdots + f(X_n)\big)\,.$$

Pour cela, nous allons nous ramener à la loi des grands nombres classique en utilisant les excursions de la chaîne entre deux passages en un même point.

Théorème VIII.6.9 (loi forte des grands nombres). *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov irréductible et récurrente positive. Soit μ l'unique mesure invariante de la chaîne. Alors, pour toute fonction $f \in \mathrm{L}^1(\mu)$ et toute loi initiale μ_0,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n} f(X_k) = \int_E f \,\mathrm{d}\mu \quad p.s.$$

Démonstration. On peut supposer la fonction f positive. On fixe $i \in \mathbb{E}$ et, pour $n > 0$, on définit $N_i(n) = \sum_{1\leq k\leq n} \mathbb{1}_{\{i\}}(X_k)$, le nombre de fois où la chaîne est passée en i avant l'instant n. On a alors $\tau_i^{N_i(n)} \leq n < \tau_i^{N_i(n)+1}$. D'où

$$\sum_{0\leq k\leq \tau_i^{N_i(n)}} f(X_k) \leq \sum_{0\leq k\leq n} f(X_k) \leq \sum_{0\leq k\leq \tau_i^{N_1(n)+1}} f(X_k).$$

Introduisons les variables aléatoires Z_n, $n\in\mathbb{N}$, définies par $Z_0 = \sum_{1\leq k\leq \tau_i} f(X_k)$ et $Z_n = \sum_{\tau_i^n+1\leq k\leq \tau_i^{n+1}} f(X_k)$ pour tout entier $n\geq 1$. Alors

$$\sum_{0\leq k\leq N_i(n)-1} Z_k \leq \sum_{0\leq k\leq n} f(X_k) \leq \sum_{0\leq k\leq N_i(n)} Z_k\,.$$

Les temps d'arrêt $(\tau_i^n)_{n\geq 1}$ étant p.s. finis, on a

$$\lim_{n\to\infty} N_i(n) = \infty \text{ p.s.}$$

Nous utiliserons pour finir le lemme suivant.

Lemme VIII.6.10. *Les variables aléatoires Z_n, $n \geq 1$, sont indépendantes, de même loi et de moyenne $\int_E f\,\mathrm{d}\mu / E_i(\tau_i)$.*

Démonstration. Remarquons tout d'abord que pour tout $n\geq 1$, $Z_n = Z_0\circ\theta^{\tau_i^n}$. Ainsi, pour toute fonction $\phi : \mathbb{E}\to\mathbb{R}$ bornée,

$$E\big(\phi(Z_n)\big) = E\big(E(\phi(Z_n) \mid \mathcal{F}_{\tau_i^n})\big) = E_i\big(\phi(Z_0)\big)\,.$$

Les Z_n, $n\geq 1$, sont donc de même loi. Montrons à présent leur indépendance. Remarquons que Z_n est $\mathcal{G}_n$-mesurable, où $\mathcal{G}_n = \mathcal{F}_{\tau_i^{n+1}}$; l'indépendance est donc une conséquence de la relation

$$E\big(\phi(Z_n) \bigm| \mathcal{G}_{n-1}\big) = E\big(\phi(Z_n)\big)\,.$$

Enfin, on a

$$E_i(Z_0) = \sum_{j\in\mathbb{E}} f(j) E_i\bigg(\sum_{0\le k\le\tau_i} \mathbb{1}_{\{j\}}(X_k)\bigg) = \sum_{j\in\mathbb{E}} f(j) E_i(N_j^i) = \sum_{j\in\mathbb{E}} f(j)\frac{\mu_j}{E_i(\tau_i)} . \qquad \square$$

Il résulte de ce lemme et de la loi forte des grands nombres (V.5.2) que

$$\lim_{n\to\infty} \frac{1}{N_i(n)} \sum_{0\le k\le N_i(n)-1} Z_k = \lim_{n\to\infty} \frac{1}{N_i(n)} \sum_{0\le k\le N_i(n)} Z_k = E_i(Z_0) = \frac{\int_E f\,\mathrm{d}\mu}{E_i(\tau_i)}$$

presque sûrement. D'où

$$\lim_{n\to\infty} \frac{1}{N_i(n)} \sum_{0\le k\le n} f(X_k) = \Big(\int_E f\,\mathrm{d}\mu\Big)\big(E_i(\tau_i)\big)^{-1} \text{ p.s.}$$

