

LICENCE de MATHÉMATIQUES
PROBABILITÉS

MIHAI GRADINARU

2001-2003

Avant propos

Ces notes sont une rédaction du cours oral en amphithéâtre donné pendant trois ans. Il s'agit d'un document de travail et pas d'un ouvrage; il est destiné à la distribution aux étudiants de Licence de Mathématiques de l'Université de Nancy. Ces notes sont inspirées librement de plusieurs notes de cours (et je remercie vivement leurs auteurs) rédigées par Ph. Barbe, J. Bertoin, J. Jacod, M. Ledoux et P. Vallois. Je remercie J. Rivat pour ses conseils en \LaTeX et S. Dabuleanu pour la lecture attentive des formes préliminaires du manuscrit.

Vandœuvre-lès-Nancy, janvier 2001 - mai 2003

M. Gradinaru

Table des matières

1	Espace de probabilité	1
1.1	Tribus	1
1.2	Variables aléatoires	2
1.3	Classes monotones	5
1.4	Probabilités	7
1.5	La construction d'une probabilité sur $(0,1]$	12
1.6	Loi d'une variable aléatoire	21
1.7	Lois de probabilités usuelles	33
1.8	Exercices	36
2	Espérance des variables aléatoires	43
2.1	Définitions et théorèmes de convergence	43
2.2	Théorème de transport	47
2.3	Moments, variance et covariance	50
2.4	Espaces L^p	54
2.5	Fonctions caractéristiques	56
2.6	Retour sur les lois de probabilités usuelles	62
2.7	Exercices	65
3	Indépendance	69
3.1	Indépendance	69
3.2	Sommes de variables aléatoires indépendantes	74
3.3	Applications de l'indépendance	76
3.4	Vecteurs gaussiens et indépendance	79
3.5	Probabilité (et espérance) conditionnelle	80
3.6	Exercices	83
4	Convergence des suites de variables aléatoires	87
4.1	Convergence presque sûre	87
4.2	Convergence en probabilité	89
4.3	Convergence dans L^p	93
4.4	Convergence en loi	97
4.5	Les lois des grands nombres et le théorème central limite	101
4.6	Exercices	108
5	Sujets d'examens 2001-2003	115

Chapitre 1

Espace de probabilité

1.1 Tribus

Une **expérience aléatoire** se décrit mathématiquement par la donnée d'un ensemble Ω (**univers**) dont les éléments notés ω sont les **résultats** (ou **issues**) possibles de l'expérience. Un **événement aléatoire** lié à l'expérience peut être représenté par une partie de Ω . Il sera toujours représenté par l'ensemble des résultats ω de l'expérience qui le réalisent. A priori il pourrait sembler naturel de considérer que toute partie de Ω représente un événement, mais cela n'est possible que si Ω est dénombrable. Pour des espaces plus grands, $\Omega = \mathbb{R}$ (ou \mathbb{R}^d ou un espace métrique E) on a besoin de la notion de tribu.

La description d'un événement comme une partie de Ω est à l'origine de la notation ensembliste. Le **contraire** de l'événement A (qui est réalisé si A ne l'est pas) correspond au complémentaire d'un ensemble $A \subset \Omega$ qui sera noté $A^c = \Omega \setminus A$. L'événement A **ou** B (qui est réalisé si au moins un des deux événements A ou B est réalisé) correspond à la réunion $A \cup B$. L'événement A **et** B (qui est réalisé si les événements A et B sont réalisés à la fois) correspond à l'intersection $A \cap B$. L'événement A **implique** l'événement B si A ne peut être réalisé sans que B le soit aussi et on note $A \subset B$. L'événement **impossible** sera noté \emptyset et l'événement **certain** sera noté Ω . A et B sont **incompatibles** si $A \cap B = \emptyset$.

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

Définition 1.1 *Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est une **tribu** (ou **σ -algèbre**) sur Ω si*

- a) $\Omega \in \mathcal{A}$
- b) \mathcal{A} est stable par passage au complémentaire $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- c) \mathcal{A} est stable par réunion dénombrable ($A_j \in \mathcal{A}, j \in \mathbb{N} \Rightarrow \cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$).

*Le couple (Ω, \mathcal{A}) est un **espace mesurable**.*

On appelle **événement** tout élément de la tribu \mathcal{A} . Si $\{\omega\} \in \mathcal{A}$ on appelle $\{\omega\}$ **événement élémentaire**.

REMARQUES: i) Par passage au complémentaire une tribu est aussi stable par intersection dénombrable.

ii) En remplaçant l'axiome (c) par

(c') \mathcal{A} est stable par réunion finie ($A_1, \dots, A_n \in \mathcal{A} \Rightarrow A_1 \cup \dots \cup A_n \in \mathcal{A}$)

on obtient la définition d'une **algèbre**. Toute tribu est une algèbre. \square

Exemples: i) $\mathcal{P}(\Omega)$ est toujours une tribu.

ii) $\{\emptyset, \Omega\}$ est une tribu appelée **tribu triviale**.

iii) La famille des ouverts de \mathbb{R}^d n'est pas une tribu car le complémentaire d'un ouvert n'est pas nécessairement un ouvert.

iv) Une réunion de deux tribus n'est pas une tribu en général. En effet, soit $\Omega = \{0,1,2\}$, $\mathcal{A}_1 = \{\emptyset, \{0\}, \{1,2\}, \Omega\}$ et $\mathcal{A}_2 = \{\emptyset, \{1\}, \{0,2\}, \Omega\}$. La réunion de $\{0\}$ et $\{1\}$ n'appartient pas à $\mathcal{A}_1 \cup \mathcal{A}_2$.

v) Une intersection d'un nombre quelconque de tribus est une tribu.

En général il n'est pas facile de décrire tous les éléments d'une tribu; on utilise le plus souvent leurs générateurs.

Définition 1.2 Soit \mathcal{E} un sous-ensemble de $\mathcal{P}(\Omega)$. La **tribu $\sigma(\mathcal{E})$ engendrée par \mathcal{E}** est l'intersection de toutes les tribus contenant \mathcal{E} ; elle est donc la plus petite tribu contenant \mathcal{E} .

REMARQUE: La tribu engendrée par deux tribus \mathcal{A}_1 et \mathcal{A}_2 est notée $\mathcal{A}_1 \vee \mathcal{A}_2 = \sigma(\mathcal{A}_1, \mathcal{A}_2) = \sigma(\mathcal{A}_1 \cup \mathcal{A}_2)$ qui est en général différente de $\mathcal{A}_1 \cup \mathcal{A}_2$. \square

Exemple: Soit A un sous-ensemble strict de Ω qui n'est pas vide. La tribu $\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$.

Définition 1.3 Si $\Omega = E$ est un espace métrique, on appelle **tribu borélienne**, notée $\mathcal{B}(E)$, la tribu engendrée par les ouverts de E . Tout élément de cette tribu est appelé **borélien**.

REMARQUE: La tribu borélienne est aussi engendrée par les fermés. Sur \mathbb{R} la tribu borélienne coïncide avec la tribu engendrée par les intervalles $]a, b[$ ou $[a, b]$, ou $]a, b]$, ou $[a, b[$, $-\infty \leq a < b \leq \infty$. \square

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

Définition 1.4 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.

REMARQUE*: En utilisant que tout ouvert de \mathbb{R}^2 peut s'écrire comme un réunion dénombrable de pavés d'intervalles ouverts, on montre que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. \square

1.2 Variables aléatoires

Une variable aléatoire sera définie par référence à une expérience aléatoire, comme une variable X dont la valeur dépend du résultat ω de cette expérience. Elle est donc une fonction définie sur l'univers Ω associé à l'expérience.

Pour définir une variable aléatoire on introduit d'abord quelques notations. Si X est une application de Ω dans E et si B est une partie de E on notera $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$. Si \mathcal{B} est une famille de parties de E , on notera $X^{-1}(\mathcal{B}) := \{X^{-1}(B) : B \in \mathcal{B}\}$.

Définition 1.5 a) Soient (Ω, \mathcal{A}) , (E, \mathcal{B}) deux espaces mesurables et $X : \Omega \rightarrow E$. On dit que X est **mesurable** (pour \mathcal{A} et \mathcal{B}) si $X^{-1}(\mathcal{B}) \subset \mathcal{A}$ (c'est-à-dire, pour tout $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$). Lorsque les deux espaces de départ et d'arrivée sont des espaces métriques munis de leurs tribus boréliennes la fonction est appelée **borélienne**.

b) Soient (E, \mathcal{B}) un espace mesurable et $X : \Omega \rightarrow E$. On appelle **tribu engendrée par X** (sur Ω) la plus petite tribu qui rend X mesurable: $\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}\} = X^{-1}(\mathcal{B})$.

c) Soit $\{X_t : t \in T\}$ une famille de fonctions de Ω dans un espace mesurable (E, \mathcal{B}) . La **tribu engendrée par $\{X_t : t \in T\}$** est la plus petite tribu qui rend mesurable toute fonction de $\{X_t : t \in T\}$: $\sigma(X_t : t \in T) := \sigma(\cup_{t \in T} \sigma(X_t))$.

REMARQUE: Dire que X est mesurable revient à dire que $\sigma(X) \subset \mathcal{A}$. □

Exemples: i) Pour A une partie de Ω , on définit la **fonction indicatrice** de A par

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

La fonction $\mathbf{1}_A$ est mesurable pour \mathcal{A} (en tant que fonction à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$) si et seulement si $A \in \mathcal{A}$.

ii) La tribu $\mathcal{B}(\mathbb{R}^2)$ est engendrée par les projections Π_1 et Π_2 sur les coordonnées ($\Pi_1(x, y) = x$ et $\Pi_2(x, y) = y$).

Définition 1.6 Une fonction mesurable X définie sur Ω muni de la tribu \mathcal{A} à valeurs dans \mathbb{R} (ou \mathbb{R}^d ou \mathbb{R}^∞) muni de sa tribu borélienne est appelée **variable aléatoire réelle** (ou **vecteur aléatoire** ou encore **suite aléatoire**).

En pratique, lorsque la tribu de l'espace d'arrivée est engendrée par un système de générateurs, pour vérifier qu'une fonction est mesurable (en particulier, variable aléatoire), il suffit de vérifier la propriété caractéristique sur les générateurs.

Proposition 1.1 1) Soient X de (Ω, \mathcal{A}) dans (E, \mathcal{B}) , et $\mathcal{C} \subset \mathcal{P}(E)$. On suppose que la tribu \mathcal{B} sur E est engendrée par \mathcal{C} , $\mathcal{B} = \sigma(\mathcal{C})$. Alors

$$\sigma(X) = X^{-1}(\mathcal{B}) = X^{-1}(\sigma(\mathcal{C})) = \sigma(X^{-1}(\mathcal{C})) := \sigma(\{X^{-1}(C) : C \in \mathcal{C}\}).$$

2) En particulier, pour qu'une fonction X de (Ω, \mathcal{A}) dans $(E, \sigma(\mathcal{C}))$ soit mesurable, il suffit que $X^{-1}(\mathcal{C}) \subset \mathcal{A}$.

Preuve: Il est clair que $X^{-1}(\sigma(\mathcal{C}))$ est une tribu contenant $X^{-1}(\mathcal{C})$, d'où $X^{-1}(\sigma(\mathcal{C})) \supset \sigma(X^{-1}(\mathcal{C}))$. Pour l'autre inclusion, soit

$$\mathcal{T} = \{B \subset E : X^{-1}(B) \in \sigma(X^{-1}(\mathcal{C}))\}.$$

On peut vérifier facilement que \mathcal{T} est une tribu. Par sa définition $X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C}))$. De plus $\mathcal{C} \subset \mathcal{T}$ car $X^{-1}(\mathcal{C}) \subset \sigma(X^{-1}(\mathcal{C}))$, et donc $\sigma(\mathcal{C}) \subset \mathcal{T}$. On déduit que $X^{-1}(\sigma(\mathcal{C})) \subset X^{-1}(\mathcal{T}) \subset \sigma(X^{-1}(\mathcal{C}))$. Notons qu'on peut traiter de la même façon le cas d'une famille quelconque de fonctions.

Si $X^{-1}(\mathcal{C}) \subset \mathcal{A}$, alors $\sigma(X^{-1}(\mathcal{C})) \subset \mathcal{A}$. Comme $\sigma(X) = \sigma(X^{-1}(\mathcal{C}))$ par le premier point, la conclusion s'ensuit. \square

REMARQUE: Une fonction $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire réelle si et seulement si pour tout réel t , $\{X \leq t\} \in \mathcal{A}$. \square

Proposition 1.2 *La composée de deux fonctions mesurables est mesurable. En particulier, la composée $Y = g(X)$ d'une fonction borélienne g avec une variable aléatoire X est une variable aléatoire.*

Preuve: Soient $X_j : (\Omega_j, \mathcal{A}_j) \rightarrow (\Omega_{j+1}, \mathcal{A}_{j+1})$, $j = 1, 2$. Si $A \in \mathcal{A}_3$, $(X_2 \circ X_1)^{-1}(A) = X_1^{-1}(X_2^{-1}(A))$. Comme X_2 est mesurable, $X_2^{-1}(A) \in \mathcal{A}_2$, et comme X_1 est mesurable, $X_1^{-1}(X_2^{-1}(A)) \in \mathcal{A}_1$. \square

Proposition 1.3 *Soient E_1 et E_2 deux espaces métriques munis de leurs tribus boréliennes. Toute fonction continue $f : E_1 \rightarrow E_2$ est borélienne.*

Preuve: On remarque que si O est un ouvert dans E_2 et f est une fonction continue, $f^{-1}(O)$ est un ouvert. Puis on applique la Proposition 1.1. \square

Proposition 1.4 *Si X, Y sont deux variables aléatoires de Ω dans \mathbb{R} , alors l'application $\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2$ est un vecteur aléatoire bi-dimensionnel. La réciproque est aussi vraie.*

Preuve: Soit $A \times B$ un pavé dans $\mathcal{B}(\mathbb{R}^2)$ et $Z(\omega) = (X(\omega), Y(\omega))$. Alors, $Z^{-1}(A \times B) = X^{-1}(A) \cap Y^{-1}(B) \in \mathcal{A}$. Comme les pavés engendrent $\mathcal{B}(\mathbb{R}^2)$, on conclut grâce à la Proposition 1.1. (voir aussi la Proposition 1.7). Pour la réciproque il suffit de voir que X et Y sont des composées des applications mesurables projections (continues) et Z . \square

Proposition 1.5 *L'espace des variables aléatoires réelles est stable pour les opérations suivantes: $(\alpha X)(\omega) = \alpha X(\omega)$ ($\alpha \in \mathbb{R}$), $(X + Y)(\omega) = X(\omega) + Y(\omega)$, $(XY)(\omega) = X(\omega)Y(\omega)$ et $(X \vee Y)(\omega) = X(\omega) \vee Y(\omega)$ (le maximum de X, Y).*

Preuve: La fonction $\omega \mapsto \alpha X(\omega)$ est la composée de la variable aléatoire X et la fonction continue $x \mapsto \alpha x$. De même, $X + Y$ (respectivement XY , respectivement $X \vee Y$) est la composée du vecteur aléatoire (X, Y) (en vertu de la Proposition 1.3) et de la fonction continue $(x, y) \mapsto x + y$ (respectivement $(x, y) \mapsto xy$, respectivement $(x, y) \mapsto x \vee y$). \square

Une limite ponctuelle de fonctions continues n'est pas nécessairement continue. Pour les fonctions mesurables (et en particulier pour les variables aléatoires) on peut montrer le suivant:

Théorème 1.1 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires de Ω dans un espace métrique*

E. Si cette suite de fonctions converge ponctuellement vers X (c'est-à-dire, pour tout $\omega \in \Omega$, $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$), alors X est une variable aléatoire à valeurs dans E .

Preuve: D'après la Proposition 1.1, il suffit de montrer que si O est un ouvert dans E , alors $X^{-1}(O) \in \mathcal{A}$. On pose

$$O_r := \left\{ x \in O : d(x, E \setminus O) > \frac{1}{r} \right\}, r \in \mathbb{N}^*.$$

L'ensemble O_r est ouvert, donc un borélien de E . Ainsi,

$$X^{-1}(O) = \bigcup_{r, m \in \mathbb{N}^*} \bigcap_{n \geq m} X_n^{-1}(O_r)$$

est un événement de \mathcal{A} . □

Définition 1.7 On appelle **variable aléatoire étagée ou simple** (à valeurs dans \mathbb{R}^d) une variable aléatoire définie sur Ω de la forme $X = \sum_{j=1}^r a_j \mathbf{1}_{A_j}$ où les A_1, \dots, A_r sont des événements disjoints dans \mathcal{A} , et où les coefficients $a_1, \dots, a_r \in \mathbb{R}^d$.

Proposition 1.6 Toute variable aléatoire X est limite simple de variables aléatoires étagées. Si de plus X est une variable aléatoire positive, la limite peut être choisie croissante.

Preuve: Soit d'abord X positive. On définit, pour $n, k \in \mathbb{N}^*$,

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

et

$$B_n := \{\omega : X(\omega) > n\}.$$

Les $A_{n,k}$ et les B_n sont éléments de \mathcal{A} en tant qu'images réciproques par la variable aléatoire X d'intervalles. La suite

$$X_n(\omega) := \sum_{k=1}^{2^n} \frac{k-1}{2^n} \mathbf{1}_{A_{n,k}}(\omega) + n \mathbf{1}_{B_n}(\omega)$$

converge en croissant vers $X(\omega)$.

Si X est quelconque, on écrit $X = X^+ - X^-$ avec $X^+ = X \vee 0$ et $X^- = (-X) \vee 0$, et on approxime les variables positives X^+ et X^- par la méthode précédente. □

1.3 Classes monotones

Dans cette section on va construire un procédé d'extension des définitions de certains objets sur les tribus après les avoir définis sur des classes restreintes d'ensembles.

Définition 1.8 Une famille \mathcal{M} de parties de Ω est appelée **classe monotone** si

a) $\Omega \in \mathcal{M}$

- b) lorsque $A, B \in \mathcal{M}$ et $B \subset A$, alors $A \setminus B \in \mathcal{M}$
c) \mathcal{M} est stable par réunion monotone croissante ($A_j \in \mathcal{M}, j \in \mathbb{N}, A_j \subset A_{j+1} \Rightarrow \bigcup_{j \in \mathbb{N}} A_j \in \mathcal{M}$). Pour $\mathcal{E} \subset \mathcal{P}(\Omega)$, on note $\mathcal{M}(\mathcal{E})$ la **classe monotone** engendrée par \mathcal{E} , c'est-à-dire l'intersection de toutes les classes monotones contenant \mathcal{E} .

REMARQUE: Une intersection d'un nombre quelconque de classes monotones est une classe monotone. \square

Exemples: i) Une tribu est une classe monotone. En effet, pour voir cela, il suffit de voir que $A \setminus B = A \cap B^c$.

ii) Une classe monotone stable par intersection finie est une tribu. En effet cette classe sera aussi stable par réunion finie en vertu de l'axiome (b) de la Définition 8, et toute réunion peut s'écrire comme une réunion croissante ($\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} (\bigcup_{k \leq j} A_k)$), pour toute famille A_j , $j \in \mathbb{N}$).

Théorème 1.2 (des classes monotones)

Soit \mathcal{E} une famille de parties de Ω stable par intersection finie. Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.

REMARQUE: On pourrait énoncer ce résultat sous la forme suivante: si \mathcal{M}_1 est une classe monotone contenant la famille de parties \mathcal{E} (stable par intersection finie), alors $\mathcal{M}_1 \supset \sigma(\mathcal{E})$.

Preuve: En vertu de l'exemple i) ci-dessus, $\sigma(\mathcal{E})$ est une classe monotone qui contient \mathcal{E} et donc $\mathcal{M}(\mathcal{E}) \subset \sigma(\mathcal{E})$. Pour prouver l'inclusion inverse, on montre que $\mathcal{M}(\mathcal{E})$ est stable par intersection finie (car alors, d'après l'exemple 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

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

L'ensemble \mathcal{M}_2 est une classe monotone. De plus il contient \mathcal{E} : on doit pour cela montrer que si $B \in \mathcal{E}$, alors $\forall C \in \mathcal{M}(\mathcal{E}), B \cap C \in \mathcal{M}(\mathcal{E})$. Or $C \in \mathcal{M}(\mathcal{E}) \subset \mathcal{M}_1$, 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 prouvé. \square

Proposition 1.7 Soit $X : \Omega \rightarrow \mathbb{R}^d$ une fonction vectorielle $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$ sur l'espace mesurable (Ω, \mathcal{A}) . X est un vecteur aléatoire si et seulement si chaque coordonnée X_j est une variable aléatoire réelle.

Preuve: On va faire la preuve pour $d = 2$. On suppose d'abord que X est un vecteur aléatoire et soit $A \in \mathcal{B}(\mathbb{R})$. Alors

$$\begin{aligned} X_1^{-1}(A) &= \{\omega \in \Omega : X_1(\omega) \in A\} \\ &= \{\omega \in \Omega : (X_1(\omega), X_2(\omega)) \in A \times \mathbb{R}\} = X^{-1}(A \times \mathbb{R}), \end{aligned}$$

donc X_1 est une variable aléatoire. De la même façon on montre que X_2 est une variable aléatoire.

Réciproquement, supposons que X_1 et X_2 sont deux variables aléatoires réelles. Soient $A, B \in \mathcal{B}(\mathbb{R})$. On a

$$\begin{aligned} X^{-1}(A \times B) &= \{\omega \in \Omega : X_1(\omega) \in A, X_2(\omega) \in B\} \\ &= X_1^{-1}(A) \cap X_2^{-1}(B) \in \mathcal{A}. \end{aligned}$$

Cette propriété d'appartenance s'étend à l'ensemble \mathcal{E} des réunions finies de pavés de \mathbb{R}^2 , deux à deux disjoints, laquelle engendre la tribu borélienne produit $\mathcal{B}(\mathbb{R}^2)$. De plus \mathcal{E} est stable par intersection finie. Donc $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}^2)$. On introduit

$$\mathcal{M}_1 = \{C \in \mathcal{B}(\mathbb{R}^2) : X^{-1}(C) \in \mathcal{A}\}$$

Puisque, $X^{-1}(\cup_{j \in \mathbb{N}} C_j) = \cup_{j \in \mathbb{N}} X^{-1}(C_j)$ pour toute famille C_j , $j \in \mathbb{N}$, on peut montrer que \mathcal{M}_1 est une classe monotone. De plus \mathcal{M}_1 contient \mathcal{E} , donc $\mathcal{M}_1 \supset \mathcal{M}(\mathcal{E}) = \mathcal{B}(\mathbb{R}^2)$. Par conséquent X est un vecteur aléatoire (voir aussi la Proposition 1.3). \square

1.4 Probabilités

Définition 1.9 Soit (Ω, \mathcal{A}) un espace mesurable. On appelle **probabilité** toute application $P : \mathcal{A} \rightarrow [0, 1]$ telle que

a) $P(\Omega) = 1$

b) P est σ -additive ($A_j \in \mathcal{A}$, $j \in \mathbb{N}$, $A_i \cap A_j = \emptyset$ si $i \neq j \Rightarrow P(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} P(A_j)$).

Un espace mesurable muni d'une probabilité (Ω, \mathcal{A}, P) est appelé **espace de probabilité**.

REMARQUE: Observons que \emptyset est la réunion dénombrable et disjointe de l'ensemble vide, donc $P(\emptyset) = \sum_{j \in \mathbb{N}} P(\emptyset)$, d'où $P(\emptyset) = 0$. Observons aussi que $\Omega = A \cup A^c \cup \emptyset \cup \dots$, d'où $P(A^c) = 1 - P(A)$. De la même façon on peut prouver que P est (finie) additive. \square

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

$$\delta_\omega : \mathcal{A} \rightarrow \{0, 1\}, \delta_\omega(A) := \mathbf{1}_A(\omega)$$

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

ii) Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ muni de la tribu des parties. On voit que $P(A) := \text{card}(A)/6$ est une probabilité. On peut voir que $P := \frac{1}{6} \sum_{j=1}^6 \delta_j$. Cette probabilité sert à modéliser le jet du dé.

iii) Soit $0 < p < 1$ et $\Omega = \{0, 1\}$. La probabilité $P := p\delta_1 + (1-p)\delta_0$ est appelée **probabilité de Bernoulli de paramètre p** . Lorsque $p = 1/2$ cette probabilité sert à modéliser le jet d'une pièce (jeu de pile ou face). Plus généralement, toute probabilité concentrée en deux points distincts sera appelée probabilité de Bernoulli.

Proposition 1.8 Soit (Ω, \mathcal{A}, P) un espace de probabilité et $(A_j)_{j \in J}$, $J \subset \mathbb{N}$ une famille finie ou dénombrable d'événements.

1) P est croissante: $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$;

2) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$;

3) Si $A_j \subset A_{j+1}$ pour tout j , alors $P(\cup_j A_j) = \lim_{j \rightarrow \infty} P(A_j)$;

4) Si $A_j \supset A_{j+1}$ pour tout j , alors $P(\cap_j A_j) = \lim_{j \rightarrow \infty} P(A_j)$;

5) P est sous-additive: $P(\cup_{j \in J} A_j) \leq \sum_{j \in J} P(A_j)$.

Preuve: 1) A_2 est la réunion disjointe des événements A_1 et $A_2 \setminus A_1$ et la Définition 9 d'une probabilité fournit $P(A_2) = P(A_1) + P(A_2 \setminus A_1) \geq P(A_1)$.

2) De la même façon

$$A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1), A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2),$$

et on en obtient l'égalité, puisque les paires d'événements A_1 et $A_2 \setminus A_1$, $A_2 \setminus A_1$ et $A_1 \cap A_2$ sont aussi disjoints.

3) Soit $B_k = A_{k+1} \setminus A_k$, $k \in \mathbb{N}$. Les ensembles B_k sont disjoints et comme $A_j = A_0 \cup \cup_{0 \leq k \leq j-1} B_k$, $j \in \mathbb{N}$, on a

$$\cup_j A_j = A_0 \cup \cup_k B_k.$$

Par σ -additivité on obtient

$$\begin{aligned} P\left(\cup_j A_j\right) &= P(A_0) + \sum_k P(B_k) = P(A_0) + \lim_{j \uparrow \infty} \sum_{k=0}^{j-1} P(B_k) \\ &= \lim_{j \uparrow \infty} \left(P(A_0) + \sum_{k=0}^{j-1} P(B_k) \right) = \lim_{j \uparrow \infty} P(A_j). \end{aligned}$$

4) Les $B_j = A_j^c$, forment une suite croissante, donc comme toute probabilité est bornée par 1, la suite $\{P(B_j) : j \in \mathbb{N}\}$ est une suite croissante bornée. Donc la limite $\lim_{j \uparrow \infty} P(B_j)$ existe et, d'après 3),

$$\begin{aligned} \lim_{j \uparrow \infty} P(B_j) &= P\left(\cup_j B_j\right) = P\left(\Omega \setminus \bigcap_j A_j\right) \\ &= P(\Omega) - P\left(\bigcap_j A_j\right). \end{aligned}$$

Donc,

$$P\left(\bigcap_j A_j\right) = 1 - \lim_{j \uparrow \infty} P(B_j) = \lim_{j \uparrow \infty} [1 - P(B_j)] = \lim_{j \uparrow \infty} P(A_j),$$

ce qui prouve l'assertion.

5) Si J est fini, par exemple $J = \{1, 2, \dots, r\}$, on procède par récurrence en remarquant que, par 2),

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2).$$

Si J est infini on peut supposer $J = \mathbb{N}$. On a alors, pour tout $k \in \mathbb{N}^*$,

$$P\left(\bigcup_{0 \leq j \leq k} A_j\right) \leq \sum_{j \in \mathbb{N}} P(A_j).$$

En considérant les ensembles croissants $B_k = \bigcup_{0 \leq j \leq k} A_j$ le résultat se déduit de 3). \square

REMARQUE: Soit $\Omega = \{\omega_0, \omega_1, \dots\}$ un ensemble (au plus) dénombrable muni de la tribu de ses parties. La formule :

$$P(A) := \sum_{\{j: \omega_j \in A\}} p_j, \quad A \in \mathcal{P}(\Omega)$$

définit une probabilité pourvu que la suite de réels $(p_j)_{j \geq 0}$ vérifie les deux conditions suivantes :

$$p_j \geq 0, j \geq 0 \text{ et } \sum_{j \geq 0} p_j = 1.$$

Inversement, si P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, en posant $p_j := P(\{\omega_j\})$, on voit que la suite de réels positifs ou nuls $(p_j)_{j \geq 0}$ satisfait $\sum_{j \geq 0} p_j = P(\Omega) = 1$, et que pour tout événement A , $P(A) = \sum_{\omega_j \in A} p_j$.

En particulier, lorsque Ω est fini, si par symétrie (donnée par la nature de l'expérience) tous les p_j sont égaux entre eux, alors nécessairement $p_j = 1/\text{card}(\Omega)$ (les résultats sont équiprobables) et la probabilité correspondante est la probabilité **uniforme** $P(A) = \text{card}(A)/\text{card}(\Omega)$, $A \in \mathcal{P}(\Omega)$. Cette probabilité uniforme modélise la plupart de jeux de hasard. Les calcul de probabilités d'événements sont des calculs de cardinaux d'ensembles (dénombrement). \square

REMARQUE: Soient $(\Omega, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et f une fonction borélienne $f: \mathbb{R}^d \rightarrow \mathbb{R}$. La formule

$$P(A) := \int_A f(x) dx, \quad A \in \mathcal{B}(\mathbb{R}^d)$$

(intégrale de Lebesgue) définit une probabilité pourvu que $f \geq 0$ et d'intégrale de Lebesgue $\int_{\mathbb{R}^d} f(x) dx = 1$. Inversement si la probabilité P sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est **absolument continue par rapport à la mesure de Lebesgue** λ , c'est-à-dire si $\lambda(A) = 0 \Rightarrow P(A) = 0$, alors (théorème de Radon-Nikodym)

$$\exists f \geq 0, \text{ borélienne} : P(A) = \int_A f(x) dx, \forall A \in \mathcal{B}(\mathbb{R}^d).$$

On dit que f est la **densité** de P par rapport à λ et elle est unique λ -p.p. \square

Preuve du théorème de Radon-Nikodym (facultatif).

Sans perte de généralité on peut supposer $d = 1$. Il suffit de prouver le résultat suivant:

(E₁) Soit P une probabilité sur $(]0,1], \mathcal{B}(]0,1])$ telle qu'elle soit absolument continue par rapport à la mesure de Lebesgue λ sur $]0,1]$. Alors il existe une fonction (unique λ -p.p.) borélienne positive f telle que $P(A) = \int_A f(x) \lambda(dx)$, pour tout $A \in \mathcal{B}(]0,1])$.

En effet, on écrit

$$\mathbb{R} = \bigcup_{n \in \mathbb{Z}} I_n, \text{ avec } I_n =]n, n+1].$$

Alors $P(\mathbb{R}) = \sum_{n \in \mathbb{Z}} P(I_n)$. On note $p_n = P(I_n)$, $n \in \mathbb{Z}$. Alors $P_n = \frac{1}{p_n} P$ est une proba sur I_n et elle est absolument continue par rapport à λ . Par (E₁) on déduit l'existence des fonctions

$f_n \geq 0$ boréliennes telles que $P_n(A) = \int_A f_n(x) \lambda(dx)$, pour tout $A \in \mathcal{B}(I_n)$, pour tout $n \in \mathbb{Z}$. Il suffit de prendre la fonction borélienne positive $f = \sum_{n \in \mathbb{Z}} \frac{1}{p_n} f_n$ pour avoir la conclusion.

On prouve donc (E₁); on commence par introduire

$$\mathcal{G} := \{g :]0,1[\rightarrow]0,\infty[\text{ borélienne} : P(A) \geq \int_A g(x) \lambda(dx), \forall A \in \mathcal{B}(]0,1[)\}.$$

Il est clair que $g \equiv 0$ est dans \mathcal{G} . On montre que \mathcal{G} est sup-stable, c'est-à-dire que si $g, h \in \mathcal{G}$ alors $\sup(g, h) \in \mathcal{G}$. En effet, on note $A_1 = \{g \geq h\}$ et $A_2 = A_1^c$ et soit $A \in \mathcal{B}(]0,1[)$. Alors

$$\int_A \sup(g, h) d\lambda = \int_{A \cap A_1} g d\lambda + \int_{A \cap A_2} h d\lambda \leq P(A \cap A_1) + P(A \cap A_2) = P(A).$$

On note

$$\gamma = \sup_{g \in \mathcal{G}} \int_{]0,1[} g d\lambda.$$

Il est clair que $\gamma \leq 1$, puisque $\int_{]0,1[} g d\lambda \leq P(]0,1[) = 1$, pour tout $g \in \mathcal{G}$. Il existe une suite $\{g_n^* : n \geq 1\} \subset \mathcal{G}$ telle que $\lim_{n \rightarrow \infty} \int_{]0,1[} g_n^* d\lambda = \gamma$. Alors la suite $\{g_n : n \geq 1\}$, avec $g_n := \sup(g_1^*, \dots, g_n^*)$ est aussi dans \mathcal{G} . Comme $g_n^* \leq g_n$, $\int_{]0,1[} g_n^* d\lambda \leq \int_{]0,1[} g_n d\lambda$, pour tout $n \geq 1$. On en déduit que $\lim_{n \rightarrow \infty} \int_{]0,1[} g_n d\lambda = \gamma$. Comme $\{g_n : n \geq 1\}$ est croissante donc par convergence monotone $\int_{]0,1[} (\lim_{n \rightarrow \infty} g_n) d\lambda = \gamma$. Notons $f = \sup_{n \geq 1} g_n \in \mathcal{G}$. Comme $\int_{]0,1[} f d\lambda = \gamma$, l'application $g \mapsto \int_{]0,1[} g d\lambda$ atteint son maximum sur \mathcal{G} en f . Montrons que $P(A) = \int_A f d\lambda$. Il est clair que $P(A) \geq \int_A f d\lambda$, car $f \in \mathcal{G}$. On définit

$$\nu(A) = P(A) - \int_A f d\lambda.$$

Si $\lambda(A) = 0$ alors $P(A) = 0$, par hypothèse et aussi $\int_A f d\lambda = 0$, donc $\nu(A) = 0$. Montrons que $\nu \equiv 0$. Supposons $\nu(]0,1[) > 0$ et on note $0 < \beta = \frac{1}{2} \nu(]0,1[)$. Alors $\nu(]0,1[) = 2\beta > \beta = \beta \lambda(]0,1[)$. On a besoin de résultat suivant:

(E₂) Soient μ et ν deux mesures finies sur un espace mesurable (Ω, \mathcal{A}) telles que $\mu(\Omega) < \nu(\Omega)$. Alors il existe $A' \in \mathcal{A}$ tel que $\mu(A') < \nu(A')$ et $\mu(A'') \leq \nu(A'')$ pour tous $A'' \in A' \cap \mathcal{A}$.

De ce résultat (avec ν et $\mu = \beta \lambda$) on déduit qu'il existe $A' \in \mathcal{B}(]0,1[)$ tel que $\nu(A') > \beta \lambda(A')$ et $\nu(A'') \geq \beta \lambda(A'')$, pour tous $A'' \in A' \cap \mathcal{B}(]0,1[)$. On note $f_0 = f + \beta \mathbb{1}_{A'}$. Alors

$$\int_A f_0 d\lambda = \int_A f d\lambda + \beta \lambda(A \cap A') \leq \int_A f d\lambda + \nu(A) = P(A), \forall A \in \mathcal{B}(]0,1[),$$

donc $f_0 \in \mathcal{G}$. De plus

$$\int_{]0,1[} f_0 d\lambda = \int_{]0,1[} f d\lambda + \beta \lambda(A') = \gamma + \beta \lambda(A') > \gamma,$$

puisque $\lambda(A') > 0$ (par $\nu(A') > \beta \lambda(A')$ et absolue continuité de ν par rapport à λ). Mais l'inégalité $\int_{]0,1[} f_0 d\lambda > \gamma$ contredit le choix de γ . L'hypothèse $\nu(]0,1[) > 0$ est donc fausse. Pour finir la preuve il faut justifier (E₂).

La fonction $\delta = \nu - \mu$ est bornée sur \mathcal{A} puisque $-\mu(\Omega) \leq \delta(A) \leq \nu(\Omega)$, pour tous $A \in \mathcal{A}$. Par récurrence on définit $\{B_n : n \geq 0\}$, $\{A_n : n \geq 0\}$:

$$B_0 := \emptyset, A_0 := \Omega = \Omega \setminus B_0$$

Si B_0, \dots, B_n et A_0, \dots, A_n sont définis, on pose

$$\alpha_n := \inf_{A \in A_n \cap \mathcal{A}} \delta(A).$$

Ainsi $\alpha_n \leq \delta(\emptyset) = 0$. Dans le cas $\alpha_n = 0$, on pose $B_{n+1} := \emptyset$ et $A_{n+1} := A_n = A_n \setminus B_{n+1}$. Si $\alpha_n < 0$, on choisit $B_{n+1} \in A_n \cap \mathcal{A}$ tel que $\delta(B_{n+1}) \leq \frac{1}{2}\alpha_n$ et on pose $A_{n+1} := A_n \setminus B_{n+1}$. La suite $\{B_n : n \geq 0\}$ est disjointe, donc la série $\sum_{n \geq 0} \delta(B_n)$ est convergente (vers $\nu(\cup B_n) - \mu(\cup B_n)$). Ainsi $\lim_{n \rightarrow \infty} \delta(B_n) = 0$, d'où $\lim_{n \rightarrow \infty} \alpha_n = 0$. On pose

$$A' := \bigcup_{n \geq 0} A_n.$$

Par les propriétés de continuité des mesures μ et ν similaires aux celles des probabilités, comme $\{A_n : n \geq 0\}$ est décroissante,

$$\delta(A') = \lim_{n \rightarrow \infty} \delta(A_n).$$

Montrons que A' satisfait aux propriétés. Comme $\delta(B_{n+1}) \leq 0$, pour tout n on a

$$\delta(A_{n+1}) = \delta(A_n) - \delta(B_{n+1}) \geq \delta(A_n) \geq \delta(A_{n-1}) \geq \dots \geq \delta(A_0) = \delta(\Omega) > 0,$$

d'où $\nu(A') - \mu(A') = \delta(A') = \lim_{n \rightarrow \infty} \delta(A_n) > 0$. Soit $A'' \in A' \cap \mathcal{A}$. Alors $A'' \in A_n \cap \mathcal{A}$ et donc $\delta(A'') \geq \alpha_n$ pour tout n . Comme $\lim_{n \rightarrow \infty} \alpha_n = 0$, on trouve $\nu(A'') - \mu(A'') = \delta(A'') \geq 0$.

Passons à la preuve de l'unicité. Soient f et g deux fonctions boréliennes positives telles que $P(A) = \int_A f(x)\lambda(dx) = \int_A g(x)\lambda(dx)$, pour tout $A \in \mathcal{B}(]0,1])$. Alors $\int_A (f-g)(x)\lambda(dx) = 0$, pour tout $A \in \mathcal{B}(]0,1])$. Soit $B = \{f \leq g\} \in \mathcal{B}(]0,1])$. Alors $\int_B (f-g)(x)\lambda(dx) = 0$ et $f-g \geq 0$ sur B . On en déduit que $f-g = 0$ p.p. sur B . De même, en considérant $B^c \in \mathcal{B}(]0,1])$, on déduit que $f-g = 0$ p.p. sur B^c . Alors $f-g = 0$ p.p. sur $]0,1]$. \square

Définition 1.10 Soient (Ω, \mathcal{A}, P) un espace de probabilité. On dit qu'un ensemble N est **négligeable** s'il existe un événement $B \in \mathcal{A}$ tel que $N \subset B$ et $P(B) = 0$.

On dit qu'une propriété $\text{Prop}(\omega)$ est vérifiée **P-presque sûrement** si l'ensemble $\{\omega : \text{Prop}(\omega) : \text{est fausse}\}$ est négligeable.

Une tribu sur Ω est dite **complète** si elle contient tous les ensembles négligeables.

REMARQUE: On peut toujours supposer, sans perte de généralité, que l'espace de probabilité est complet. \square

Exemples: i) Soit $X : \Omega = \{1,2,3\} \rightarrow \{0,1\}$ mesurable définie par $X(1) = X(2) = 1$, $X(3) = 0$ et soit P la probabilité définie par $P(\{1\}) = P(\{2\}) = 1/2$, $P(\{3\}) = 0$. Alors X est constante et égale à 1 p.s.

ii) Soit l'intervalle $\Omega = [0,1]$ muni de la tribu borélienne et de la mesure de Lebesgue λ . Soit

$$X(\omega) := \begin{cases} 1, & \text{si } \omega \in \mathbb{Q} \cap [0,1] \\ 0, & \text{sinon.} \end{cases}$$

Alors $X = 0$, p.s. (car $\lambda(\mathbb{Q} \cap [0,1]) = 0$).

De même, la variable aléatoire

$$Y(\omega) := \text{sign}(\omega - \frac{1}{2}) = \begin{cases} 1, & \text{si } \omega > 1/2 \\ 0, & \text{si } \omega = 1/2 \\ -1, & \text{si } \omega < 1/2. \end{cases}$$

Alors Y est continue p.s. (car son seul point de discontinuité est $\frac{1}{2}$ et $\lambda(\{1/2\}) = 0$).

1.5 La construction d'une probabilité sur $(0,1]$

i) Soit

$$\Omega = (0,1], \mathcal{B} = \mathcal{B}_{(0,1]} \text{ et } \mathcal{S} = \{(a,b] : 0 \leq a \leq b \leq 1\}.$$

Il est clair que $\emptyset, \Omega \in \mathcal{S}$.

Si $I_1, I_2 \in \mathcal{S}$, alors $I_1 \cap I_2 \in \mathcal{S}$. Donc \mathcal{S} est stable par intersection finie.