Prenant $f \equiv 1$, on obtient

$$\lim_{n\to\infty} \frac{N_i(n)}{n} = E_i(\tau_i)$$

ce qui conduit au résultat annoncé. $\square$

Exercices

Exercice VIII.1. À quelles conditions deux matrices

$$\mathbb{P} = (P_{ij})_{1\le i\le n, 1\le j\le m} \qquad \text{et} \qquad \mathbb{Q} = (Q_{ij})_{1\le i\le m, 1\le j\le n}$$

sont-elles les lois conditionnelles $\mathcal{L}(X \mid Y)$ et $\mathcal{L}(Y \mid X)$ de deux variables aléatoires X et Y prenant respectivement n et m valeurs ? Montrer que si l'on connaît $\mathcal{L}(X \mid Y) = \mathbb{P}$ et $\mathcal{L}(Y \mid X) = \mathbb{Q}$, alors on connaît la loi du couple (X, Y).

Exercice VIII.2. Montrer que $(X_0, \ldots, X_n)$ est une chaîne de Markov à valeurs dans un ensemble fini $\mathbb{E}$ si et seulement si il existe des fonctions $g_i : \mathbb{E}\times\mathbb{E} \to [\,0,\infty\,[$, $0 \le i \le n-1$, telles que, pour tous $x_0, \ldots, x_n \in \mathbb{E}$,

$$P\{\, X_0 = x_0, \ldots, X_n = x_n \,\} = g_0(x_0, x_1) g_1(x_1, x_2) \cdots g_{n-1}(x_{n-1}, x_n) \,.$$

Exercice VIII.3. Sur l'ensemble fini $\mathbb{E} = \mathbf{Z}/m\mathbf{Z}$, on considère la chaîne $(X_n)_{n\ge 0}$ de générateurs $P_{i,i+k} = P_{i,i-k} = 1/2$, $P_{i,j} = 0$ sinon, où $1 \le k < m$. Pour quelles valeurs de m et k la chaîne est-elle récurrente irréductible ? Donner, dans tous les cas, ses classes de récurrence et la mesure invariante de ses classes. Lorsque la chaîne est récurrente irréductible, déterminer quand elle est apériodique.

Montrer que l'on peut réaliser la chaîne $(X_n)_{n\ge 0}$ sous la forme $X_{n+1} = f(X_n, \varepsilon_n)$ avec une fonction f et une suite $(\varepsilon_n)_{n\ge 0}$ de variables aléatoires dans $\{\, -1, +1 \,\}$ que l'on déterminera.

Exercice VIII.4. Soit $(X_n)_{n\geq 0}$ une chaine de Markov de matrice de transition P_{ij} avec $P_{ij} > 0$ pour tout couple (i, j). On suppose que $X_0 = i$ p.s. et l'on choisit $j \neq i$. Soit

$$T = \inf\{\, n \geq 1 \,:\, X_n = j \,\} \,.$$

Démontrer qu'il existe $\rho \in]\,0, 1[$ tel que $P\{\, T > n \,\} \leq \rho^n$ pour tout $n \geq 1$.

Exercice VIII.5. Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires réelles et de même loi de fonction de répartition continue F. Considérons les temps de record T_n, $n \geq 1$, et les records X_{T_n}, définis par $T_0 = 0$ et

$$T_{n+1} = \min\{\, i > T_n \,:\, X_i \geq X_{T_n} \,\} \,.$$

Démontrer que $(T_n)_{n\in\mathbb{N}}$ et $(X_{T_n})_{n\in\mathbb{N}}$ sont deux chaînes de Markov non homogènes.

Exercice VIII.6. Soit $(V, \mathcal{E})$ un graphe connexe non orienté d'ensemble de sommets fini V et d'ensemble d'arêtes $\mathcal{E} \in V \times V$. On associe à chaque arête (i, j) un poids $w_{i,j} = w_{j,i} > 0$ et l'on pose $w_i = \sum_j w_{i,j}$. Déterminer la mesure invariante de la chaîne de Markov sur V de matrice de transition $P_{i,j} = w_{i,j}/w_i$.