Enfin, si $I \in \mathcal{S}$, alors I^c est une union disjointe de deux intervalles de \mathcal{S} (faire un dessin).

De plus on sait que $\sigma(\mathcal{S}) = \mathcal{B}$ la tribu borélienne sur Ω .

ii) On définit sur \mathcal{S} la fonction $\lambda : \mathcal{S} \rightarrow [0,1]$, par :

$$\lambda(\emptyset) = 0 \text{ et } \lambda((a,b]) = b - a.$$

On a $\lambda(\Omega) = \lambda((0,1]) = 1$.

Montrons que λ est une fonction additive sur \mathcal{S} . Soit $(a,b] \in \mathcal{S}$ et supposons que :

$$(a,b] = \bigcup_{i=1}^r (a_i, b_i],$$

où les intervalles dans le membre droit sont disjoints. Supposons aussi que ces intervalles ont été numérotés convenablement :

$$a_1 = a, b_r = b, b_i = a_{i+1}, i = 1, \dots, r-1.$$

Alors $\lambda((a,b]) = b - a$ et

$$\sum_{i=1}^r \lambda((a_i, b_i]) = \sum_{i=1}^r (b_i - a_i) = b_1 - a_1 + b_2 - a_2 + \dots + b_r - a_r = b_r - a_1 = b - a.$$

iii) Montrons maintenant que λ est σ -additive. Soit $(a,b] \in \mathcal{S}$ et supposons que :

$$(a,b] = \bigcup_{i=1}^{\infty} (a_i, b_i],$$

où les intervalles dans le membre droit sont à nouveau disjoints.

iii)-a) Montrons d'abord que

$$\lambda((a,b]) = b - a \leq \sum_{i=1}^{\infty} (b_i - a_i) = \sum_{i=1}^{\infty} \lambda((a_i, b_i]). \quad (1.1)$$

On choisit $\varepsilon < b - a$ et on observe que

$$[a + \varepsilon, b] \subset \bigcup_{i=1}^{\infty} \left(a_i, b_i + \frac{\varepsilon}{2^i} \right).$$

Le membre droit est un recouvrement ouvert du compact $[a + \varepsilon, b]$, donc on peut en extraire un recouvrement fini : il existe un entier N tel que

$$[a + \varepsilon, b] \subset \bigcup_{i=1}^N \left(a_i, b_i + \frac{\varepsilon}{2^i} \right). \quad (1.2)$$

Pour terminer la preuve de (1.1) il suffit de prouver que

$$b - a - \varepsilon \leq \sum_{i=1}^N \left(b_i - a_i + \frac{\varepsilon}{2^i} \right). \quad (1.3)$$

En effet on aurait alors

$$b - a - \varepsilon \leq \sum_{i=1}^N \left(b_i - a_i + \frac{\varepsilon}{2^i} \right) \leq \sum_{i=1}^{\infty} (b_i - a_i) + \varepsilon$$

d'où

$$b - a \leq \sum_{i=1}^{\infty} (b_i - a_i) + 2\varepsilon$$

et comme ε est choisit arbitraire on obtient (1.1). Il reste à prouver que (1.2) implique (1.3).

iii)-b) Avec un petit changement de notations, on va montrer que

$$[c, d] \subset \bigcup_{i=1}^N (c_i, d_i). \quad (1.4)$$

implique

$$d - c \leq \sum_{i=1}^N (d_i - c_i). \quad (1.5)$$

On fait une récurrence : c'est clair pour $N = 1$. Supposons que (1.4) pour $N - 1$ implique (1.5) pour $N - 1$ et on vérifie l'implication pour N . Supposons que

$$c_N = \max_{i=1, \dots, N} c_i \text{ et } c_N < d \leq d_N.$$

On considère deux cas (faire un dessin) :

1. Supposons $c_N \leq c$. Alors $d - c \leq d_N - c_N \leq \sum_{i=1}^N (d_i - c_i)$.
2. Supposons $c < c_N$. Alors par (1.4) on a

$$[c, c_N] \subset \bigcup_{i=1}^{N-1} (c_i, d_i),$$

et par l'hypothèse de récurrence

$$c_N - c \leq \sum_{i=1}^{N-1} (d_i - c_i),$$

donc

$$d - c = d - c_N + c_N - c \leq d - c_N + \sum_{i=1}^{N-1} (d_i - c_i) \leq d_N - c_N + \sum_{i=1}^{N-1} (d_i - c_i) = \sum_{i=1}^N (d_i - c_i).$$

Cela achève la preuve de (1.1).

iii)-c) Pour terminer la preuve de la σ -additivité on va vérifier l'inégalité inverse. On reprend $(a, b] = \bigcup_{i=1}^{\infty} (a_i, b_i]$ avec une union d'intervalles disjoints. Comme pour tout n , $\bigcup_{i=1}^n (a_i, b_i]$ est une union d'intervalles disjoints la même chose est vrai pour

$$(a, b] \setminus \bigcup_{i=1}^n (a_i, b_i] =: \bigcup_{j=1}^m I_j.$$

Comme λ est additive on peut écrire

$$\lambda((a, b]) = \lambda\left(\bigcup_{i=1}^n (a_i, b_i] \cup \bigcup_{j=1}^m I_j\right) = \sum_{i=1}^n \lambda((a_i, b_i]) + \sum_{j=1}^m \lambda(I_j) \geq \sum_{i=1}^n \lambda((a_i, b_i])$$

Lorsque $n \rightarrow \infty$ on trouve

$$\lambda((a, b]) \geq \sum_{i=1}^{\infty} \lambda((a_i, b_i])$$

donc λ est σ -additive sur \mathcal{S} .

iv) Une définition et le resultat clé en cadre général :

Définition 1.11 (*semi-algèbre*)

Soit Ω un ensemble non-vide quelconque. Une famille de sous-ensembles \mathcal{S} de Ω est une **semi-algèbre** si

- $\emptyset, \Omega \in \mathcal{S}$;
- \mathcal{S} est stable par intersections finies;
- si $A \in \mathcal{S}$, alors il existe un n fini et des ensembles C_1, \dots, C_n disjoints dans \mathcal{S} tels que $A^c = C_1 \cup \dots \cup C_n$.

Théorème 1.3 (*d'extension*)

Supposons que \mathcal{S} est une semi-algèbre sur un ensemble non-vide quelconque Ω et que P est une fonction σ -additive d'ensembles définie sur \mathcal{S} , à valeurs dans $[0, 1]$, telle que $P(\Omega) = 1$. Alors il existe une unique mesure de probabilité \mathbb{P} sur $\sigma(\mathcal{S})$ qui prolonge P (c'est-à-dire que $\mathbb{P}|_{\mathcal{S}} = P$).

Ce résultat s'applique pour $\Omega = (0,1]$ avec \mathcal{S} définie en i) et λ définie en ii). La probabilité ainsi obtenue est la **mesure de Lebesgue sur** $(0,1]$.

v) Donnons les idées de preuve du résultat clé. On considère Ω un ensemble non-vide quelconque. Le Théorème 1.3 est une conséquence des trois lemmes suivants. Avant de les énoncer on a besoin d'une définition (rappel) :

Définition 1.12 (*algèbre et algèbre engendrée*)

1. Une famille de sous-ensembles \mathcal{A} de Ω est une **algèbre** si

- $\Omega \in \mathcal{A}$;
- $A \in \mathcal{A}$ implique $A^c \in \mathcal{A}$;
- $A, B \in \mathcal{A}$ implique $A \cup B \in \mathcal{A}$.

2. L'**algèbre engendrée** par une famille de sous-ensembles \mathcal{S} de Ω est la plus petite algèbre contenant \mathcal{S} (on note $\mathcal{A}(\mathcal{S})$).

Lemme 1.1 (*algèbre engendrée par une semi-algèbre*)

Supposons que \mathcal{S} est une semi-algèbre sur Ω . Alors, l'algèbre engendrée par \mathcal{S} est la famille des unions finies d'éléments disjoints de \mathcal{S} :

$$\mathcal{A}(\mathcal{S}) := \left\{ \bigcup_{i \in I} S_i : I \text{ fini, } S_i \in \mathcal{S} \text{ disjoints} \right\}. \quad (1.6)$$

Lemme 1.2 (*première extension*)

Supposons que \mathcal{S} est une semi-algèbre sur Ω et que P est une fonction σ -additive définie sur \mathcal{S} , à valeurs dans $[0,1]$, telle que $P(\Omega) = 1$. Alors il existe une unique extension P' de P sur $\mathcal{A}(\mathcal{S})$, telle que P' est σ -additive sur $\mathcal{A}(\mathcal{S})$ et $P'(\Omega) = 1$. Cette extension est définie par

$$P' \left(\bigcup_{i \in I} S_i \right) = \sum_{i \in I} P(S_i). \quad (1.7)$$

Lemme 1.3 (*deuxième extension*)

Supposons que \mathcal{A} est une algèbre sur Ω et que P' est une fonction σ -additive définie sur \mathcal{A} , à valeurs dans $[0,1]$, telle que $P'(\Omega) = 1$. Alors il existe une unique mesure de probabilité \mathbb{P} sur $\sigma(\mathcal{A})$ (tribu engendrée par \mathcal{A}) qui prolonge P' .

Preuve du Lemme 1.1. On note Λ l'ensemble du membre droit de (1.6). Il est clair que $\Lambda \supset \mathcal{S}$ et on montre que Λ est une algèbre. On vérifie les trois axiomes de la Définition 1.12.

1. Définition 1.11 implique $\Omega \in \mathcal{S}$, donc $\Omega \in \Lambda$.
2. Si $\bigcup_{i \in I} S_i$ et $\bigcup_{j \in J} S'_j$ sont deux éléments de Λ , alors

$$\left(\bigcup_{i \in I} S_i \right) \cap \left(\bigcup_{j \in J} S'_j \right) = \bigcup_{(i,j) \in I \times J} S_i \cap S'_j \in \Lambda$$

puisque $\{S_i \cap S'_j : (i,j) \in I \times J\}$ est une famille finie disjointe d'éléments de \mathcal{S} (qui est stable par intersection finie).

3. Enfin, vérifions la stabilité au passage au complémentaire. Soit $\bigcup_{i \in I} S_i \in \Lambda$ dont le complémentaire est $\bigcap_{i \in I} S_i^c$. Par la troisième axiome d'une semi-algèbre, comme $S_i \in \mathcal{S}$, on a

$$S_i^c = \bigcup_{j \in J_i} S_{ij},$$

avec une famille finie disjointe $\{S_{ij} : j \in J_i\}$ d'éléments de \mathcal{S} . Par le point précédent on conclut que $\bigcap_{i \in I} S_i^c \in \Lambda$.

Cela montre que Λ est une algèbre contenant \mathcal{S} , donc $\Lambda \supset \mathcal{A}(\mathcal{S})$. Par ailleurs, $\bigcup_{i \in I} S_i \in \Lambda$ implique $\bigcup_{i \in I} S_i \in \mathcal{A}(\mathcal{S})$, d'où $\Lambda \subset \mathcal{A}(\mathcal{S})$. \square

Preuve du Lemme 1.2. On commence par vérifier que P' est bien définie par (1.7), ensuite que P' est σ -additive sur $\mathcal{A}(\mathcal{S})$ et enfin que l'extension est unique.

1. Supposons que $A \in \mathcal{A}(\mathcal{S})$ admet deux représentations différentes $A = \bigcup_{i \in I} S_i = \bigcup_{j \in J} S'_j$ et on a besoin de vérifier que $\sum_{i \in I} P(S_i) = \sum_{j \in J} P(S'_j)$ pour que P' ait une unique valeur en A . Comme $S_i \in \mathcal{A}$,

$$\sum_{i \in I} P(S_i) = \sum_{i \in I} P(S_i \cap A) = \sum_{i \in I} P(S_i \cap \bigcup_{j \in J} S'_j) = \sum_{i \in I} P(\bigcup_{j \in J} S_i \cap S'_j)$$

et comme $S_i = \bigcup_{j \in J} S_i \cap S'_j \in \mathcal{S}$ on peut utiliser l'additivité de P . Ainsi

$$\sum_{i \in I} P(S_i) = \sum_{i \in I} \sum_{j \in J} P(S_i \cap S'_j) = \sum_{j \in J} \sum_{i \in I} P(S_i \cap S'_j) = \sum_{j \in J} P(S'_j).$$

La dernière égalité s'obtient en faisant le chemin inverse.

2. Supposons que pour $i \geq 1$ les ensembles disjoints sont $A_i = \bigcup_{j \in J_i} S_{ij} \in \mathcal{A}(\mathcal{S})$, $S_{ij} \in \mathcal{S}$ et $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}(\mathcal{S})$. Par ailleurs, comme $A \in \mathcal{A}(\mathcal{S})$, A admet la représentation $A = \bigcup_{k \in K} S_k$, $S_k \in \mathcal{S}$, $k \in K$ fini. Alors

$$\mathcal{S} \ni S_k = S_k \cap A = \bigcup_{i=1}^{\infty} S_k \cap A_i = \bigcup_{i=1}^{\infty} \bigcup_{j \in J_i} S_k \cap S_{ij} \text{ avec } S_{ij} \in \mathcal{S}.$$

Par (1.7)

$$\begin{aligned} P'(A) &= \sum_{k \in K} P(S_k) = \sum_{k \in K} \sum_{i=1}^{\infty} \sum_{j \in J_i} P(S_k \cap S_{ij}) = \sum_{i=1}^{\infty} \sum_{j \in J_i} \sum_{k \in K} P(S_k \cap S_{ij}) \\ &= \sum_{i=1}^{\infty} \sum_{j \in J_i} P(S_{ij}) = \sum_{i=1}^{\infty} P(\bigcup_{j \in J_i} S_{ij}) = \sum_{i=1}^{\infty} P'(A_i), \end{aligned}$$

puisque $\sum_{k \in K} S_k \cap S_{ij} = A \cap S_{ij} = S_{ij} \in \mathcal{S}$.

3. Soient P'_1 et P'_2 deux extensions additives. Alors, pour tout $A = \bigcup_{i \in I} S_i \in \mathcal{A}(\mathcal{S})$ on a $P'_1(A) = \sum_{i \in I} P(S_i) = P'_2(A)$.

On ainsi construit l'extension de P à une algèbre. \square

Preuve du Lemme 1.3. On divise la preuve en trois parties : en 1ère partie on étend P' à une fonction d'ensemble Π σ -additive sur une famille $\mathcal{G} \supset \mathcal{A}$. En 2ème partie on étend Π à une fonction d'ensemble Π^* sur $\mathcal{P}(\Omega) \supset \sigma(\mathcal{A})$ et en 3ème partie on fait la restriction de Π^* à $\sigma(\mathcal{A})$ et on obtient la probabilité recherchée.

1. On définit d'abord la famille \mathcal{G} :

$$\mathcal{G} = \left\{ \bigcup_{j=1}^{\infty} A_j : A_j \in \mathcal{A} \right\} = \left\{ \lim_n \uparrow B_n : B_n \in \mathcal{A}, B_n \subset B_{n+1}, \forall n \right\}$$

et ensuite la fonction $\Pi : \mathcal{G} \rightarrow [0,1]$, par : si $G = \lim_n \uparrow B_n \in \mathcal{G}$, où $B_n \in \mathcal{A}$, alors

$$\Pi(G) = \lim_{n \rightarrow \infty} P'(B_n). \quad (1.8)$$

Cette dernière définition est bien justifiée car P' est σ -additive donc la propriété de continuité sur des suites croissantes est vraie. On dit que $\{B_n\}$ est une **suite approchante** de G . Il reste à voir que Π est bien définie, c'est-à-dire

Fait 1 Si G admet deux suites approchantes $\{B_n\}$ et $\{B'_n\}$,

$$G = \lim_n \uparrow B_n = \lim_n \uparrow B'_n \text{ alors } \lim_{n \rightarrow \infty} P'(B_n) = \lim_{n \rightarrow \infty} P'(B'_n). \quad (1.9)$$

Listons quelques propriétés de Π et \mathcal{G} :

Fait 2 On a

$$\Omega, \emptyset \in \mathcal{G} \text{ et } \Pi(\omega) = 1, \Pi(\emptyset) = 0,$$

et pour $G \in \mathcal{G}$

$$0 \leq \Pi(G) \leq 1. \quad (1.10)$$

Enfin, on a $\mathcal{A} \subset \mathcal{G}$ et $\Pi|_{\mathcal{A}} = P'$.

Fait 3 Si $G_i \in \mathcal{G}$ pour $i = 1, 2$, alors $G_1 \cup G_2 \in \mathcal{G}$, $G_1 \cap G_2 \in \mathcal{G}$ et Π est additive :

$$\Pi(G_1 \cup G_2) + \Pi(G_1 \cap G_2) = \Pi(G_1) + \Pi(G_2). \quad (1.11)$$

Fait 4 Π est monotone sur \mathcal{G} : si $G_i \in \mathcal{G}$ pour $i = 1, 2$ et $G_1 \subset G_2$, alors $\Pi(G_1) \leq \Pi(G_2)$.

Fait 5 \mathcal{G} est stable par des limites des suites croissantes et Π est continue sur des suites croissantes : si $G_n \in \mathcal{G}$ et $G_n \uparrow G$, alors $G \in \mathcal{G}$ et $\Pi(G) = \lim_{n \rightarrow \infty} \Pi(G_n)$.

Par les Faits 3 et 5 on déduit que Π est σ -additive sur \mathcal{G} donc la 1ère partie est vérifiée.

2. On définit $\Pi^* : \mathcal{P}(\Omega) \rightarrow [0,1]$ par

$$\forall A \in \mathcal{P}(\Omega) : \Pi^*(A) = \inf\{\Pi(G) : A \subset G \in \mathcal{G}\}. \quad (1.12)$$

$\Pi^*(A)$ est le plus petit majorant des valeurs $\Pi(G)$ sur des ensembles $G \in \mathcal{G}$ contenant A . C'est la **mesure extérieure** de A . Comme pour Π , on va lister les propriétés de Π^* :

Fait 6 On a

$$\Pi^*|_{\mathcal{G}} = \Pi \quad (1.13)$$

et $0 \leq \Pi^*(A) \leq 1$, pour tout $A \in \mathcal{P}(\Omega)$. En particulier, $\Pi^*(\Omega) = \Pi(\Omega) = 1$ et $\Pi^*(\emptyset) = \Pi(\emptyset) = 0$.

Fait 7 Π^* est sous-additive : on a pour $A_1, A_2 \in \mathcal{P}(\Omega)$

$$\Pi^*(A_1 \cup A_2) + \Pi^*(A_1 \cap A_2) \leq \Pi^*(A_1) + \Pi^*(A_2). \quad (1.14)$$

En particulier

$$1 = \Pi^*(\Omega) \leq \Pi^*(A) + \Pi^*(A^c). \quad (1.15)$$

Fait 8 Π^* est monotone sur $\mathcal{P}(\Omega)$.

Fait 9 Π^* est continue sur des suites croissantes : si $A_n \uparrow A$, alors $\Pi^*(A_n) \uparrow \Pi^*(A)$.

3. On introduit une sous-famille \mathcal{D} de $\mathcal{P}(\Omega)$:

$$\mathcal{D} := \{D \in \mathcal{P}(\Omega) : \Pi^*(D) + \Pi^*(D^c) = 1\}. \quad (1.16)$$

Fait 10 \mathcal{D} est une tribu et $\Pi^*_{|\mathcal{D}}$ est une probabilité sur (Ω, \mathcal{D}) .

Fait 11 $\mathcal{D} \supset \mathcal{A}$, donc $\mathcal{D} \supset \sigma(\mathcal{A})$, et alors $\Pi^*_{|\sigma(\mathcal{A})}$ est la probabilité (unique par un argument de classe monotone) désirée.