BIBLIOGRAPHIE

Dacunha-Castelle, D., Duflo, M. (1982). *Probabilités et statistiques*, Masson.

Dudley, R.M. (1989). *Real Analysis and Probability*, Wadsworth.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Wiley.

Fine, T.L. (1973). *Theories of Probabilities : An Examination of Foundations*, Academic press.

Foata, D., Fuchs, A. (1998). *Cours de probabilités pour la licence* (2e édition), Dunod.

Grimmett, G. (1993). *Probability Theory*, Cambridge University Press.

Malliavin, P. (1982). *Intégration et probabilités, analyse de Fourier et analyse spectrale*, Masson.

Neveu, J. (1964). *Bases mathématiques du calcul des probabilités*, Masson.

Norris, J.R. (1997). *Markov Chains*, Cambridge University Press.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer.

Revuz, D. (1975). *Markov Chains*, North Holland.

Revuz, D. (1997). *Probabilités*, Hermann.

Rudin, W. (1975). *Functional Analysis*, McGraw-Hill.

Williams, D. (1991). *Probability with Martingales*, Cambridge University Press.

APPENDICE

LOIS DE PROBABILITÉS USUELLES

1. Loi de Bernoulli

Définition. Une variable aléatoire X, à valeurs dans $\{0,1\}$, suit une loi de Bernoulli de paramètre $p \in [0,1]$, notée $\mathcal{B}(1,p)$, si

$$P\{X=1\} = p = 1 - P\{X=0\}.$$

Espérance : p
Variance : $p(1-p)$
Fonction caractéristique : $1 - p + pe^{it}$

2. Loi binomiale

Soit C_n^k le coefficient binomial $\dfrac{n!}{k!(n-k)!}$.

Définition. Une variable aléatoire X, à valeurs entières, suit une loi binomiale de taille $n \geq 1$ et de paramètre $p \in [0,1]$, notée $\mathcal{B}(n,p)$, si

$$P\{X=k\} = \mathrm{C}_n^k\, p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Espérance : np
Variance : $np(1-p)$
Fonction caractéristique : $(1 - p + pe^{it})^n$

Stabilité par convolution : $\mathcal{B}(n,p)*\mathcal{B}(m,p) = \mathcal{B}(n+m,p)$, ou de façon équivalente, si X, Y sont indépendantes et de lois respectives $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$, alors $X+Y$ est de loi $\mathcal{B}(n+m,p)$. En particulier, si $(X_i)_{1\leq i\leq n}$ est une suite de variables de Bernoulli indépendantes, alors $X_1+\cdots+X_n$ est de loi $\mathcal{B}(n,p)$. Il en résulte que si une urne contient n boules, une proportion p d'entre elles étant noires, $1-p$ étant blanches, et si l'on tire au hasard sans remise n boules, le nombre de boules noires tirées suit une loi $\mathcal{B}(n,p)$.

Proposition. *(i) Si X_n suit une loi $\mathcal{B}(n,p_n)$ avec $\lim_{n\to\infty} np_n = \lambda$, $\lambda > 0$, alors X_n converge en loi quand $n\to\infty$ vers une variable de loi de Poisson de paramètre λ.*
(ii) Si X_n suit une loi $\mathcal{B}(n,p)$, alors $(X_n - np)/\sqrt{np(1-p)}$ converge en loi quand $n\to\infty$ vers une variable de loi normale centrée réduite $\mathcal{N}(0,1)$.

3. Loi de Poisson

Définition. Une variable aléatoire X, à valeurs entières, suit une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$, si

$$P\{\, X = k \,\} = e^{-\lambda}\frac{\lambda^k}{k!}\,, \qquad k \in \mathbb{N}\,.$$

Espérance : λ

Variance : λ

Fonction caractéristique : $\exp\big(\lambda(e^{it}-1)\big)$

Stabilité par convolution : $\mathcal{P}(\lambda)*\mathcal{P}(\mu) = \mathcal{P}(\lambda+\mu)$. Autrement dit, si X et Y sont indépendantes et suivent respectivement des lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$, alors $X+Y$ est de loi $\mathcal{P}(\lambda+\mu)$.

Proposition. *Si X_λ suit une loi $\mathcal{P}(\lambda)$, alors $(X-\lambda)/\sqrt{\lambda}$ converge en loi quand $\lambda\to\infty$ vers une variable de loi normale $\mathcal{N}(0,1)$.*

4. Loi multinomiale

Définition. Un vecteur aléatoire $X = (X_1,\ldots,X_d)$, à valeurs dans $\mathbb{N}^d$, suit une loi multinomiale de paramètres $n \in \mathbb{N}$, $p_1,\ldots,p_d \in [\,0,1\,]$, $p_1+\cdots+p_d = 1$, notée $\mathcal{M}(n,p_1,\ldots,p_d)$, si

$$P\big\{\, X = (n_1,\ldots,n_d) \,\big\} = \frac{n!}{n_1!\ldots n_d!}\,p_1^{n_1}\ldots p_d^{n_d}\,, \qquad \begin{array}{l} n_1+\cdots+n_d = n\,, \\ n_1,\ldots,n_d \in \mathbb{N}\,. \end{array}$$

Espérance : $(np_1, \ldots, np_d)$

Covariance : $\operatorname{cov}(X_i, X_j) = -np_ip_j$, $i \neq j$

Variance : $\operatorname{Var}(X_i) = np_i(1-p_i)$

Fonction caractéristique : $\left(\sum_{1\leq j\leq d} p_j e^{it_j}\right)^n$

Si l'on dispose de n boules que l'on jette une par une aléatoirement dans d boîtes différentes, chaque boule ayant la probabilité p_i d'être jetée dans la i-ème boîte, les nombres $(N_1, \ldots, N_d)$ de boules dans les boîtes $1, \ldots, d$, suivent une loi multinomiale $\mathcal{M}(n, p_1, \ldots, p_d)$.

5. Loi hypergéométrique

Définition. Une variable aléatoire X, à valeurs entières, suit une loi hypergéométrique de paramètres (N, n, p) avec $Np \in \mathbb{N}^*$ et $p \in [\,0,1\,]$ si

$$P\{\,X = k\,\} = \frac{\mathrm{C}_{Np}^{k}\,\mathrm{C}_{N(1-p)}^{n-k}}{\mathrm{C}_{N}^{n}}\,, \qquad \max\big(0, n - N(1-p)\big) \leq k \leq \min(n, Np)\,.$$

Espérance : np

Variance : $\dfrac{N-n}{N-1}\,np(1-p)$

Si on tire n boules sans remise dans une urne en contenant N, une proportion p étant noires, $1-p$ étant blanches, le nombre de boules noires tirées suit une loi hypergéométrique de paramètres (N, n, p).

Proposition. *Si X_N suit une loi hypergéométrique de paramètres (N, n, p), alors X_N converge en loi quand $N \to \infty$ vers une variable de loi binomiale $\mathcal{B}(n, p)$.*

6. Loi binomiale négative

Définition. Une variable aléatoire X, à valeurs entières, suit une loi binomiale négative de paramètres $(n, p) \in \mathbb{N}^* \times [\,0,1\,]$ si

$$P\{\,X = k\,\} = \mathrm{C}_{n+k-1}^{n-1}\,p^n(1-p)^k\,, \qquad k \in \mathbb{N}\,.$$

Espérance : $n(1-p)/p$

Variance : $n(1-p)/p^2$

Fonction caractéristique : $\left(\dfrac{p}{1-(1-p)e^{it}}\right)^n$

Si $(X_i)_{i\geq 1}$ est une suite de variables aléatoires indépendantes et de même loi de Bernoulli $\mathcal{B}(1, p)$, représentant un succès si $X_i = 1$, un échec si $X_i = 0$, le nombre total d'échecs avant le n-ième succès suit une loi binomiale négative de paramètres (n, p).

Lorsque $n = 1$, on parle aussi de loi géométrique.