Preuve du Fait 1 : Il suffit de montrer que :

$$\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} B'_n \text{ implique } \lim_{n \rightarrow \infty} P'(B_n) \leq \lim_{n \rightarrow \infty} P'(B'_n). \quad (1.17)$$

Pour m fixé $\lim_n \uparrow (B_m \cap B'_n) = B_m$ et on a aussi $B_m \cap B'_n \subset B'_n$. On sait que la σ -additivité de P implique la continuité sur des suites croissantes. On en déduit :

$$\lim_{n \rightarrow \infty} P'(B'_n) \geq \lim_{n \rightarrow \infty} P'(B_m \cap B'_n) = P'(B_m).$$

Comme cette inégalité a lieu pour tout m , on en déduit que

$$\lim_{n \rightarrow \infty} P'(B'_n) \geq \lim_{m \rightarrow \infty} P'(B_m).$$

□

Preuve du Fait 2 : Si on pose $B_n = \Omega$ pour tout n , alors

$$\mathcal{A} \ni B_n = \Omega \uparrow \Omega \text{ et } \Pi(\Omega) = \lim_{n \rightarrow \infty} P'(\Omega) = 1.$$

Le même argument marche pour \emptyset . (1.10) s'obtient par le fait que $0 \leq P'(B_n) \leq 1$, pour toute suite approchante $\{B_n\}$ de \mathcal{A} . Enfin, pour montrer que $\Pi(A) = P'(A)$ pour tout $A \in \mathcal{A}$, on prend la suite approchante identiquement égale à A . □

Preuve du Fait 3 : Soient les suites approchantes $B_{n1}, B_{n2} \in \mathcal{A}$, telles que $B_{ni} \uparrow G_i$ pour $i = 1, 2$. Comme \mathcal{A} est une algèbre, on voit que

$$\mathcal{A} \ni B_{n1} \cup B_{n2} \uparrow G_1 \cup G_2, \mathcal{A} \ni B_{n1} \cap B_{n2} \uparrow G_1 \cap G_2,$$

qui montrent que $G_1 \cup G_2, G_1 \cap G_2 \in \mathcal{G}$. De plus

$$P'(B_{n1} \cup B_{n2}) + P'(B_{n1} \cap B_{n2}) = P'(B_{n1}) + P'(B_{n2}),$$

et pour $n \rightarrow \infty$ on trouve (1.11). □

Preuve du Fait 4 : C'est une conséquence directe de (1.17). □

Preuve du Fait 5 : Pour chaque n , G_n admet une suite approchante $B_{m,n} \in \mathcal{A}$ telle que $\lim_m \uparrow B_{m,n} = G_n$. On définit $D_m = \cup_{n=1}^m B_{m,n} \in \mathcal{A}$ (car \mathcal{A} est stable par union finie). On va montrer que

$$\lim_m \uparrow D_m = G \quad (1.18)$$

et alors G admet une suite approchante croissante d'éléments de \mathcal{A} , donc $G \in \mathcal{G}$. Montrons d'abord que $\{D_m\}$ est croissante :

$$D_m = \bigcup_{n=1}^m B_{m,n} \subset \bigcup_{n=1}^m B_{m+1,n} \subset \bigcup_{n=1}^{m+1} B_{m+1,n} = D_{m+1}$$

Calculons la limite de $\{D_m\}$. Si $n \geq m$, on a par la définition de D_m :

$$B_{m,n} \subset D_m = \bigcup_{j=1}^m B_{m,j} \subset \bigcup_{j=1}^m G_j = G_m.$$

donc $B_{m,n} \subset D_m \subset G_m$. On prend la limite en m :

$$G_n = \lim_m \uparrow B_{m,n} \subset \lim_m \uparrow D_m \subset \lim_m \uparrow G_m = G$$

et ensuite la limite en n :

$$G = \lim_n \uparrow G_n \subset \lim_m \uparrow D_m \subset \lim_m \uparrow G_m = G.$$

Donc $G \in \mathcal{G}$ et par la définition de Π , on sait que $\Pi(G) = \lim_{m \rightarrow \infty} \Pi(D_m)$. Il reste à prouver que $\Pi(G_n) \uparrow \Pi(G)$. Par les trois inclusions précédentes :

$$\Pi(B_{m,n}) \leq \Pi(D_m) \leq \Pi(G_m).$$

On fait $m \rightarrow \infty$ et comme $G_n = \lim_m \uparrow B_{m,n}$,

$$\Pi(G_n) \leq \lim_{m \rightarrow \infty} \Pi(D_m) \leq \lim_{m \rightarrow \infty} \Pi(G_m), \forall n.$$

On fait $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \Pi(G_n) \leq \lim_{m \rightarrow \infty} \Pi(D_m) \leq \lim_{m \rightarrow \infty} \Pi(G_m)$$

donc $\lim_{n \rightarrow \infty} \Pi(G_n) = \lim_{m \rightarrow \infty} \Pi(D_m) = \Pi(G)$. □

Preuve du Fait 6 : Il est clair que si $A \in \mathcal{G}$, alors $A \in \{G : A \subset G \in \mathcal{G}\}$ et donc l'infimum est atteint en A . □

Preuve du Fait 7 : Pour vérifier (1.14), on fixe $\varepsilon > 0$ et on trouve $G_i \in \mathcal{G}$ tels que $G_i \supset A_i$ et pour $i = 1, 2$,

$$\Pi^*(A_i) + \frac{\varepsilon}{2} \geq \Pi(G_i).$$

On somme ces deux inégalités et on trouve

$$\Pi^*(A_1) + \Pi^*(A_2) + \varepsilon \geq \Pi(G_1) + \Pi(G_2) = \Pi(G_1 \cup G_2) + \Pi(G_1 \cap G_2).$$

par le Fait 3 pour Π . Comme $G_1 \cup G_2 \supset A_1 \cup A_2$ et $G_1 \cap G_2 \supset A_1 \cap A_2$, par la définition de Π^* qu'on peut encore minorer par

$$\Pi^*(A_1 \cup A_2) + \Pi^*(A_1 \cap A_2).$$

□

Preuve du Fait 8 : Cette propriété est une conséquence du fait que Π est monotone sur \mathcal{G} (Fait 4). \square

Preuve du Fait 9 : On fixe $\varepsilon > 0$. Pour chaque $n \geq 1$ on trouve $G_n \in \mathcal{G}$ tels que $G_n \supset A_n$ et

$$\Pi^*(A_n) + \frac{\varepsilon}{2^n} \geq \Pi(G_n).$$

On pose $G'_n = \cup_{m=1}^n G_m$. Comme \mathcal{G} est stable par réunion finie, $G'_n \in \mathcal{G}$ et $\{G'_n\}$ est croissante. On montre par récurrence

$$\Pi^*(A_n) + \varepsilon \sum_{m=1}^n 2^{-m} \geq \Pi(G'_n). \quad (1.19)$$

Pour $n = 1$ c'est le choix de G_n . On montre que " $n \Rightarrow n + 1$ ". On a

$$A_n \subset G_n \subset G'_n \text{ et } A_{n+1} \subset G_{n+1} \subset G'_{n+1}$$

et donc $A_n \subset G'_n$ et $A_{n+1} \subset G'_{n+1}$, donc $A_n \subset G'_n \cap G_{n+1} \in \mathcal{G}$. Ainsi

$$\Pi(G'_{n+1}) = \Pi(G'_n \cup G_{n+1}) = \Pi(G'_n) + \Pi(G_{n+1}) - \Pi(G'_n \cap G_{n+1})$$

par (1.11). On peut alors majorer le membre de droite de l'égalité précédente par

$$\leq \left(\Pi^*(A_n) + \varepsilon \sum_{m=1}^n 2^{-m} \right) + \Pi^*(A_{n+1}) + \frac{\varepsilon}{2^{n+1}} - \Pi^*(A_n) = \varepsilon \sum_{m=1}^{n+1} 2^{-m} + \Pi^*(A_{n+1})$$

qui est (1.19) pour $n + 1$. On fait $n \rightarrow \infty$ dans (1.19). Par la monotonie de Π sur \mathcal{G} et celle de Π^* sur $\mathcal{P}(\Omega)$, et comme \mathcal{G} est stable par des unions croissantes, on obtient

$$\lim_{n \rightarrow \infty} \Pi^*(A_n) + \varepsilon \geq \lim_{n \rightarrow \infty} \Pi(G'_n) = \Pi\left(\bigcup_{j=1}^{\infty} G'_j\right).$$

Comme $A = \lim_n \uparrow A_n \subset \bigcup_{j=1}^{\infty} G'_j \in \mathcal{G}$, on déduit que $\lim_{n \rightarrow \infty} \Pi^*(A_n) \geq \Pi^*(A)$. Par ailleurs, la monotonie donne $\Pi^*(A_n) \leq \Pi^*(A)$, d'où $\lim_{n \rightarrow \infty} \Pi^*(A_n) \leq \Pi^*(A)$. \square

Preuve du Fait 10 : D'abord on prouve que \mathcal{D} est une algèbre. Il est clair que $\Omega \in \mathcal{D}$, puisque $\Pi^*(\Omega) = 1$ et $\Pi^*(\emptyset) = 0$. Le passage au complémentaire est évident donc il reste à vérifier la stabilité aux unions et intersections finies. Si $A_1, A_2 \in \mathcal{D}$, alors par (1.14) on trouve :

$$\Pi^*(D_1 \cup D_2) + \Pi^*(D_1 \cap D_2) \leq \Pi^*(D_1) + \Pi^*(D_2) \quad (1.20)$$

$$\Pi^*((D_1 \cup D_2)^c) + \Pi^*((D_1 \cap D_2)^c) \leq \Pi^*(D_1^c) + \Pi^*(D_2^c). \quad (1.21)$$

On additionne (1.20) et (1.21) pour obtenir

$$\Pi^*(D_1 \cup D_2) + \Pi^*((D_1 \cup D_2)^c) + \Pi^*(D_1 \cap D_2) + \Pi^*((D_1 \cap D_2)^c) \leq 2 \quad (1.22)$$

où le membre de droite est obtenu parce que $D_1, D_2 \in \mathcal{D}$. Par (1.15) le membre de gauche est ≥ 2 donc en (1.22) on a égalité. En combinant cette égalité avec (1.15) on trouve

$$\Pi^*(D_1 \cup D_2) + \Pi^*((D_1 \cup D_2)^c) = 1$$

$$\Pi^*(D_1 \cap D_2) + \Pi^*((D_1 \cap D_2)^c) = 1,$$

donc $D_1 \cup D_2, D_1 \cap D_2 \in \mathcal{D}$ et \mathcal{D} est une algèbre. De plus on obtient des égalités en (1.20) et (1.21) (sinon on contredit (1.22)), donc Π^* est additive sur \mathcal{D} .

Pour montrer que \mathcal{D} est une tribu il suffit de vérifier que \mathcal{D} est une classe monotone (et utiliser ensuite le théorème de classe monotone). Comme \mathcal{D} est stable par passage au complémentaire il suffit de montrer que $D_n \in \mathcal{D}, D_n \uparrow D$, implique $D \in \mathcal{D}$. Par le Fait 9

$$\lim_{n \rightarrow \infty} \Pi^*(D_n) = \Pi^*\left(\bigcup_{n=1}^{\infty} D_n\right) = \Pi^*(D).$$

Par ailleurs, pour tout $m \geq 1$,

$$\Pi^*\left(\left(\bigcup_{n=1}^{\infty} D_n\right)^c\right) = \Pi^*\left(\bigcap_{n=1}^{\infty} D_n^c\right) \leq \Pi^*(D_m^c),$$

d'où, par (1.15),

$$1 \leq \Pi^*\left(\bigcup_{n=1}^{\infty} D_n\right) + \Pi^*\left(\left(\bigcup_{n=1}^{\infty} D_n\right)^c\right) \leq \lim_{n \rightarrow \infty} \Pi^*(D_n) + \Pi^*(D_m^c). \quad (1.23)$$

Si on fait $m \rightarrow \infty$, comme $D_n \in \mathcal{D}$,

$$1 \leq \lim_{n \rightarrow \infty} \Pi^*(D_n) + \lim_{m \rightarrow \infty} \Pi^*(D_m^c) = \lim_{n \rightarrow \infty} (\Pi^*(D_n) + \Pi^*(D_n^c)) = 1,$$

donc (1.23) est une égalité et donc $D = \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$. Ainsi \mathcal{D} est une algèbre et une classe monotone, donc une tribu.

Montrons que $\Pi^*_{|\mathcal{D}}$ est σ -additive. Soit D_n une suite disjointe dans \mathcal{D} . Comme \mathcal{D} est une algèbre et par le Fait 9 et l'additivité de Π^* on a

$$\Pi^*\left(\bigcup_{n=1}^{\infty} D_n\right) = \Pi^*\left(\lim_n \bigcup_{i=1}^n D_i\right) = \lim_{n \rightarrow \infty} \Pi^*\left(\bigcup_{i=1}^n D_i\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \Pi^*(D_i) = \sum_{i=1}^{\infty} \Pi^*(D_i).$$

□

Preuve du Fait 11. Tout élément $A \in \mathcal{A}$ est élément de \mathcal{G} (suite approchante constante) et alors $\Pi^*(A) = \Pi(A) = P'(A)$ et la même chose pour A^c . Mais alors, par (1.15) $1 \leq \Pi^*(A) + \Pi^*(A^c) = P'(A) + P'(A^c) = 1$, d'où $A \in \mathcal{D}$. Ainsi, $\mathcal{D} \supset \mathcal{A}$, donc la tribu \mathcal{D} contient $\sigma(\mathcal{A})$. La restriction $\Pi^*_{|\sigma(\mathcal{A})}$ est la probabilité désirée. L'unicité de cette extension de \mathcal{A} à $\sigma(\mathcal{A})$ s'obtient par un argument de classe monotone. □

1.6 Loi d'une variable aléatoire

Définition 1.13 Soient (Ω, \mathcal{A}, P) un espace de probabilité et X une variable aléatoire Ω dans un espace mesurable (E, \mathcal{B}) . L'application $P_X : \mathcal{B} \rightarrow [0, 1]$ définie par $P_X(B) := P(X^{-1}(B))$, définit une probabilité sur (E, \mathcal{B}) appelée **loi** de X .

Soit une autre variable aléatoire $Y : (\Omega', \mathcal{A}', P') \rightarrow (E, \mathcal{B})$. On dit que X et Y ont la même loi lorsque $P_X = P_Y \Leftrightarrow P_X(B) = P_Y(B), \forall B \in \mathcal{B}$.

REMARQUE: On peut alléger les notations:

$$P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X \in B).$$

En ce qui concerne la deuxième partie, remarquer que les deux variables aléatoires ne sont pas nécessairement définies sur le même espace de probabilité. \square

Exemple: Sur $\Omega = \{1,2,3,4,5,6\}$ muni de la tribu de ses parties et de la probabilité $P(A) = \text{card}(A)/6$ on définit X à valeurs dans $\{0,1\}$, par $X(\omega) = 1$, si ω est pair et $X(\omega) = 0$ si ω est impair. On vérifie que $P_X(\{0\}) = P_X(\{1\}) = 1/2$.

REMARQUE: Si l'on se donne une probabilité sur (E, \mathcal{B}) (une loi) on peut toujours l'écrire comme la loi d'une variable aléatoire (prendre l'identité pour la variable aléatoire). Pour les applications, en général, seule compte la loi et on explicite plus rarement la variable aléatoire et l'espace de probabilité (Ω, \mathcal{A}, P) . Par exemple, on dit que X est de loi de Bernoulli si $P(X = 1) = 1 - P(X = 0) = p$ au lieu de dire que $X : (\Omega, \mathcal{A}, P) \rightarrow \{0,1\}$ avec $P_X(\{1\}) = 1 - P_X(\{0\}) = p$. Aussi la représentation d'une loi par une variable aléatoire n'est pas unique. En reprenant l'exemple de la loi de Bernoulli, on peut choisir

$$\Omega = \{0,1\}, \mathcal{A} = \mathcal{P}(\Omega), P = p\delta_1 + (1-p)\delta_0, X(\omega) = \omega$$

ou

$$\Omega' = [0,1], \mathcal{A}' = \mathcal{B}([0,1]), P' = \lambda, Y(\omega) = \mathbf{1}_{[0,p]}(\omega).$$

\square

Définition 1.14 On dit qu'une loi Q sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est **discrète** si c'est une combinaison linéaire finie ou dénombrable à coefficients positifs ou nuls de masses de Dirac

$$Q = \sum_{j \in J} p_j \delta_{x_j}, x_j \in \mathbb{R}^d, j \in J \subset \mathbb{N}.$$

Une variable aléatoire X est **discrète** si sa loi est $P_X = \sum_{j \in J} p_j \delta_{x_j}$. X ne prend (presque sûrement) qu'un nombre fini ou dénombrable de valeurs.

X est discrète si et seulement si $X = \sum_{j \in J} x_j \mathbf{1}_{A_j}$ avec $A_j = X^{-1}(\{x_j\}) = \{\omega \in \Omega : X(\omega) = x_j\}$ disjoints deux à deux et dont la réunion est Ω .

On dit qu'une loi Q sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ **admet une densité** s'il existe une fonction f borélienne (presque sûrement) positive ou nulle, telle que

$$Q(B) = \int_{\mathbb{R}^d} f(x) \mathbf{1}_B(x) dx, B \in \mathcal{B}(\mathbb{R}^d).$$

On notera la densité de la loi d'une variable aléatoire X , par f_X et on dira qu'elle est la **densité** de X :

$$P_X(B) = P(X \in B) = \int_{\mathbb{R}^d} f_X(x) \mathbf{1}_B(x) dx, B \in \mathcal{B}(\mathbb{R}^d).$$

REMARQUE: (cas discret) On peut écrire que $p_j = P(X = x_j) = p_X(j)$ pour tous $j \in J$. Les coefficients p_j sont positifs ou nuls et $\sum_{j \in J} p_j = 1$. \square

REMARQUE: (cas à densité) Si f est la densité d'une loi alors elle est presque sûrement unique (c'est-à-dire, si g est aussi une densité de la même loi alors $f = g$ p.p.). De plus

$$\int_{\mathbb{R}^d} f(x) dx = 1.$$

□

REMARQUE: On peut montrer que si f est une fonction borélienne positive p.p. sur \mathbb{R}^d et telle que son intégrale sur tout l'espace est égale à 1, alors il existe un espace de probabilité (Ω, \mathcal{A}, P) et une variable aléatoire X dont la loi admet f pour densité.

Une condition suffisante (et nécessaire) pour que la loi de X admette une densité est la condition d'absolue continuité de la probabilité P_X par rapport à la mesure de Lebesgue. □

REMARQUE: Soit X une variable aléatoire dont la loi est discrète donnée par $p_X(j) = P(X = x_j)$, $j \in J$ et soit $g : J \rightarrow I$ une application mesurable de l'ensemble dénombrable J dans l'ensemble dénombrable I . Alors $Y = g(X)$ est une variable aléatoire dont la loi est discrète donnée par

$$p_Y(i) = P(Y = y_i) = \sum_{j \in J: g(x_j) = y_i} p_X(j).$$

On verra que ce résultat ne se généralise pas tel quel à des fonctions de plusieurs variables dans le cas discret. □

Proposition 1.9 Soit X une variable aléatoire à valeurs dans \mathbb{R}^d de densité f_X et soit $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ un difféomorphisme de \mathbb{R}^d , c'est-à-dire, une bijection continûment différentiable ainsi que son inverse. Alors $Y = g(X)$ est une variable aléatoire à valeurs dans \mathbb{R}^d qui admet pour densité la fonction

$$f_Y(y) = f_X(g^{-1}(y)) |\text{Jac } g^{-1}(y)|, y \in \mathbb{R}^d,$$

où $\text{Jac } g(x) = \det \left(\frac{\partial g_i}{\partial x_j} : 1 \leq i, j \leq d \right)$ désigne le jacobien de g .

Preuve: Soit $B \in \mathcal{B}(\mathbb{R}^d)$ et on calcule

$$P(g(X) \in B) = P(X \in g^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{\{x \in g^{-1}(B)\}} f_X(x) dx = \int_{\mathbb{R}^d} \mathbb{1}_B(g(x)) f_X(x) dx.$$

On pose $y = g(x) \Leftrightarrow x = g^{-1}(y)$ et on applique le théorème de changement de variable

$$P(g(X) \in B) = \int_{\mathbb{R}^d} \mathbb{1}_B(y) f_X(g^{-1}(y)) |\text{Jac } g^{-1}(y)| dy.$$

□

Définition 1.15 Soit X une variable aléatoire réelle définie sur un espace de probabilité (Ω, \mathcal{A}, P) . 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

$$F_X(t) = P_X([-\infty, t]) = P(\{\omega \in \Omega : X(\omega) \leq t\}) = P(X \leq t), t \in \mathbb{R}.$$

Proposition 1.10 *La fonction de répartition F_X de la variable aléatoire réelle X vérifie les propriétés suivantes:*

- 1) $0 \leq F_X \leq 1$;
- 2) F_X est croissante, continue à droite avec une limite à gauche en tout point;
- 3) $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow \infty} F_X(t) = 1$.

Réciproquement, une fonction F vérifiant 1)-3) est la fonction de répartition d'une variable aléatoire réelle.

Preuve: 1) vient du fait que P est à valeurs dans $[0,1]$. La croissance découle de la croissance de P .

La continuité à droite peut être vue comme une conséquence de la Proposition 8, 4) en remarquant que

$$\{X \leq t\} = \bigcap_{n \in \mathbb{N}^*} \{X \leq t + \frac{1}{n}\}$$

et que la croissance de F_X implique

$$\lim_{h \downarrow 0} F_X(t+h) = \lim_{n \uparrow \infty} F_X(t + \frac{1}{n}) = F_X(t).$$

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

La propriété 3) vient encore de la Proposition 8, 4) en remarquant que $\emptyset = \bigcap_{n \in \mathbb{N}^*} \{X \leq -n\}$ et donc

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

tandis que

$$1 = P(\Omega) = \lim_{n \uparrow \infty} P(X \leq n),$$

d'après la Proposition 8, 3).