7. Loi uniforme continue

Définition. Une variable aléatoire X, à valeurs réelles, suit une loi uniforme sur $[\,a, b\,]$, $a < b$, notée $\mathcal{U}_{[a,b]}$, si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

$$f(x) = \frac{1}{b-a}\, \mathbb{1}_{[a,b]}(x)\,.$$

Espérance : $(a+b)/2$

Variance : $(b-a)^2/12$

Fonction caractéristique : $e^{ita}\dfrac{e^{itb}-e^{ita}}{it(b-a)}$

8. Loi de Paréto

Définition. Une variable aléatoire X, à valeurs positives, suit une loi de Paréto de paramètre $p > 1$ si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

$$f(x) = \frac{(p-1)}{x^p}\, \mathbb{1}_{[1,\infty[}(x)\,.$$

Espérance : $\dfrac{p-1}{p-2}$ si $p > 2$

Variance : $\dfrac{p-1}{(p-3)(p-2)^2}$ si $p > 3$

9. Loi gamma

Pour $p > 0$, on définit l'intégrale « gamma », $\Gamma(p) = \displaystyle\int_0^\infty x^{p-1}e^{-x}\,\mathrm{d}x$.

Définition. Une variable aléatoire X, à valeurs positives, suit une loi gamma de paramètres $p > 0$ et $\theta > 0$, notée $\gamma(p,\theta)$, si sa densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est

$$f(x) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1}\, \mathbb{1}_{[0,\infty[}(x)\,.$$

La loi $\gamma(1,\theta)$ est appelée loi exponentielle de paramètre θ, notée $\mathcal{E}xp(\theta)$.

Espérance : p/θ
Variance : p/θ^2
Fonction caractéristique : $\dfrac{1}{(1-i\theta t)^p}$
Stabilité par convolution : $\gamma(p,\theta) * \gamma(q,\theta) = \gamma(p+q,\theta)$. Autrement dit, si X et Y sont deux variables aléatoires indépendantes, de lois respectives $\gamma(p,\theta)$ et $\gamma(q,\theta)$, alors $X+Y$ est de loi $\gamma(p+q,\theta)$.

Proposition. *Si X_p suit une loi $\gamma(p,1)$, alors $(X_p - p)/\sqrt{p}$ converge en loi quand $p \to \infty$ vers une variable aléatoire de loi $\mathcal{N}(0,1)$.*

10. Loi béta

Pour $p, q > 0$, on définit l'intégrale « béta » par

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\, \mathrm{d}x = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\,.$$

Définition. Une variable aléatoire X, à valeurs sur $]\,0,1\,[$, suit une loi béta de première espèce de paramètres $p, q > 0$, notée $\beta(p,q)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}\, \mathbb{1}_{]0,1[}(x)\,.$$

Espérance : $B(p+1,q)/B(p,q)$
Variance : $\dfrac{pq}{(p+q)^2(p+q+1)}$

Proposition. *Si X et Y sont indépendantes et suivent respectivement une loi $\gamma(p)$ et $\gamma(q)$, alors $X/(X+Y)$ suit une loi $\beta(p,q)$. En outre $X/(X+Y)$ et $X+Y$ sont indépendantes.*

11. Loi de Laplace

Définition. Une variable aléatoire X, à valeurs réelles, suit une loi de Laplace (ou double exponentielle) si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{2}\, e^{-|x|}\,, \quad x \in \mathbb{R}\,.$$

Espérance : 0

Variance : 2

Fonction caractéristique : $1/(1+t^2)$

12. Loi normale unidimensionnelle

Définition. Une variable aléatoire X, à valeurs réelles, suit une loi normale $\mathcal{N}(m,\sigma^2)$ de moyenne m et variance $\sigma^2 > 0$ si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-m)^2}{2\sigma^2}\Big)\,, \qquad x \in \mathbb{R}\,.$$

Espérance : m

Variance : σ^2

Fonction caractéristique : $\exp\Big(itm - \dfrac{\sigma^2 t^2}{2}\Big)$

Stabilité par convolution : $\mathcal{N}(m_1,\sigma_1^2) * \mathcal{N}(m_2,\sigma_2^2) = \mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$. Autrement dit, si X_1 et X_2 sont indépendantes, de lois respectives $\mathcal{N}(m_1,\sigma_1^2)$ et $\mathcal{N}(m_2,\sigma_2^2)$, alors X_1+X_2 est de loi $\mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$.