Soit maintenant F une fonction vérifiant 1)-3). On définit, pour $a < b$ la fonction d'ensemble $\mu(]a,b]) = F(b) - F(a)$. La définition de μ s'étend à l'ensemble \mathcal{C} des réunions finies d'intervalles. On peut aussi montrer que pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{C} , de réunion A , on ait $\lim_{n \uparrow \infty} \mu(A_n) = \mu(A)$. Alors, par un théorème de prolongement (admis), μ se prolonge en une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. \square

Théorème 1.4 *La fonction de répartition caractérise la loi, c'est-à-dire, $F_X = F_Y$ si et seulement si $P_X = P_Y$.*

Preuve: En effet, si $F_X = F_Y$, alors P_X et P_Y coïncident sur les intervalles, donc sur la tribu engendrée par l'ensemble \mathcal{E} des intervalles; cette tribu est la tribu borélienne. De plus \mathcal{E} est stable par intersection finie. Donc, par le théorème de classe monotone $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$. Soit

$$\mathcal{M}_1 = \{B \in \mathcal{B}(\mathbb{R}) : P_X(B) = P_Y(B)\}.$$

Par la propriété de continuité monotone de P on prouve que \mathcal{M}_1 est une classe monotone contenant \mathcal{E} . Le résultat s'ensuit. \square

Proposition 1.11 *La fonction de répartition admet au plus un nombre dénombrable de points de discontinuité.*

Preuve: Soit D_n l'ensemble des points de discontinuité avec un saut d'amplitude plus grande que $1/n$:

$$D_n := \left\{ t \in \mathbb{R} : F(t) - F(t-) \geq \frac{1}{n} \right\}.$$

Comme $0 \leq F \leq 1$ on a nécessairement $\text{card}(D_n) \leq n$. L'ensemble des points de discontinuité est $\cup_{n \in \mathbb{N}^*} D_n$, et donc est dénombrable. \square

REMARQUE: Supposons que la loi de X est discrète et que X ne prend qu'un nombre fini de valeurs $X(\Omega) = \{x_1, \dots, x_r\}$, avec $x_1 < \dots < x_r$. Alors

$$F_X(t) = \begin{cases} 0, & \text{si } t < x_1 \\ p_1 + \dots + p_j, & \text{si } x_j \leq t < x_{j+1} \text{ avec } j < n, \\ 1, & \text{si } t \geq x_r \end{cases} \quad t \in \mathbb{R},$$

où $p_j = P(X = x_j)$, $j = 1, \dots, r$; on peut voir que $p_j = F_X(x_j) - F_X(x_{j-})$.

REMARQUE: Supposons que la loi de X admet une densité f_X . Alors

$$F_X(t) = \int_{-\infty}^t f_X(x) dx, t \in \mathbb{R}.$$

On peut voir que F_X est continue: $F_X(t-) = F_X(t)$ et en particulier

$$P(X = t) = 0, \forall t \in \mathbb{R}.$$

De plus F_X est dérivable p.p. sur \mathbb{R} . Ainsi il y a une bijection entre l'ensemble des fonctions boréliennes positives p.p. sur \mathbb{R} et d'intégrale 1 et l'ensemble des fonctions de répartition continues sur \mathbb{R} et dérivables p.p. sur \mathbb{R} . Si de plus f_X est continue, alors F_X est dérivable. \square

Fonctions monotones. Dérivabilité de l'intégrale par rapport à la borne supérieure. (facultatif).

Rappel: si f est une fonction continue et si F est une fonction continuellement différentiable on a

$$\frac{d}{dx} \int_a^x f(t) dt = f(x) \quad \text{et} \quad \int_a^b F'(t) dt = F(b) - F(a).$$

Questions: La première égalité est-elle vraie pour des fonctions intégrables au sens de Lebesgue? Quelle est la classe (aussi vaste que possible) des fonctions pour lesquelles la deuxième égalité a lieu?

Le but est d'étudier les propriétés de l'intégrale de Lebesgue $F(x) := \int_a^x f(t) dt$, comme fonction de sa borne supérieure. Si $f \geq 0$ alors F est monotone non décroissante. On sait

que toute fonction intégrable f est différence de deux fonctions intégrables non négatives $f = f^+ - f^-$ donc F se décompose en une différence de deux fonctions monotones non décroissantes. Ainsi l'étude de F peut être réduite à l'étude des fonctions monotones du même type. Les propriétés **5** et **7** ci-dessous, donnent la réponse à la première question. La réponse à la deuxième est contenue dans les propriétés **8**, **11** et **12**.

Définition 1.16 Une fonction g sur un intervalle $[a, b]$ est dite monotone non décroissante si $t \leq t'$ implique $g(t) \leq g(t')$. La limite $\lim_{h \downarrow 0} g(t_0 + h)$ (lorsqu'elle existe) s'appelle limite à droite de g au point t_0 et se note $g(t_0+)$. De la même façon on définit la limite à gauche de g en t_0 notée $g(t_0-)$. Le point où ces deux limites existent, mais sont inégales, s'appelle point de discontinuité de première espèce. La différence $g(t_0+) - g(t_0-)$ s'appelle saut de g en t_0 . Si $g(t_0) = g(t_0+)$ on dit que la fonction est continue à droite en t_0 .

Propriété 1 Une fonction monotone non décroissante sur un intervalle $[a, b]$ est borélienne et bornée, donc intégrable.

En effet si g est monotone non décroissante sur $[a, b]$, on a $g(a) \leq g(t) \leq g(b)$ sur $[a, b]$. D'autre part $\{t : g(t) < c\}$ est soit un segment, soit un intervalle semi-ouvert (soit l'ensemble vide). En effet, supposons qu'il existe des points t tels que $g(t) < c$. On note α la borne supérieure de ces points. Alors $\{t : g(t) < c\}$ est soit $[a, \alpha]$, soit $[a, \alpha)$.

Propriété 2 Une fonction monotone ne peut avoir que des discontinuités de première espèce.

En effet, soit t_0 un point quelconque de $[a, b]$ et soit $t_n \uparrow t_0$. La suite $\{g(t_n)\}$ est bornée par $g(a)$ et $g(b)$, donc elle a au moins un point d'accumulation. L'existence de plusieurs points d'accumulation pour une telle suite est en contradiction avec la monotonie de g . Ainsi $g(t_0-)$ existe et on fait de la même façon pour $g(t_0+)$.

Propriété 3 L'ensemble des points de discontinuité d'une fonction monotone est au plus dénombrable.

En effet, la somme de tout nombre fini de sauts de g sur $[a, b]$ ne dépasse pas $g(b) - g(a)$. Par conséquent, pour chaque n , le nombre de sauts plus grand que $\frac{1}{n}$ est fini. En faisant leur somme pour tous les $n = 1, 2, \dots$, on trouve que le nombre total de sauts est fini ou dénombrable.

Propriété 4 Toute fonction monotone continue à droite peut être représentée de façon unique comme la somme d'une fonction continue monotone et d'une fonction de sauts (continue à droite).

En effet, soient t_1, t_2, \dots les points de discontinuité de la fonction non décroissante g et soient p_1, p_2, \dots ses sauts en ces points. On pose $p(t) = \sum_{\{n: t_n \leq t\}} p_n$. Soit $\gamma(t) = g(t) - p(t)$. Cette fonction est non décroissante car, pour $t' < t''$

$$\gamma(t'') - \gamma(t') = [g(t'') - g(t')] - [p(t'') - p(t')] \geq 0$$

comme somme de l'accroissement de g sur $[t', t'']$ et la somme de ses sauts sur le même intervalle. Pour t^* un point arbitraire on a (voir!)

$$\gamma(t^*+) - \gamma(t^*-) = g(t^*+) - g(t^*-) - p^* = 0$$

(où p^* est le saut de p au point t^*), par la continuité à droite de g et de p .

Propriété 5 Une fonction monotone sur un intervalle $[a,b]$ admet presque partout sur cet intervalle une dérivée finie.

On pose $R(t) := \frac{g(t)-g(t_0)}{t-t_0}$ et on introduit les notations suivantes :

$$\Lambda_{dr}(t_0) := \limsup_{t \downarrow t_0} R(t), \lambda_{dr}(t_0) := \liminf_{t \downarrow t_0} R(t), \Lambda_{ga}(t_0) := \limsup_{t \uparrow t_0} R(t), \lambda_{ga}(t_0) := \liminf_{t \uparrow t_0} R(t).$$

On a toujours $\lambda_{dr} \leq \Lambda_{dr}$ et $\lambda_{ga} \leq \Lambda_{ga}$. (Pour simplifier on ne fait pas apparaître la dépendance en t_0 .)

Il faut prouver que $-\infty < \lambda_{ga} = \lambda_{dr} = \Lambda_{ga} = \Lambda_{dr} < \infty$ a lieu pour presque tous $t_0 \in [a,b]$.

On commence par prouver le résultat pour g continue non décroissante. Il suffit de prouver que, presque partout, on a

$$\Lambda_{dr} < \infty \text{ et } \lambda_{ga} \geq \Lambda_{dr}.$$

En effet, si l'on pose $\hat{g}(t) = -g(-t)$, g sera continue non décroissante sur $[-b, -a]$. Alors, on peut voir que pour tout $t_0 \in (a,b)$, on a

$$\Lambda_{ga}(t_0) = \hat{\Lambda}_{dr}(-t_0), \lambda_{dr}(t_0) = \hat{\lambda}_{ga}(-t_0).$$

Donc en appliquant la deuxième inégalité à \hat{g} on obtient $\lambda_{dr} \geq \Lambda_{ga}$, d'où, presque partout, $\Lambda_{dr} \leq \lambda_{ga} \leq \Lambda_{ga} \leq \lambda_{dr} \leq \Lambda_{dr}$.

Montrons d'abord que $\Lambda_{dr} < \infty$ presque partout. Supposons que $\Lambda_{dr} = \infty$ en un point t_0 . Alors, pour toute constante C il existe $\tau > t_0$ tel que $R(\tau) > C$ ou

$$g(\tau) - g(t_0) > C(\tau - t_0) \Leftrightarrow g(\tau) - C\tau > g(t_0) - Ct_0.$$

Soit la fonction continue $h(t) = g(t) - Ct$. On voit que t_0 est un point invisible à droite pour h (voir le lemme suivant admis) :

Lemme 1 (Riesz)

Soit h une fonction continue sur $[a,b]$. Un point t_0 est dit invisible à droite pour h s'il existe un point τ , $t_0 < \tau \leq b$ tel que $h(t_0) < h(\tau)$. Alors, l'ensemble des points invisibles à droite pour h est un ouvert de $[a,b]$ et, par conséquent est une réunion finie ou dénombrable d'intervalles ouverts deux à deux disjoints (a_k, b_k) (et peut être un intervalle semi-ouvert d'extrémité a). Pour chacun de ces intervalles on a $h(a_k) \leq h(b_k)$.

Par le lemme on déduit que

$$g(a_k) - Ca_k \leq g(b_k) - Cb_k \Leftrightarrow g(b_k) - g(a_k) \geq C(b_k - a_k),$$

d'où

$$\sum_k (b_k - a_k) \leq \sum_k \frac{g(b_k) - g(a_k)}{C} \leq \frac{g(b) - g(a)}{C}.$$

Comme C peut être aussi grande que l'on veut, on en déduit que l'ensemble des points pour lesquels Λ_{dr} peut être recouvert par une famille d'intervalles dont la somme des longueurs est aussi petite que l'on veut. Par conséquent, cet ensemble est de mesure de Lebesgue nulle.

Montrons maintenant que $\lambda_{ga} \geq \Lambda_{dr}$ presque partout. On considère deux rationnels $0 < c < C < \infty$ et on pose $\rho = c/C$. On désigne par $E_{c,C}$ l'ensemble des t pour lesquels $\Lambda_{dr} > C$

et $\lambda_{ga} < c$. On va montrer que $E_{c,C}$ est de mesure de Lebesgue nulle et comme l'ensemble des points pour lesquels $\lambda_{ga} < \Lambda_{dr}$ est une réunion finie ou dénombrable d'ensembles de la forme $E_{c,C}$, on aura la conclusion. Pour montrer que $\ell(E_{c,C}) = 0$ (la mesure de Lebesgue sur \mathbb{R} sera notée ici ℓ) on a besoin du lemme suivant (admis) :

Lemme 2 *Soit un sous ensemble E de l'intervalle $[a,b]$ tel que pour tout intervalle $(\alpha,\beta) \subset [a,b]$ on a $\ell(E \cap (\alpha,\beta)) \leq \rho(\beta - \alpha)$, où $0 < \rho < 1$. Alors $\ell(E) = 0$.*

Soit l'ensemble des $t \in (\alpha,\beta)$ pour lesquels $\lambda_{ga} > C$. À tout point de cet ensemble on peut associer un $\tau < t$ tel que

$$\frac{g(\tau) - g(t)}{\tau - t} < c \Leftrightarrow g(\tau) - cg(\tau) > g(t) - cg(t).$$

Ainsi t est invisible à gauche pour $g(t) - ct$ (définition identique), donc d'après un résultat similaire au Lemme 1, l'ensemble de tous ces t est une réunion finie ou dénombrable d'intervalles $(\alpha_k, \beta_k) \subset (\alpha, \beta)$ et $g(\beta_k) - c\beta_k \leq g(\alpha_k) - c\alpha_k$, c'est-à-dire

$$g(\beta_k) - g(\alpha_k) \leq c(\beta_k - \alpha_k).$$

Sur chacun des intervalles (α_k, β_k) on considère l'ensemble G_k des points t pour lesquels $\Lambda_{dr} > C$. On reprend un raisonnement identique et on déduit que G_k est une réunion finie ou dénombrable d'intervalles $(\alpha_{kj}, \beta_{kj})$ et que

$$\beta_{kj} - \alpha_{kj} \leq \frac{1}{C}[g(\beta_{kj}) - g(\alpha_{kj})].$$

Il est clair qu'on peut recouvrir $E_{c,C} \cap (\alpha, \beta)$ par des intervalles $(\alpha_{kj}, \beta_{kj})$ et on a

$$\sum_{k,j} (\beta_{kj} - \alpha_{kj}) \leq \frac{1}{C} \sum_{k,j} [g(\beta_{kj}) - g(\alpha_{kj})] \leq \frac{1}{C} \sum_k [g(\beta_k) - g(\alpha_k)] \leq \frac{c}{C} \sum_k (\beta_k - \alpha_k) \leq \rho(\beta - \alpha).$$

Le Lemme 2 s'applique et le résultat est prouvé dans le cas d'une fonction g continue non décroissante (à l'exception des Lemmes 1 et 2). Si g est monotone discontinue on peut utiliser un résultat identique au Lemme 1 à condition de bien définir la notion de point t_0 invisible à droite dans ce cas : il existe $\tau > t_0$ tel que $\max\{g(t_0-), g(t_0), g(t_0+)\} < g(\tau)$.

Propriété 6 *La fonction de sauts d'une fonction monotone a une dérivée nulle presque partout.*

En effet, une telle fonction est la somme d'une série convergente de fonctions non décroissantes de la forme $q_n(t) = p_n \mathbb{1}_{\{t > t_n\}}$ dont chacune a une dérivée presque partout nulle. Il reste à utiliser le théorème de dérivation terme à terme d'une série de fonctions monotones (admis).

Propriété 7 *Pour toute fonction intégrable au sens de Lebesgue f on a presque partout l'égalité suivante : $\frac{d}{dx} \int_a^x f(t)dt = f(x)$.*

On note $F(x) = \int_a^x f(t)dt$. Montrons d'abord que $f(x) \geq F'(x)$ presque partout. En effet, si $f(x) < F'(x)$, il existe deux rationnels tels que $f(x) < \alpha < \beta < F'(x)$. Soit $E_{\alpha,\beta}$ l'ensemble des x pour lesquels a lieu cette inégalité. Cet ensemble est borélien car f et F' le sont. On va montrer que la mesure de Lebesgue de chacun des ensembles $E_{\alpha,\beta}$ est nulle. Comme leur nombre est au plus dénombrable on en déduit que $\ell(\{x : f(x) < F'(x)\}) = 0$.

Soit $\varepsilon > 0$ et soit $\delta > 0$ tel que $|\int_A f(t)dt| < \varepsilon$, dès que $\ell(A) < \delta$. On choisit l'ouvert $G \subset [a,b]$ tel que $G \supset E_{\alpha,\beta}$ et $\ell(G) < \ell(E_{\alpha,\beta}) + \delta$. Si $x \in E_{\alpha,\beta}$, alors pour tous les $\xi > x$ voisins de x ,

$$\frac{F(\xi) - F(x)}{\xi - x} > \beta \Leftrightarrow F(\xi) - \beta\xi > F(x) - \beta x.$$

Donc x est un point invisible à droite pour la fonction $F(x) - \beta x$ sur chacun des intervalles composant G . Par le Lemme 1, on peut indiquer un ouvert $S = \cup_k (a_k, b_k)$ tel que $E_{\alpha,\beta} \subset S \subset G$ et

$$F(b_k) - \beta b_k \geq F(a_k) - \beta a_k \Leftrightarrow F(b_k) - F(a_k) \geq \beta(b_k - a_k) \Leftrightarrow \int_{a_k}^{b_k} f(t)dt \geq \beta(b_k - a_k),$$

d'où

$$\int_S f(t)dt \geq \beta \ell(S).$$

D'autre part

$$\int_S f(t)dt = \int_{E_{\alpha,\beta}} f(t)dt + \int_{S \setminus E_{\alpha,\beta}} f(t)dt \leq \alpha \ell(E_{\alpha,\beta}) + \varepsilon \leq \alpha \ell(S) + \varepsilon + |\alpha| \delta.$$

On en déduit

$$\alpha \ell(S) + \varepsilon + |\alpha| \delta \geq \beta \ell(S) \Leftrightarrow \ell(S) \leq \frac{\varepsilon + |\alpha| \delta}{\beta - \alpha}.$$

Ainsi, l'ensemble $E_{\alpha,\beta}$ peut être enfermé dans un ensemble ouvert de mesure arbitrairement petite donc $\ell(E_{\alpha,\beta}) = 0$.

Nous avons prouvé que $f(x) \geq F'(x)$ presque partout. En remplaçant $f(x)$ par $-f(x)$ on peut montrer que $-f(x) \geq -F'(x)$ donc $f(x) \leq F'(x)$ presque partout. L'égalité est prouvée.

Propriété 8 La dérivée F' d'une fonction monotone non décroissante F est intégrable au sens de Lebesgue et $\int_a^b F'(x)dx \leq F(b) - F(a)$.

Par définition, la dérivée de F au point x est la limite de $\Phi_h(x) = \frac{1}{h}(F(x+h) - F(x))$ quand $h \rightarrow 0$. La monotonie de F entraîne son intégrabilité donc l'intégrabilité de chacune des fonctions Φ_h . On peut donc intégrer :

$$\int_a^b \Phi_h(x)dx = \frac{1}{h} \int_a^b F(x+h)dx - \frac{1}{h} \int_a^b F(x)dx = \frac{1}{h} \int_a^{b+h} F(x)dx - \frac{1}{h} \int_a^{a+h} F(x)dx.$$

Le membre de droite tend vers $F(b) - F(a+)$ quand $h \downarrow 0$. Par le lemme de Fatou on obtient

$$\int_a^b F'(x)dx \leq \liminf_{h \rightarrow 0} \int_a^b \Phi_h(x)dx = F(b) - F(a+) \leq F(b) - F(a)$$

(l'existence de l'intégrale de F' est assurée également par le lemme de Fatou). Il est facile de voir que $F : [0,1] \rightarrow \mathbb{R}$ donnée par $F(x) = \mathbb{1}_{(\frac{1}{2},1]}(x)$ satisfait l'inégalité stricte!

Définition 1.17 Une fonction F sur un intervalle $[a,b]$ est dite à variation bornée s'il existe une constante C telle que pour toute subdivision $a = x_0 < x_1 < \dots < x_n = b$ on ait la variation totale $\mathbb{V}_a^b[F] := \sum_k |F(x_k) - F(x_{k-1})| \leq C$.

Propriété 9 (admise)

Toute fonction monotone est à variation bornée. Toute fonction à variation bornée est différence de deux fonctions monotones non décroissantes.

Définition 1.18 Une fonction F sur un intervalle $[a, b]$ est dite absolument continue sur $[a, b]$, si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour toute famille au plus dénombrable d'intervalles deux à deux disjoints (a_k, b_k) satisfaisant $\sum_k (b_k - a_k) < \delta$, on a $\sum_k |F(b_k) - F(a_k)| < \varepsilon$.

Propriété 10 (admise)

Toute fonction absolument continue est uniformément continue. Toute fonction absolument continue est à variation bornée. Toute fonction absolument continue est différence de deux fonctions absolument continues monotones non décroissantes.

Propriété 11 Si f est une fonction intégrable au sens de Lebesgue alors $F(x) = \int_a^x f(t)dt$ est une fonction absolument continue.

Si $\{(a_k, b_k)\}$ est une famille quelconque d'intervalles deux à deux disjoints alors

$$\sum_k |F(b_k) - F(a_k)| = \sum_k \left| \int_{a_k}^{b_k} f(t)dt \right| \leq \sum_k \int_{a_k}^{b_k} |f(t)|dt = \int_{\cup_k (a_k, b_k)} |f(t)|dt.$$

En vertu de la continuité absolue de l'intégrale de Lebesgue, la dernière expression tend vers zéro quand la longueur totale des intervalles (a_k, b_k) tend vers zéro.

Propriété 12 La dérivée F' d'une fonction absolument continue sur un intervalle $[a, b]$ est intégrable au sens de Lebesgue sur cet intervalle et pour tout x , $a \leq x \leq b$ on a $\int_a^x F'(t)dt = F(x) - F(a)$.

Il suffit de prouver le résultat pour le cas où F est monotone non décroissante. Alors $\Phi(x) = F(x) - \int_a^x F'(t)dt$ est aussi monotone non décroissante: si $x'' > x'$ on a

$$\Phi(x'') - \Phi(x') = F(x'') - F(x') - \int_{x'}^{x''} F'(t)dt \geq 0.$$

D'autre part Φ est absolument continue comme différence de deux fonctions absolument continues et $\Phi'(x) = 0$ presque partout (par 7). On a besoin d'un lemme (admis):

Lemme 3 Si la dérivée d'une fonction absolument continue monotone non décroissante est nulle presque partout, alors cette fonction est une constante.

On en déduit que Φ est une constante. En faisant $x = a$ dans sa définition on trouve que cette constante est égale à $F(a)$.

Cela termine les idées sur les fonctions monotones. □

Exemples: i) Soit $\lambda > 0$ et $F(t) = (1 - e^{-\lambda t})\mathbb{1}_{[0, \infty[}(t)$. C'est une fonction de répartition dérivable et la densité est $f(t) = \lambda e^{-\lambda t}\mathbb{1}_{[0, \infty[}(t)$ (loi exponentielle).

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

Exemple: (lois gaussiennes)

Soit la fonction borélienne positive

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}.$$

On vérifie que $\int_{\mathbb{R}} f(x) dx = 1$, donc que f est une densité de probabilité. Le graphe de cette fonction est la “cloche de Gauss”, symétrique par rapport à 0. C'est la densité de la loi **gaussienne** (ou **normale**) **centrée réduite**, notée $\mathcal{N}(0,1)$.

Sa fonction de répartition est notée

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx, t \in \mathbb{R}.$$

Cette fonction n'est pas connue explicitement, mais les valeurs sont tabulées. On peut écrire, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} \Phi(-t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx = 1 - \Phi(t). \end{aligned}$$

En particulier $\Phi(0) = \frac{1}{2}$.

Si X a la fonction de répartition Φ alors $X \sim \mathcal{N}(0,1)$. De plus, par le calcul précédent $-X \sim \mathcal{N}(0,1)$, donc X et $-X$ ont la même loi (loi symétrique).

Soit X une variable aléatoire gaussienne centrée réduite. Y est une **variable aléatoire gaussienne** s'il existe deux réels m, σ tels que $Y = \sigma X + m$. Si $\sigma = 0$, la gaussienne Y est dite dégénérée. Si $\sigma > 0$, la fonction de répartition de Y est

$$F_Y(t) = P(Y \leq t) = P(\sigma X + m \leq t) = P\left(X \leq \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right).$$

C'est une fonction dérivable et donc la densité de Y est

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right), y \in \mathbb{R}.$$

On écrit $Y \sim \mathcal{N}(m, \sigma^2)$.

Proposition 1.12 *Soit F une fonction de répartition. On appelle **fonction de quantile** la fonction*

$$G(u) = \inf\{t : F(t) > u\}, u \in]0,1[.$$

Soit U une variable aléatoire uniforme sur $]0,1[$, c'est-à-dire de densité $f_U(x) = \mathbb{1}_{]0,1[}(x)$. Alors $G(U)$ a pour fonction de répartition F .

Preuve: On voit que pour tout $t \in \mathbb{R}$

$$G(U) \leq t \Leftrightarrow \inf\{s : F(s) > U\} \leq t.$$

Cette dernière assertion est satisfaite dès que $F(t) > U$. Autrement dit on a $\{U < F(t)\} \subset \{G(U) \leq t\}$. D'autre part si l'inégalité $\inf\{s : F(s) > U\} \leq t$ est vérifiée, alors $F(s) > U$

pour tout $s > t$, c'est-à-dire on a $\{G(U) \leq t\} \subset \{U < F(s)\}$. Comme U suit la loi uniforme, il en découle des deux inclusions que, pour tout $s > t$

$$F(t) = P(U < F(t)) \leq P(G(U) \leq t) \leq P(U \leq F(s)) = F(s).$$

Comme F est continue à droite, en faisant tendre s vers t , on obtient $P(G(U) \leq t) = F(t)$ pour tout $t \in \mathbb{R}$, ce qui est le résultat. \square

REMARQUE: On peut voir que G est croissante, continue à droite et admet une limite à gauche en tout point de $]0,1[$. Pour voir la continuité à droite il suffit de remarquer que $\{t : F(t) > u\} = \cup_{n \in \mathbb{N}^*} \{t : F(t) > u + \frac{1}{n}\}$. On peut aussi prouver que $F(t) = \inf\{u : G(u) > t\}$ (d'ici aussi le nom, **inverse continue à droite de F**). On peut vérifier que si F est inversible, G est l'inverse de F au sens usuel. La proposition précédente permet de générer des variables aléatoires de loi arbitraire à partir d'une variable uniforme (fournie par l'ordinateur sous forme des nombres pseudo-aléatoires). \square

Définition 1.19 Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à valeurs dans \mathbb{R}^d . On appelle **fonction de répartition** de X ou de sa loi P_X , la fonction sur \mathbb{R}^d définie par

$$F_X(t_1, \dots, t_d) = P_X(]-\infty, t_1] \times \dots \times]-\infty, t_d]) = P(X_1 \leq t_1, \dots, X_d \leq t_d).$$

La loi de la variable aléatoire réelle X_j est appelée la **j -ème marginale** de X . Elle est donnée par sa fonction de répartition:

$$F_{X_j}(t_j) = \lim_{t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_d \rightarrow \infty} F_X(t).$$

REMARQUE: La loi d'un vecteur aléatoire détermine chacune des lois marginales, mais la réciproque est fautive en général. Voici un exemple: soit deux couples aléatoires (X, Y) et (U, V) dont les lois discrètes sont données par

$$P_{X,Y} = \frac{1}{6}\delta_{(0,0)} + \frac{1}{3}\delta_{(0,1)} + \frac{1}{12}\delta_{(1,0)} + \frac{5}{12}\delta_{(1,1)}$$

et

$$P_{U,V} = \frac{1}{4}\delta_{(0,0)} + \frac{1}{4}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,1)}.$$

Les lois marginales sont données par

$$P_X = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 = P_U \quad \text{et} \quad P_Y = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1 = P_V.$$

On produit deux couples aléatoires de lois différentes mais ayant les mêmes lois marginales. \square

REMARQUE: Supposons que la loi du couple (X, Y) est discrète à valeurs dans $\{x_i : i \in I\} \times \{y_j : j \in J\}$:

$$p_{i,j} = P(X = x_i, Y = y_j), i \in I, j \in J.$$

Alors les lois de X et de Y sont données par:

$$p_i = P(X = x_i) = \sum_{j \in J} p_{i,j} \quad \text{et} \quad q_j = P(Y = y_j) = \sum_{i \in I} p_{i,j}.$$

En effet

$$\begin{aligned} p_i &= \mathbb{P}(X = x_i) = \mathbb{P}(X = x_i, Y \in \{y_j : j \in J\}) \\ &= \mathbb{P}\left(\bigcup_{j \in J} (X = x_i, Y = y_j)\right) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j). \end{aligned}$$

□

REMARQUE: Supposons que la loi du couple (X, Y) admet une densité $f_{X, Y}$. Alors les deux lois marginales admettent des densités :

$$f_X(x) = \int_{\mathbb{R}} f_{X, Y}(x, y) dy \text{ et } f_Y(y) = \int_{\mathbb{R}} f_{X, Y}(x, y) dx.$$

En effet,

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) \\ &= \iint_{\mathbb{R}^2} \mathbb{1}_A(x) f_{X, Y}(x, y) dx dy = \int_{\mathbb{R}} \mathbb{1}_A(x) \left(\int_{\mathbb{R}} f_{X, Y}(x, y) dy \right) dx. \end{aligned}$$

□

1.7 Lois de probabilités usuelles

1. Loi de Bernoulli.

$$Q = (1 - p)\delta_0 + p\delta_1, p \in [0, 1].$$

X suit la loi de Bernoulli, notée $\mathcal{B}(1, p)$ si

$$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0).$$

Si $p = 0$ (respectivement $p = 1$) X est presque sûrement égale à 0 (respectivement 1).

Il s'agit de modéliser une expérience à deux issues possibles (succès et échec) dont probabilité de succès vaut p . Le résultat X de cette expérience suit la loi de Bernoulli de paramètre p .

2. Loi binomiale.

$$Q = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} \delta_k, p \in]0, 1[.$$

X suit la loi binomiale, notée $\mathcal{B}(n, p)$ si

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \text{ si } k = 0, \dots, n$$

et $\mathbb{P}(X = k) = 0$ sinon.

Il s'agit de répéter n fois l'expérience à deux issues possibles dont probabilité de succès vaut p , en assurant chaque fois les mêmes conditions initiales (indépendance des répétitions). Le nombre X de succès obtenus suit la loi binomiale.

3. Loi uniforme.

$$Q = \frac{1}{r} \sum_{j=1}^r \delta_{x_j}, r \in \mathbb{N}^*$$

X suit la loi uniforme, notée $\mathcal{U}(x_1, \dots, x_r)$ si

$$P(X = x_j) = \frac{1}{r}, j = 1, \dots, r.$$

On modélise une expérience à r issues possibles dont chacune a la même probabilité $1/r$. Le résultat X de cette expérience suit la loi uniforme.

4. Loi géométrique.

$$Q = \sum_{k=1}^{\infty} p(1-p)^{k-1} \delta_k, p \in]0,1[.$$

X suit la loi géométrique, notée $\mathcal{G}(p)$ si

$$P(X = k) = p(1-p)^{k-1}, k \geq 1.$$

On reprend la modélisation de la loi binomiale, mais cette fois-ci X note la première fois où le succès est obtenu. Alors X suit la loi géométrique.

5. Loi de Poisson.

$$Q = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k, \lambda > 0.$$

X suit la loi de Poisson, notée $\mathcal{P}(\lambda)$ si

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}.$$

6. Loi hypergéométrique.

X suit la loi hypergéométrique, notée $\mathcal{H}(N, n, p)$, $N \in \mathbb{N}^*$, $1 \leq n \leq N$, $p \in]0,1[$, si

$$P(X = k) = \frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n}, \text{ si } \max\{0, n - N(1-p)\} \leq k \leq \min\{n, Np\}$$

et = 0 sinon.

Si on tire n boules en une seule fois, sans remise, dans une urne contenant N boules dont Np sont blanches et $N(1-p)$ sont rouges, le nombre X de boules blanches tirées suit la loi hypergéométrique.

7. Loi multinomiale.

(X_1, \dots, X_d) à valeurs dans \mathbb{N}^d , suit la loi multinomiale, notée $\mathcal{M}(n, p_1, \dots, p_n)$, $n \in \mathbb{N}$, $p_1 + \dots + p_d = 1$, $p_1, \dots, p_d \in [0,1]$:

$$P((X_1, \dots, X_d) = (n_1, \dots, n_d)) = \frac{n!}{n_1! \dots n_d!} p_1^{n_1} \dots p_d^{n_d}, \text{ si } n_1 + \dots + n_d = n$$

et = 0 sinon.

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 (X_1, \dots, X_d) de boules dans les boîtes suivent la loi multinomiale.

8. Loi uniforme.

Une variable aléatoire réelle X suit une loi uniforme sur $[a, b]$, $a < b$, notée $\mathcal{U}_{[a, b]}$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{1}{b-a} \mathbf{1}_{[a, b]}(x).$$

9. Loi exponentielle.

Une variable aléatoire réelle X suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty[}(x).$$

10. Loi gamma.

Une variable aléatoire réelle X suit une loi gamma de paramètres $p > 0$ et $\lambda > 0$, notée $\gamma(p, \lambda)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbf{1}_{[0, \infty[}(x),$$

où $\Gamma(p) := \int_0^\infty x^{p-1} e^{-x} dx$ est la fonction d'Euler de seconde espece.

10. Loi de chi-deux.

Une variable aléatoire réelle X suit une loi de chi-deux de paramètres $n \in \mathbb{N}^*$, notée $\chi^2(n)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{[0, \infty[}(x)$$

11. Loi beta.

Une variable aléatoire réelle X suit une loi beta de paramètres $p > 0$ et $q > 0$, notée $\beta(p, q)$, si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{x^{p-1}(1-x)^{q-1}}{B(p, q)} \mathbf{1}_{[0, 1]}(x),$$

où $B(p, q) := \int_0^1 x^{p-1}(1-x)^{q-1} dx$ est la fonction d'Euler de première espece.

11. Loi de Laplace.

Une variable aléatoire réelle X suit une loi de Laplace si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{1}{2} e^{-|x|}.$$

12. Loi de Cauchy.

Une variable aléatoire réelle X suit une loi de Cauchy si sa densité par rapport à la mesure de Lebesgue est

$$f(x) := \frac{1}{\pi(1+x^2)}.$$

13. Loi de Gauss (ou normale).

Une variable aléatoire réelle X suit une loi de Gauss (ou normale) de paramètres $m \in \mathbb{R}$ et σ^2 , $\sigma > 0$, notée $\mathcal{N}(m, \sigma^2)$, si sa densité par rapport à la mesure de Lebesgue est

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

13. Loi de Gauss (ou normale) multidimensionnelle.

Une variable aléatoire (X_1, \dots, X_d) à valeurs dans \mathbb{R}^d suit une loi de Gauss (ou normale) de paramètres $m \in \mathbb{R}^d$ et Σ , matrice inversible, notée $\mathcal{N}(m, \Sigma)$, si sa densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d est

$$f(x) := \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-m)^* \Sigma^{-1} (x-m)\right), \quad x \in \mathbb{R}^d.$$

1.8 Exercices

1.1. Montrer que $\mathcal{P}(\Omega)$ est une tribu et dire combien d'éléments elle a si Ω possède r éléments.

1.2. Montrer qu'une tribu est stable par intersection dénombrable, par réunion finie, par intersection finie. Montrer qu'on peut remplacer l'axiome de stabilité par réunion dénombrable par l'axiome de stabilité par intersection dénombrable.

1.3. Montrer que l'intersection d'un nombre quelconque de tribus est une tribu; en déduire qu'il existe une seule tribu, minimale pour l'inclusion, dans l'ensemble des tribus contenant un sous-ensemble donné \mathcal{E} de $\mathcal{P}(\Omega)$.

1.4. Soit $\Omega = \{1, 2, 3, 4, 5\}$. Déterminer la tribu $\sigma(\{A, B, C\})$ où $A = \{1, 2\}$, $B = \{2, 3\}$ et $C = \emptyset$. Quelles sont les variables aléatoires réelles définies sur Ω ?

1.5. Montrer que sur \mathbb{R} , la tribu engendrée par les parties fermées est égale à $\mathcal{B}(\mathbb{R})$.

1.6. Montrer que la tribu borélienne coïncide avec la tribu engendrée par les intervalles $]a, b[$ ou par les intervalles $] - \infty, b]$, où $-\infty < a < b < \infty$. Même question lorsqu'on suppose seulement $a, b \in \mathbb{Q}$, $a < b$.

1.7. On considère sur \mathbb{R} la tribu engendrée par les singletons. Montrer que cette tribu coïncide avec la tribu

$$\mathcal{T} = \{A \subset \mathbb{R} : A \text{ ou } A^c \text{ soit au plus dénombrable}\}.$$

1.8. Montrer qu'une fonction réelle est une variable aléatoire par rapport à la tribu triviale (sur Ω) si et seulement si elle est constante.

1.9. Soit X une fonction de l'espace mesurable (Ω, \mathcal{A}) dans l'espace mesurable (E, \mathcal{B}) . Montrer que $\mathcal{A}' = \{X^{-1}(B) : B \in \mathcal{B}\}$ et $\mathcal{B}' = \{B : X^{-1}(B) \in \mathcal{A}\}$ sont des tribus mais que $X(\mathcal{A}) = \{X(A) : A \in \mathcal{A}\}$ n'est pas une tribu en général.

1.10. Soit $(X_n)_{n \in \mathbb{N}}$ une suite bornée de variables aléatoires réelles définies sur un espace mesurable (Ω, \mathcal{A}) . Montrer que $\sup_{n \in \mathbb{N}} X_n$, $\inf_{n \in \mathbb{N}} X_n$, $\limsup_{n \rightarrow \infty} X_n$ et $\liminf_{n \rightarrow \infty} X_n$ sont des variables aléatoires. En déduire que l'ensemble

$$\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe} \}$$

est un événement (c'est-à-dire est un élément de \mathcal{A}).

1.11. Donner un exemple de fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ et d'une tribu sur l'ensemble de départ telles que $|f|$ soit mesurable sans que f le soit.

1.12. Montrer que si $(A_j)_{j \in \mathbb{N}}$ est une famille d'événements, alors pour tout $r \in \mathbb{N}^*$,

$$\sum_{j=0}^r P(A_j) - \sum_{0 \leq j < k \leq r} P(A_j \cap A_k) \leq P\left(\bigcup_{0 \leq j \leq r} A_j\right) \leq \sum_{j=0}^r P(A_j).$$

1.13. Soit l'espace mesurable (Ω, \mathcal{A}) et $P : \mathcal{A} \rightarrow [0, \infty[$ une application additive (c'est-à-dire $P(A \cup B) = P(A) + P(B)$, lorsque $A, B \in \mathcal{A}$ et $A \cap B = \emptyset$), telle que $P(\Omega) = 1$. Montrer que les quatre affirmations suivantes sont équivalentes:

- (i) P est une probabilité (c'est-à-dire elle est σ -additive);
- (ii) P est continue sur des suites croissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}, A_n \subset A_{n+1} \Rightarrow P(\cup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

(on notera pour des telles suites croissantes $\lim_n A_n := \cup_{n \in \mathbb{N}} A_n$);

- (iii) P est continue sur des suites décroissantes :

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}, A_n \supset A_{n+1} \Rightarrow P(\cap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} P(A_n)$$

(on notera pour des telles suites décroissantes $\lim_n A_n := \cap_{n \in \mathbb{N}} A_n$);

- (iv)

$$(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}, A_n \supset A_{n+1} \text{ et } \cap_{n \in \mathbb{N}} A_n = \emptyset \Rightarrow \lim_{n \rightarrow \infty} P(A_n) = 0.$$

1.14. Soit (Ω, \mathcal{A}, P) un espace de probabilité. On considère une suite d'ensembles mesurables $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$ et on note

$$\liminf_n A_n := \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} A_m, \quad \limsup_n A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m.$$

Montrer que:

$$P(\liminf_n A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_n A_n).$$

On dit que la suite $(A_n)_{n \in \mathbb{N}}$ est convergente si $\liminf_n A_n = \limsup_n A_n$. Montrer que si la suite est croissante (respectivement décroissante) alors elle est convergente et

$$\lim_n A_n = \bigcup_{n \in \mathbb{N}} A_n \text{ (respectivement } \lim_n A_n = \bigcap_{n \in \mathbb{N}} A_n).$$

Montrer que si la suite $(A_n)_{n \in \mathbb{N}}$ est convergente on a la propriété de continuité de la mesure μ :

$$P(\lim_n A_n) = \lim_{n \rightarrow \infty} P(A_n).$$

1.15. On dit qu'un événement $A \in \mathcal{A}$ est presque sûr si A est presque sûrement égal Ω , c'est-à-dire $\Omega = A \cup N$, avec N un ensemble négligeable. Soit $(A_j)_{j \in J}$, $J \subset \mathbb{N}$, une famille d'événements presque sûrs. Montrer que $\bigcap_{j \in J} A_j$ est presque sûr.

1.16. Soit (Ω, \mathcal{A}, P) un espace de probabilité. Soit la famille d'ensembles suivante:

$$\mathcal{A}_P = \{C \in \mathcal{P}(\Omega) : \exists A_1, A_2 \in \mathcal{A}, \text{ tels que } A_1 \subset C \subset A_2 \text{ et } P(A_2 \setminus A_1) = 0\}.$$

- (i) Montrer que $\mathcal{A}_P = \sigma(\mathcal{A} \cup \mathcal{N})$, où \mathcal{N} est la classe des ensembles P -négligeables.
- (ii) On définit \bar{P} sur \mathcal{A}_P par $\bar{P}(C) = P(A_1) = P(A_2)$. Montrer que \bar{P} est bien définie (c'est-à-dire que sa valeur ne dépend pas du choix de A_1 et A_2). Montrer que \bar{P} est la seule mesure finie sur $\mathcal{A}_P = \sigma(\mathcal{A} \cup \mathcal{N})$ qui prolonge P (c'est-à-dire qui coïncide avec P sur \mathcal{A}).
- (iii) Montrer que pour toute fonction X réelle \mathcal{A}_P -mesurable, il existe deux variables aléatoires U, V réelles telles que $U \leq X \leq V$ et $V - U = 0$ $P - p.s.$

1.17. Soient X, Y deux variables aléatoires réelles sur un espace de probabilité (Ω, \mathcal{A}, P) .

- (i) $\{X \neq Y\}$ est-il un événement? Montrer que pour tout $B \in \mathcal{B}(\mathbb{R})$ on a

$$|P(X \in B) - P(Y \in B)| \leq P(X \neq Y).$$

- (ii) On suppose que X et Y sont presque sûrement égales. Montrer qu'elles ont la même loi.