13. Loi du chi-deux

Définition. Une variable aléatoire X, à valeurs positives, suit une loi du chi-deux $\chi^2(d)$ à $d \in \mathbb{N}^*$ degrés de liberté si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{2^{d/2}\Gamma(d/2)}\, x^{(d/2)-1} e^{-x/2}\, \mathbb{1}_{[0,\infty[}(x)\,.$$

En particulier, $X/2$ suit une loi $\gamma(d/2, 1/2)$.

Espérance : d

Variance : $2d$

Fonction caractéristique : $(1-2it)^{-d/2}$

Stabilité par convolution : $\chi^2(d_1) * \chi^2(d_2) = \chi^2(d_1+d_2)$. Autrement dit, si X_1 et X_2 sont des variables aléatoires indépendantes, de lois respectives $\chi^2(d_1)$ et $\chi^2(d_2)$, alors X_1+X_2 est de loi $\chi^2(d_1+d_2)$.

Proposition. *(i) Si $X_1, \ldots, X_d$ sont des variables gaussiennes centrées réduites indépendantes, alors $X_1^2 + \cdots + X_d^2$ suit une loi du chi-deux à d degrés de liberté.*

(ii) Si X_d suit une loi du chi-deux à d degrés de liberté, alors $(X_d - d)/\sqrt{d}$ converge en loi quand $d \to \infty$ vers une variable aléatoire de loi $\mathcal{N}(0,1)$.

14. Loi de Student

Définition. Une variable aléatoire X, à valeurs réelles, suit une loi de Student à $d \in \mathbb{N}^*$ degrés de liberté si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{\sqrt{d}B(1/2, d/2)}\left(1+\frac{x^2}{d}\right)^{-\frac{d+1}{2}}, \quad x \in \mathbb{R}.$$

Lorsque $d=1$, la loi de Student à d degrés de liberté s'appelle loi de Cauchy, et sa densité est

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}.$$

Espérance : 0 pour $d > 1$

Variance : $d/(d-2)$ pour $d > 2$

Fonction caractéristique : $e^{-|t|}$ pour la loi de Cauchy

Proposition. *Si Y est une variable normale centrée réduite, si Z suit une loi du chi-deux à d degrés de liberté, et si Y et Z sont indépendantes, alors $Y/\sqrt{Z/d}$ suit une loi de Student à d degrés de liberté. En particulier, lorsque $d=1$, si Y et Y' sont indépendantes de loi $\mathcal{N}(0,1)$, la variable aléatoire $Y/|Y'|$ suit une loi de Cauchy. Par symétrie, il en va de même de Y/Y'.*

15. Loi normale multidimensionnelle

Définition. Un vecteur aléatoire $X = (X_1, \dots, X_d)$, à valeurs dans $\mathbb{R}^d$, suit une loi normale $\mathcal{N}(m, \Gamma)$ de moyenne $m \in \mathbb{R}^d$ et de matrice de covariance inversible Γ si sa densité par rapport à la mesure de Lebesgue est

$$f(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\text{dét}\Gamma}} \exp\left(-\frac{1}{2}{}^{\text{t}}(x-m)\Gamma^{-1}(x-m)\right), \quad x \in \mathbb{R}^d .$$

Espérance : m

Covariance : Γ

Fonction caractéristique : $\exp\big(i\langle t, m\rangle - \frac{1}{2}{}^{\text{t}}t\,\Gamma t\big)$

Stabilité par convolution : $\mathcal{N}(m_1, \Gamma_1) * \mathcal{N}(m_2, \Gamma_2) = \mathcal{N}(m_1 + m_2, \Gamma_1 + \Gamma_2)$. Autrement dit, si X et Y sont deux vecteurs aléatoires indépendants, de lois respectives $\mathcal{N}(m_1, \Gamma_1)$ et $\mathcal{N}(m_2, \Gamma_2)$, alors $X + Y$ est de loi $\mathcal{N}(m_1 + m_2, \Gamma_1 + \Gamma_2)$.

INDEX TERMINOLOGIQUE

INDEX DES NOTATIONS

La référence est celle du premier emploi de la notation.

resp. signifie respectivement
$\mathbb{N}$ = ensemble des entiers naturels
$\mathbb{Z}$ = ensemble des entiers relatifs
$\mathbb{Q}$ = ensemble des rationnels
$\mathbb{R}$ = ensemble des réels
$\vee$ = maximum
$\wedge$ = minimum