1.18. (i) Soient X, Y deux variables aléatoires réelles sur un espace de probabilité (Ω, \mathcal{A}, P) telles que

$$P(Y \leq t < X) = 0, \quad \forall t \in \mathbb{R}.$$

Montrer que $P(Y < X) = 0$.

- (ii) On suppose cette fois que X et Y ont la même loi. Montrer que si $X \geq Y$ $p.s.$ alors X et Y sont presque sûrement égales.

1.19. Une urne contient des boules noires et des boules rouges. On tire au hasard et avec remise à chaque fois, une boule de l'urne. On effectue une série "infinie" de tirages. Montrer que l'événement "obtenir à chaque fois une boule de couleur noire" est négligeable.

1.20. Soit X une variable aléatoire vérifiant

$$P(X = 0) = \frac{1}{6}, P(X = 1) = \frac{1}{2}, P(X = 4) = \frac{1}{3}.$$

Tracer le graphe de la fonction de répartition de X .

1.21. Soit l'espace de probabilité $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Montrer que les applications suivantes sont des variables aléatoires et calculer leur fonction de répartition:

- (i) $X_1(\omega) = 2\omega$.
- (ii) $X_2(\omega) = 2 - 2\omega$.

(iii) $Y(\omega) = \omega^2$.

(iv) $Z_1 = X_1 + X_2$.

(v) $Z_2 = X_1 + Y$.

(vi) $Z_3 = X_1 \wedge 1$.

(vii) $W(\omega) = [10\omega]$ ($[\cdot]$ désigne la partie entière).

(viii)* $U_n(\omega) = f(n\omega)$, avec $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne et périodique de période 1. Montrer de plus que les U_n ont toutes la même loi.

1.22. (i) Trouver la fonction de répartition de la variable aléatoire réelle

$$Y(\omega) := a, \quad \forall \omega \in \Omega.$$

(ii) Supposons que pour tout événement $A \in \mathcal{A}$, $P(A) = 0$ ou $P(A) = 1$. Montrer alors que toute variable aléatoire réelle définie sur (Ω, \mathcal{A}, P) est p.s. constante.

1.23. (i) Montrer qu'il existe une variable aléatoire X à valeurs dans \mathbb{N} telle que:

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

pour une unique valeur de a que l'on déterminera.

(ii) Soit Y une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$. On lui associe la variable aléatoire

$$Z := \begin{cases} \frac{Y}{2}, & \text{si } Y \text{ est pair,} \\ \frac{1-Y}{2}, & \text{si } Y \text{ est impair.} \end{cases}$$

Trouver la loi de Z .

(iii) Soit T une variable aléatoire de loi géométrique de paramètre $p \in]0, 1[$. On considère $U = 4[T/2] - 2T + 1$, où $[\cdot]$ désigne la partie entière. Trouver la loi de U .

1.24. (i) Soit

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

Calculer c pour que f soit la densité de probabilité d'une variable aléatoire X . Calculer sa fonction de répartition. En déduire $P(1 \leq X \leq 2)$.

(ii) Mêmes questions pour

$$g(x) = \frac{c}{x^p} \mathbb{1}_{[1, \infty[}(x), \quad \text{où } p > 1.$$

1.25. Montrer qu'il existe une variable aléatoire X dont la fonction de répartition vaut

$$F(t) = \frac{1}{1 + e^{-t}}, \quad \forall t \in \mathbb{R}.$$

Calculer la densité de X . On pose $Y = e^X$, $Z = X \mathbb{1}_{\{0 < X < 1\}}$ et $U = \mathbb{1}_{\{0 < X < 1\}}$. Trouver les lois de Y , Z et U .

1.26. La durée T d'une communication téléphonique est une variable aléatoire de fonction de répartition

$$F(t) = (1 - e^{-\lambda t}) \mathbb{1}_{[0, \infty[}(t), \quad \text{où } \lambda > 0.$$

Calculer, pour $a, b > 0$, $P(T < 0)$, $P(a \leq T)$, $P(a \leq T \leq b)$. T possède-t-elle une densité?

1.27. Soit X variable aléatoire positive de densité f telle que $f(x) > 0$ si $x > 0$, et de fonction de répartition F . On regarde X comme la durée de vie d'un composant; son taux de panne à l'instant $t > 0$ est défini par

$$r(t) := \lim_{h \downarrow 0} \frac{P(t \leq X < t + h)}{hP(t \leq X)}.$$

(i) Calculer r en termes de F et f et réciproquement. Donner une condition nécessaire et suffisante sur r pour que $\lim_{t \rightarrow \infty} F(t) = 1$.

(ii) Déterminer les densités correspondant aux taux de panne constants.

(iii) Calculer f et F lorsque $r(t) = kt^{\alpha-1}$, $t > 0$, $k, \alpha > 0$ (loi de Weibull).

1.28. Soit $X \sim \mathcal{U}_{[-2,1]}$ et on pose $Y = |X|$. Trouver la fonction de répartition et la densité de Y .

1.29. Quelle est la densité de la variable aléatoire $Y = aX + b$, $a \in \mathbb{R}^*$, $b \in \mathbb{R}$, lorsque la densité de la variable aléatoire réelle X est f ?

1.30. On jette un point D au hasard sur le cercle centré en $O' = (\frac{1}{2}, 0)$ de rayon $\frac{1}{2}$, c'est-à-dire, que pour tout arc \mathcal{A} , $P(D \in \mathcal{A})$ est proportionnel à la longueur de \mathcal{A} . On note aussi le point $C = (1, 0)$.

(i) Calculer le coefficient de proportionnalité. On note Θ la mesure de l'angle $(\widehat{O'C, O'D})$ qui appartient à $] -\pi, \pi]$. Quelle est la loi de Θ ?

(ii) On pose $D = (X, Y)$. Trouver la loi de X (loi arcsinus).

1.31. Soient f, g deux fonctions boréliennes, bornées, définies sur $[0, 1]$. On suppose que pour tout intervalle $I \subset [0, 1]$,

$$\int_0^1 \mathbb{1}_I(x) f(x) dx = \int_0^1 \mathbb{1}_I(x) g(x) dx.$$

Montrer que $f = g$ presque partout. Que peut-on dire lorsque f, g sont continues (à droite)?

1.32. Soient $n \geq 1$ un entier fixé et p_1, p_2, p_3 trois réels positifs tels que $p_1 + p_2 + p_3 = 1$. On note :

$$p_{ij} = \begin{cases} \frac{n!}{i!j!(n-i-j)!} p_1^i p_2^j p_3^{n-i-j}, & \text{si } i + j \leq n, \\ 0, & \text{sinon.} \end{cases}$$

(i) Montrer qu'il existe un couple (X, Y) tel que $P(X = i, Y = j) = p_{ij}$.

(ii) Trouver les lois de X et de Y .

1.33. Le couple aléatoire (X, Y) a la densité, par rapport à la mesure de Lebesgue sur \mathbb{R}^2 ,

$$f(x, y) = cy \mathbb{1}_{[0,1]}(x) \mathbb{1}_{[0,1]}(y).$$

(i) Calculer c .

(ii) Trouver les densités marginales de X et de Y .

Mêmes questions pour

$$g(x, y) = c(x + 3y)e^{-x-2y} \mathbb{1}_{[0, \infty[}(x) \mathbb{1}_{[0, \infty[}(y).$$

1.34. Le couple aléatoire (X, Y) a la densité, par rapport à la mesure de Lebesgue sur \mathbb{R}^2 ,

$$f(x, y) = \frac{1}{\pi} \mathbf{1}_D(x, y), \text{ où } D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

- (i) Calculer $P(X \geq \frac{1}{\sqrt{2}})$, $P(Y \geq \frac{1}{\sqrt{2}})$, puis $P(X \geq \frac{1}{\sqrt{2}}, Y \geq \frac{1}{\sqrt{2}})$.
- (ii) Calculer $P(X \geq Y)$ et $P(X \geq \lambda Y)$.
- (iii) Trouver les marginales de X et de Y .
- (iv) Soient a, b deux réels égaux à ± 1 . Montrer que (X, Y) a la même loi que (aX, bY) . Montrer que (X, Y) a la même loi que (Y, X) .
- (v) On note (R, Θ) les coordonnées polaires de (X, Y) . Trouver la loi du couple (R, Θ) et ses marginales.

1.35. Soit le couple aléatoire (X, Y) . On note $\varphi(x, y) = f(x)g(y)$, avec f, g boréliennes positives. Quelles hypothèses doivent vérifier f et g pour que φ puisse être la densité du couple (X, Y) ? Calculer alors les densités marginales de X et Y et vérifier que φ est le produit de ces densités.

1.36. Soit

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

i) Vérifier que, pour tout $x > 0$:

$$\frac{1}{x\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} (1 + 1/t^2) dt.$$

ii) En déduire, pour $x > 0$:

$$1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

iii) Prouver que :

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} \left(1 - \frac{3}{t^4} \right) dt.$$

En déduire que, pour $x > 0$:

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} \leq 1 - \Phi(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

iv) Prouver que :

$$1 - \Phi(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \text{ lorsque } x \uparrow \infty.$$

Chapitre 2

Espérance des variables aléatoires

2.1 Définitions et théorèmes de convergence

Dans tout ce chapitre nous considérons des variables aléatoires d'un espace de probabilité complet (Ω, \mathcal{A}, P) à valeurs dans \mathbb{R} ou \mathbb{R}^d .

Définition 2.1 Soit $A \in \mathcal{A}$ un événement. La variable aléatoire $X = \mathbb{1}_A$ est **intégrable** et son **espérance** est donnée par:

$$E(X) = E(\mathbb{1}_A) = \int_{\Omega} X(\omega)P(d\omega) := P(A).$$

Définition 2.2 Si X est une variable aléatoire étagée positive $X = \sum_{j=1}^r a_j \mathbb{1}_{A_j}$ avec les A_1, \dots, A_r événements disjoints, alors X est **intégrable** et son espérance est donnée par:

$$E(X) := \sum_{j=1}^r a_j P(A_j) = \sum_{j=1}^r a_j \int_{\Omega} \mathbb{1}_{A_j} dP.$$

REMARQUE: On peut vérifier la linéarité de l'espérance sur les variables aléatoires étagées positives. \square

Définition 2.3 Soit X une variable aléatoire positive sur (Ω, \mathcal{A}, P) . On définit et on note son espérance par:

$$E(X) = \int_{\Omega} X dP = \int_{\Omega} X(\omega)P(d\omega) \\ := \sup \{E(Y) : Y \text{ variable aléatoire étagée positive}, Y \leq X\}.$$

REMARQUE: On peut voir que l'espérance d'une variable aléatoire positive peut être infinie. \square

Proposition 2.1 (i) Si $0 \leq X \leq Y$, alors $0 \leq E(X) \leq E(Y)$.

(ii) Si $A \subset B$ et $X \geq 0$, alors $E(X\mathbb{1}_A) \leq E(X\mathbb{1}_B)$.

(iii) Si $X \geq 0$ et $\alpha \geq 0$, alors $E(\alpha X) = \alpha E(X)$.

(iv) Si $X = 0$, alors $E(X) = 0$.

(v) Si $P(B) = 0$, alors $E(X\mathbb{1}_B) = 0$.

Preuve: L'idée est d'établir la proposition sur des variables aléatoires étagées, puis passer au supremum pour les variables positives. Prouvons, suivant ce schéma, (iii). On voit que si $X = \sum_{j=1}^r a_j \mathbb{1}_{A_j}$, alors

$$E(\alpha X) = \sum_{j=1}^r \alpha a_j P(A_j) = \alpha \sum_{j=1}^r a_j P(A_j) = \alpha E(X).$$

□

Théorème 2.1 (convergence monotone)

1) Soit $(X_n)_{n \in \mathbb{N}}$ une suite croissante de variables aléatoires positives sur (Ω, \mathcal{A}, P) , convergeant ponctuellement vers X . Alors l'espérance de la variable aléatoire positive X satisfait:

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

2) Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires positives. Si on note $Y = \sum_{n \geq 0} Y_n < \infty$, alors

$$E(Y) = \sum_{n \geq 0} E(Y_n).$$

Preuve: X est une variable aléatoire par le Théorème 1.1. Comme X_n est croissante positive, d'après la Proposition 2.1 $E(X_n)$ est une suite croissante et positive, donc admet une limite $\ell \geq 0$ (éventuellement $+\infty$). Comme $X_n \leq X$, par la Proposition 2.1 on voit que $\ell \leq E(X)$. Soit

$$0 \leq Y = \sum_{j=1}^r b_j \mathbb{1}_{B_j} \leq X$$

une variable aléatoire positive étagée. Soit $0 \leq c < 1$. On a

$$\begin{aligned} E(X_n) &\geq E[\mathbb{1}_{\{X_n \geq cY\}} X_n] \geq cE[Y \mathbb{1}_{\{X_n \geq cY\}}] \\ &= c \sum_{j=1}^r b_j P(B_j \cap \{X_n \geq cY\}) \end{aligned}$$

d'après la Proposition 2.1 et la Définition 2.2. Quand $n \rightarrow \infty$,

$$\begin{aligned} \ell &\geq c \sum_{j=1}^r b_j \lim_{n \rightarrow \infty} P(B_j \cap \{X_n \geq cY\}) \\ &= c \sum_{j=1}^r b_j P(B_j) = cE(Y), \end{aligned}$$

la seconde égalité résultant de la Proposition 1.8 3) et du fait que $\cup_n \{X_n \geq cY\} = \Omega$. Mais c était arbitraire dans $[0, 1[$, donc $\ell \geq E(Y)$ pour toute variable étagée $0 \leq Y \leq X$. Donc, par définition $\ell \geq E(X)$ et le théorème est prouvé. □

Théorème 2.2 (lemme de Fatou)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires positives. Alors

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

Preuve: On note $Y_n = \inf_{m \geq n} X_m$. Cette suite est croissante et converge simplement vers $\liminf_{n \rightarrow \infty} X_n$. De plus $Y_n \leq X_n$. Il suffit d'appliquer le Théorème 2.1 et d'utiliser la Proposition 2.1 (i). \square

Soit X une variable aléatoire réelle. On note $X^+ = X \vee 0$ sa partie positive et $X^- = (-X) \vee 0$ sa partie négative. Alors $X = X^+ - X^-$ et $|X| = X^+ + X^-$. On sait que X^+ et X^- sont aussi des variables aléatoires.

Définition 2.4 Soit $X = X^+ - X^-$ une variable aléatoire réelle. On dit que X est **intégrable** si $E(X^+) < \infty$ et $E(X^-) < \infty$, ou, équivalent, si $E(|X|) < \infty$. Dans ce cas, son **espérance** est:

$$E(X) := E(X^+) - E(X^-).$$

Une variable aléatoire d'espérance 0 est dite **centrée**.

Exemple: Soit l'espace de probabilité $(\Omega, \mathcal{A}, \delta_{\omega_0})$ où δ_{ω_0} est la masse de Dirac dans un point (quelconque) fixé $\omega_0 \in \Omega$. Soit X une variable aléatoire quelconque. Alors X est intégrable par rapport à la masse de Dirac et

$$\int X(\omega) \delta_{\omega_0}(d\omega) = X(\omega_0).$$

Plus généralement, si $P = \sum_{j=1}^r a_j \delta_{\omega_j}$ avec $a_j \geq 0$, $\sum_{j=1}^r a_j = 1$ et $\omega_j \in \Omega$, alors $E(X) = \sum_{j=1}^r a_j X(\omega_j)$.

Définition 2.4 Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d , $X = (X_1, \dots, X_d)$. X est **intégrable** si et seulement si X_j sont intégrables, $j = 1, \dots, d$, et

$$E(X) = (E(X_1), \dots, E(X_d)).$$

REMARQUE: En particulier une variable aléatoire à valeurs complexes est intégrable si et seulement si les variables aléatoires parties réelle et imaginaire sont intégrables. \square

Proposition 2.2 Si X, Y sont intégrables et si $\alpha, \beta \in \mathbb{R}$, alors

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

De plus, si $X \leq Y$, alors $E(X) \leq E(Y)$.

Preuve: On suppose d'abord $X, Y \geq 0$ et $\alpha, \beta \geq 0$. D'après la Proposition 1.6 il existe des suites $\{X_n\}_{n \in \mathbb{N}}$ et $\{Y_n\}_{n \in \mathbb{N}}$ de variables aléatoires étagées qui convergent en croissant vers X et Y respectivement. Alors la suite $\alpha X_n + \beta Y_n$ converge en croissant vers $\alpha X + \beta Y$, et le résultat se déduit du théorème de convergence monotone. En général on sépare les parties positive et négative et on distingue selon les signes de α et β . Si $X \leq Y$, alors $Y - X \geq 0$, donc d'après la Proposition 2.1, $E(Y - X) \geq 0$ et la conclusion s'ensuit par linéarité. \square

Proposition 2.3 Soit Y intégrable et soit $\{X_n\}_{n \in \mathbb{N}}$ une suite de variables aléatoires intégrables.

i) Si $Y \leq X_n$, alors $E(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} E(X_n)$.

ii) Si $X_n \leq Y$, alors $\limsup_{n \rightarrow \infty} E(X_n) \leq E(\limsup_{n \rightarrow \infty} X_n)$.

Preuve: Par le lemme de Fatou on a

$$E[\liminf_{n \rightarrow \infty} (X_n - Y)] \leq \liminf_{n \rightarrow \infty} E(X_n - Y)$$

ce qui prouve (i) par linéarité de l'espérance. De même

$$E[\liminf_{n \rightarrow \infty} (Y - X_n)] \leq \liminf_{n \rightarrow \infty} E(Y - X_n).$$

□

Théorème 2.3 (convergence dominée de Lebesgue)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires telles que $|X_n| \leq Y$, où Y est intégrable et la suite converge simplement vers X . Alors

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

Preuve: Comme $\lim_{n \rightarrow \infty} X_n = X$ et $-Y \leq X_n \leq Y$, la Proposition 2.3 donne

$$\begin{aligned} \limsup_{n \rightarrow \infty} E(X_n) &\leq E(\limsup_{n \rightarrow \infty} X_n) = E(X) \\ &= E(\liminf_{n \rightarrow \infty} X_n) \leq \liminf_{n \rightarrow \infty} E(X_n). \end{aligned}$$

□

Proposition 2.4 (inégalité de Jensen)

Si $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction convexe et si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire, telle que X et $\varphi(X)$ soient intégrables, alors

$$\varphi(E(X)) \leq E(\varphi(X)).$$

Preuve: La convexité de φ assure qu'en tout point son graphe est au-dessus de sa tangente: pour tout $t \in \mathbb{R}$, il existe δ (on peut prendre pour δ la dérivée à gauche ou à droite de φ en t) tel que

$$\varphi(x) \geq \varphi(t) + \delta(x - t).$$

On applique cette inégalité à $t = E(X)$ et $x = X(\omega)$, pour tout ω , et on intègre les deux membres. Par la Proposition 2.2 on déduit la conclusion. □

REMARQUE: Si φ est strictement convexe, l'égalité $\varphi(E(X)) = E(\varphi(X))$ n'a lieu que si X est p.s. constante. De plus, si l'égalité a lieu pour toute variable aléatoire X , alors φ est linéaire. □

REMARQUE: Dans la pratique, l'inégalité de Jensen est le plus souvent utilisée pour les fonctions $\varphi(x) = |x|$, x^2 , et $1/x$ lorsque $x > 0$. En particulier, une variable aléatoire dont le carré est intégrable est intégrable, et si X est à valeurs strictement positives,

$$E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}.$$

□

2.2 Théorème de transport

Théorème 2.4 (de transport)

Soit X un vecteur aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et soit ϕ une fonction borélienne de \mathbb{R}^d dans \mathbb{R} . Alors $\phi(X)$ est intégrable si et seulement si ϕ est intégrable sur \mathbb{R}^d par rapport à la loi de X , c'est-à-dire

$$\mathbb{E}(|\phi(X)|) < \infty \iff \int_{\mathbb{R}^d} |\phi(t)| \mathbb{P}_X(dt) < \infty.$$

Si tel est le cas,

$$\mathbb{E}(\phi(X)) = \int_{\Omega} \phi(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}^d} \phi(t) \mathbb{P}_X(dt).$$

En particulier, lorsque X est réelle intégrable,

$$\mathbb{E}(X) = \int_{\mathbb{R}} t \mathbb{P}_X(dt).$$

Preuve: On reprend l'idée générale de construction de l'espérance. On suppose $d = 1$. Si $\phi = \mathbb{1}_B$ pour un borélien B de \mathbb{R} ,

$$\int_{\mathbb{R}} \mathbb{1}_B(t) \mathbb{P}_X(dt) = \mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{E}(\mathbb{1}_B \circ X)$$

et la formule est vraie dans ce cas. Si ϕ est une fonction étagée, la formule est valide par linéarité par rapport à \mathbb{P} . Si ϕ est positive, soit $(\phi_n)_{n \in \mathbb{N}}$ une suite de fonctions positives étagées convergeant en croissant vers ϕ (Proposition 1.6). Alors $\phi_n \circ X$ est une suite de variables aléatoires étagées qui converge ponctuellement en croissant vers $\phi \circ X$. En utilisant le théorème de convergence monotone pour l'intégrale par rapport à \mathbb{P}_X et pour l'espérance,

$$\begin{aligned} \int_{\mathbb{R}} \phi(t) \mathbb{P}_X(dt) &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \phi_n(t) \mathbb{P}_X(dt) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(\phi_n \circ X) = \mathbb{E}(\phi \circ X). \end{aligned}$$

Dans le cas général, on voit que

$$\int_{\mathbb{R}} |\phi(t)| \mathbb{P}_X(dt) = \mathbb{E}(|\phi \circ X|)$$

et donc $\phi \circ X$ est intégrable par rapport à \mathbb{P} si et seulement si ϕ est intégrable par rapport à \mathbb{P}_X . En posant $\phi = \phi^+ - \phi^-$, on conclut que

$$\begin{aligned} \int_{\mathbb{R}} \phi(t) \mathbb{P}_X(dt) &= \int_{\mathbb{R}} \phi^+(t) \mathbb{P}_X(dt) - \int_{\mathbb{R}} \phi^-(t) \mathbb{P}_X(dt) \\ &= \mathbb{E}(\phi^+ \circ X) - \mathbb{E}(\phi^- \circ X) = \mathbb{E}(\phi \circ X). \end{aligned}$$

Le théorème est démontré. □

Dans la pratique, souvent, la loi de X est discrète ou à densité et le théorème de transport peut s'écrire comme dans les deux propositions suivantes :

Proposition 2.5 *Si X est une variable aléatoire discrète de loi donnée par :*

$$P_X = \sum_{j \in J} p_j \delta_{x_j},$$

alors,

$$E(|\phi(X)|) < \infty \iff \sum_{j \in J} |\phi(x_j)| p_j < \infty,$$

et, sous cette condition,

$$\begin{aligned} E(\phi(X)) &= \int_{\mathbb{R}^d} \phi(t) P_X(dt) = \sum_{j \in J} \phi(x_j) p_j \\ &= \sum_{j \in J} \phi(x_j) P(X = x_j). \end{aligned}$$

Proposition 2.6 *Si X est une variable aléatoire à densité f_X par rapport à la mesure de Lebesgue sur \mathbb{R}^d :*

$$P_X(B) = \int_B f_X(t) dt, B \in \mathcal{B}(\mathbb{R}^d),$$

alors,

$$E(|\phi(X)|) < \infty \iff \int_{\mathbb{R}^d} |\phi(t)| f_X(t) dt < \infty,$$

et, sous cette condition,

$$E(\phi(X)) = \int_{\mathbb{R}^d} \phi(t) P_X(dt) = \int_{\mathbb{R}^d} \phi(t) f_X(t) dt.$$

On peut trouver la densité d'une variable aléatoire à l'aide du résultat suivant:

Proposition 2.7 *Supposons qu'il existe une fonction borélienne positive f telle que pour toute fonction ϕ borélienne positive (ou borélienne bornée) on ait*

$$E(\phi(X)) = \int_{\mathbb{R}^d} \phi(t) f(t) dt.$$

Alors f est la densité de X par rapport à la mesure de Lebesgue.

REMARQUE: Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d admettant une densité f_X . Soit g un difféomorphisme sur \mathbb{R}^d , de jacobien $\text{Jac } g(x)$. Alors le vecteur $Y = g(X)$ a pour densité

$$f_Y(y) = (f_X \circ g^{-1})(y) |\text{Jac } g^{-1}(y)|.$$

En effet, si ϕ est une fonction borélienne bornée (par exemple une indicatrice d'un borélien), par la Proposition 2.6 et la formule de changement de variables pour des intégrales de Lebesgue, on a

$$\begin{aligned} \mathbb{E}((\phi \circ g)(X)) &= \int_{\mathbb{R}^d} (\phi \circ g)(t) f_X(t) dt \\ &= \int_{\mathbb{R}^d} \phi(y) (f_X \circ g^{-1})(y) |\text{Jac } g^{-1}(y)| dy, \end{aligned}$$

et il reste à appliquer la Proposition 2.7. \square

Exemples: i) Soit X de loi $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Alors $\mathbb{E}(X) = \frac{1}{2}$. Dans un jeu de pile ou face, on obtient en moyenne une fois sur deux pile ($X = 1$) et une fois sur deux face ($X = 0$).

ii) Soient x_1, \dots, x_r des réels et soit $P_X = \frac{1}{r} \sum_{j=1}^r \delta_{x_j}$. Alors $\mathbb{E}(X) = \frac{1}{r} \sum_{j=1}^r x_j$ est la moyenne des x_j .

iii) Soit $X \sim \mathcal{E}(\lambda)$ Alors,

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} t f_X(t) dt = \int_{\mathbb{R}} t \lambda e^{-\lambda t} \mathbb{1}_{[0, \infty[}(t) dt \\ &= \int_0^{\infty} \lambda t e^{-\lambda t} dt = \frac{1}{\lambda}. \end{aligned}$$

iv) (**loi gaussienne**) Si $X \sim \mathcal{N}(0,1)$ alors, par symétrie

$$\mathbb{E}(X) = \int_{\mathbb{R}} t f_X(t) dt = \int_{\mathbb{R}} t e^{-\frac{t^2}{2}} \frac{dt}{\sqrt{2\pi}} = 0.$$

v) Si X est une variable aléatoire réelle, intégrable, la linéarité de l'espérance implique

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b, \forall a, b \in \mathbb{R}.$$

Donc si $Y \sim \mathcal{N}(m, \sigma^2)$ alors $\mathbb{E}(Y) = m$.

Il est parfois plus commode de calculer l'espérance à partir de la fonction de répartition:

Proposition 2.8 *Si X est une variable aléatoire réelle positive de fonction de répartition F_X , alors*

$$\mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X > t) dt = \int_0^{\infty} (1 - F_X(t)) dt.$$

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

$$\sum_{n \geq 0} \mathbb{P}(X > \varepsilon n) < \infty \text{ ou } \sum_{n \geq 0} 2^n \mathbb{P}(X > \varepsilon 2^n) < \infty.$$

Preuve: Par le théorème de Fubini on peut écrire

$$\begin{aligned} \int_0^{\infty} \mathbb{P}(X > t) dt &= \int_0^{\infty} \mathbb{E}(\mathbb{1}_{]t, \infty[}(X)) dt \\ &= \mathbb{E}\left(\int_0^X dt\right) = \mathbb{E}(X). \end{aligned}$$

Pour la deuxième partie, on voit que

$$\sum_{n \geq 0} \mathbb{P}(X > n + 1) \leq \int_0^{\infty} \mathbb{P}(X > t) dt \leq \sum_{n \geq 0} \mathbb{P}(X > n),$$

en divisant $[0, \infty[$ en intervalles $[n, n + 1[$. De la même façon

$$\begin{aligned} \sum_{n \geq 0} 2^n \mathbb{P}(X > 2^{n+1}) &\leq \int_0^{\infty} \mathbb{P}(X > t) dt \\ &\leq 1 + \sum_{n \geq 0} 2^n \mathbb{P}(X > 2^n). \end{aligned}$$

On conclut en remplaçant X par $\frac{X}{\varepsilon}$. □

Proposition 2.9 (*inégalité de Markov*)

Si X est intégrable et $t > 0$, alors

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

Preuve: On observe que

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

et on intègre cette inégalité par rapport à \mathbb{P} . □

REMARQUE: 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. □

REMARQUE: On peut reformuler le point (iv) de la Proposition 2.1:

(iv') $\mathbb{E}(X) = 0 \Leftrightarrow X = 0$ p.s. □

2.3 Moments, variance et covariance

Pour définir les moments et la variance d'une variable aléatoire réelle X il suffit de considérer dans le théorème de transport $\phi(t) = t^k$, $k \in \mathbb{N}^*$ ou $\phi(t) = (t - \mathbb{E}(X))^2$.

Définition 2.5 *Soit X une variable aléatoire réelle. On dit que X admet un moment d'ordre k , où $k \in \mathbb{N}^*$, si $\mathbb{E}(|X|^k) < \infty$. Sous cette condition le moment d'ordre k de X est*

$$\mu_k(X) = \mathbb{E}(X^k), k \in \mathbb{N}^*.$$

En particulier, $\mu_1(X) = \mathbb{E}(X)$.

Proposition 2.10 *Supposons que la variable aléatoire réelle X admet un moment d'ordre k , $k \in \mathbb{N}^*$. Alors*

$$\mu_k(X) = \int_{\mathbb{R}} t^k \mathbb{P}_X(dt).$$

Si X est réelle discrète

$$\mu_k(X) = \sum_{j \in J} x_j^k p_j$$

et si X admet la densité f_X par rapport à la mesure de Lebesgue sur \mathbb{R}

$$\mu_k(X) = \int_{\mathbb{R}} t^k f_X(t) dt.$$

Lorsque X est positive ayant la fonction de répartition F_X , alors

$$\mu_k(X) = k \int_0^{\infty} t^{k-1} \mathbb{P}(X > t) dt = k \int_0^{\infty} t^{k-1} (1 - F_X(t)) dt.$$

Définition 2.6 Soit X une variable aléatoire réelle dont le carré est intégrable (ayant un moment d'ordre 2). On appelle **variance** de X , ou de sa loi, et on note $\text{Var}(X)$, la quantité

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

La racine $\sqrt{\text{Var}(X)}$ est appelée **écart type**, parfois noté $\sigma(X)$. Une variable aléatoire de variance 1 est dite **réduite**.

REMARQUE: Une expression équivalente de la variance est

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mu_2(X) - \mu_1(X)^2,$$

car grâce à la linéarité de l'espérance

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2 - 2XE(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2. \end{aligned}$$

□

REMARQUE: Si $\text{Var}(X) = 0$ alors $X = \mathbb{E}(X)$ p.s.

□

Proposition 2.11 Supposons que la variable aléatoire réelle X a le carré intégrable. Alors

$$\text{Var}(aX + b) = a^2 \text{Var}(X), a, b \in \mathbb{R}.$$

Exemple: (loi gaussienne) Soit $X \sim \mathcal{N}(0,1)$. Alors X admet des moments de tous ordres; tous les moments impairs sont nuls et

$$\mu_{2k}(X) = \frac{(2k)!}{2^k k!}, k \in \mathbb{N}^*$$

par intégration par parties. Par exemple, le moment d'ordre deux est

$$\mu_2(X) = 1,$$

et donc $\text{Var}(X) = 1$ (d'ici la terminologie gaussienne réduite). Si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $\text{Var}(Y) = \sigma^2$ et $\mu_2(Y) = m^2 + \sigma^2$. Ainsi les paramètres de la loi $\mathcal{N}(m, \sigma^2)$ sont l'espérance et la variance de cette loi.

Proposition 2.12 (*inégalité de Tchebychev*)

Si X est une variable aléatoire réelle dont le carré est intégrable, et $t > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

REMARQUE: Si X admet un moment d'ordre k , $k \in \mathbb{N}^*$, et $t > 0$, alors,

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

□

Preuve: Si X est une variable aléatoire réelle dont le carré est intégrable, alors par l'inégalité de Markov

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(|X|^2)}{t^2},$$

pour tout $t > 0$, puisque $\{X \geq t\} \subset \{|X|^2 \geq t^2\}$. De la même façon on vérifie la remarque ci-dessus.

Il suffit maintenant d'appliquer cette inégalité à la variable $|X - \mathbb{E}(X)|$ pour obtenir l'inégalité de Tchebychev. □

Exemple: (fonction génératrice d'une v.a. à valeurs entières)

La **fonction génératrice** d'une variable aléatoire à valeurs dans \mathbb{N} est définie sur l'intervalle réel $[-1, 1]$ (soit sur le disque unité $\{|z| \leq 1\}$ du plan complexe) par

$$G_X(s) = \mathbb{E}(s^X) = \sum_{n \geq 0} s^n p_n = \sum_{n \geq 0} s^n \mathbb{P}(X = n), |s| \leq 1.$$

La fonction génératrice est une fonction continue sur $\{|s| \leq 1\}$, indéfiniment différentiable sur $\{|s| < 1\}$ et elle détermine la loi qui a servi à la définir, puisque, pour tout $n \in \mathbb{N}$,

$$\left[\frac{d^n}{ds^n} G_X(s) \right]_{s=0} = n! \mathbb{P}(X = n).$$

Ainsi G_X caractérise la loi de X . La terminologie génératrice vient du fait qu'elle doit permettre de calculer les moments de X .

Ainsi, X est intégrable si et seulement si G_X est dérivable à gauche en $s = 1$ et dans ce cas (la dérivée à gauche ici) :

$$\mu_1(X) = \mathbb{E}(X) = G'_X(1).$$

De la même façon, X est de carré intégrable si et seulement si G_X est deux fois dérivable à gauche en $s = 1$ et dans ce cas :

$$\mathbb{E}[X(X-1)] = G''_X(1), \text{ d'où } \mu_2(X) = G'_X(1) + G''_X(1).$$

Plus généralement,

$$\mathbb{E}[X(X-1)\dots(X-k+1)] = \left[\frac{d^k}{ds^k} G_X(s) \right]_{s \uparrow 1}.$$

□

Définition 2.7 Soient X et Y deux variables aléatoires réelles dont les carrés sont intégrables. On appelle **la covariance** de X et Y la quantité

$$\begin{aligned} \text{Cov}(X,Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

En particulier, si $X = Y$, $\text{Cov}(X,X) = \text{Var}(X)$. On appelle **coefficient de corrélation linéaire** la quantité

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}},$$

lorsque les deux variables ne sont pas presque sûrement constantes.

REMARQUE: Comme $|xy| \leq \frac{1}{2}(x^2 + y^2)$, la variable XY est intégrable et la covariance est bien définie. □

REMARQUE: Par l'inégalité Cauchy-Schwarz (voir §2.4) on voit que

$$|\rho(X,Y)| \leq 1$$

avec égalité (respectivement 1 ou -1) si et seulement si Y est une fonction affine de X (respectivement $aX + b$ ou $-aX + b$ avec $a > 0$). □

Définition 2.9 Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d . Sa **matrice (carrée) de covariance** est la matrice:

$$K_X = (\text{Cov}(X_j, X_k))_{1 \leq j, k \leq d}$$

REMARQUE: A la variance se substitue à présent une matrice. Il est facile de voir que :

$$K_X = \mathbb{E}(XX^*) - \mathbb{E}(X)\mathbb{E}(X)^*,$$

(ici, si $x \in \mathbb{R}^d$, xx^* est une matrice $d \times d$).

Cette matrice est symétrique (par la symétrie de la covariance) et semi-définie positive, puisque pour tous réels $\alpha_1, \dots, \alpha_d$,

$$\sum_{1 \leq j, k \leq d} K_X(j,k) \alpha_j \alpha_k = \mathbb{E} \left[\left(\sum_{j=1}^d \alpha_j (X_j - \mathbb{E}(X_j)) \right)^2 \right] \geq 0.$$

Elle est définie positive si aucune combinaison linéaire des composantes X_j n'est p.s. constante. □

2.4 Espaces L^p

Nous avons défini la classe des variables aléatoires (réelles) intégrables sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, notée dorénavant par $\mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Pour $0 < p < \infty$, on note $\mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ (ou simplement \mathcal{L}^p si le contexte est clair) l'ensemble des variables aléatoires X telles que $E(|X|^p) < \infty$. \mathcal{L}^0 est simplement l'ensemble des variables aléatoires. Enfin, on définit \mathcal{L}^∞ l'ensemble des variables aléatoires pour lesquelles il existe $c > 0$ avec

$$\mathbb{P}(\{\omega : |X(\omega)| > c\}) = 0.$$

C'est l'ensemble des variables aléatoires presque sûrement bornées.

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

$$\|X\|_p := (E(|X|^p))^{1/p}.$$

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

$$\|X\|_\infty := \inf \{c > 0 : \mathbb{P}(|X| > c) = 0\}$$

(supremum essentiel de X).

Deux réels $p, q \geq 1$ sont **conjugués** si $\frac{1}{p} + \frac{1}{q} = 1$. On convient que 1 et ∞ sont conjugués.

Proposition 2.13 (*inégalité de Hölder*)

Soient p, q conjugués avec $1 \leq p \leq \infty$, $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$. Alors $XY \in \mathcal{L}^1$ et

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Preuve: Si $p = 1$ ou $p = \infty$ l'inégalité est évidente. Si $\|X\|_p \|Y\|_q = 0$, alors $XY = 0$ p.s. et l'inégalité de Hölder est triviale. Supposons donc $\|X\|_p \|Y\|_q \neq 0$. Par homogénéité, on peut supposer que $\|Y\|_q = 1$. Il suffit donc de prouver que

$$(E(|XY|))^p \leq E(|X|^p).$$

On considère la mesure de probabilité \mathbb{Q} ayant la densité $|Y|^q$ par rapport à la mesure de probabilité \mathbb{P} ,

$$\mathbb{Q}(A) = \int_A |Y(\omega)|^q \mathbb{P}(d\omega) = E(\mathbb{1}_A |Y|^q), A \in \mathcal{A}.$$

L'inégalité à prouver devient

$$\left(\int_\Omega |X(\omega)| |Y(\omega)|^{1-q} \mathbb{Q}(d\omega) \right)^p \leq \int_\Omega |X(\omega)|^p |Y(\omega)|^{-q} \mathbb{Q}(d\omega)$$

ou encore, avec une notation évidente

$$(E^{\mathbb{Q}}(|X| |Y|^{1-q}))^p \leq E^{\mathbb{Q}}(|X|^p |Y|^{-q})$$

qui est une conséquence de l'inégalité de Jensen pour la fonction convexe $\varphi(x) = x^p$, car $(1-q)p = -q$. \square

REMARQUE: Pour $p = q = 2$ on obtient **l'inégalité de Cauchy-Schwarz** :

$$\|XY\|_1 \leq \|X\|_2 \|Y\|_2 \Leftrightarrow E(|XY|) \leq \sqrt{E(|X|^2)} \sqrt{E(|Y|^2)}.$$

□

REMARQUE: Les variables aléatoires p.s. constantes appartiennent à tous les espace \mathcal{L}^p , $1 \leq p \leq \infty$ et (par l'inégalité de Jensen ou de Hölder),

$$\mathcal{L}^{p'}(\Omega, \mathcal{A}, P) \subset \mathcal{L}^p(\Omega, \mathcal{A}, P), \text{ lorsque } 1 \leq p \leq p' \leq \infty.$$

Autrement dit, l'application $p \mapsto \|\cdot\|_p$ est croissante. En effet, soit $p < p'$ et on applique l'inégalité de Hölder à $|X|^p$ et 1 avec les exposants $\frac{p'}{p}$ et $\frac{p'}{p'-p}$:

$$\|X\|_p^p = E(|X|^p \cdot 1) \leq \left(E(|X|^{p \frac{p'}{p}}) \right)^{\frac{p}{p'}} = \|X\|_{p'}^p.$$

De plus

$$\lim_{p \rightarrow \infty} \|X\|_p = \|X\|_\infty.$$

En effet, on voit d'abord que $|X| \leq \|X\|_\infty$, d'où $\|X\|_p \leq \|X\|_\infty$. mais par la croissance de $p \mapsto \|\cdot\|_p$ on déduit que $\lim_{p \rightarrow \infty} \|X\|_p \leq \|X\|_\infty$.

Soit ensuite a arbitraire tel que $a < \|X\|_\infty$. Alors l'événement $A := \{|X| > a\}$ est de probabilité strictement positive. En particulier $P(A)^{\frac{1}{p}} \rightarrow 1$, lorsque $p \rightarrow \infty$. On a $|X|^p \geq a^p \mathbf{1}_A$, d'où $\|X\|_p \geq a P(A)^{\frac{1}{p}}$, donc $\lim_{p \rightarrow \infty} \|X\|_p \geq a$, pour tout $a < \|X\|_\infty$. Comme a était arbitraire, on obtient $\lim_{p \rightarrow \infty} \|X\|_p \geq \|X\|_\infty$. □

Proposition 2.14 (*inégalité de Minkowski*)

Soit $p \geq 1$. Si $X, Y \in \mathcal{L}^p$, alors $X + Y \in \mathcal{L}^p$ et

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Preuve: On observe que p et $\frac{p}{p-1}$ sont conjugués. En utilisant l'inégalité triangulaire et l'inégalité de Hölder,

$$\begin{aligned} \|X + Y\|_p^p &= E(|X + Y|^p) \\ &\leq E(|X||X + Y|^{p-1} + |Y||X + Y|^{p-1}) \\ &\leq (\|X\|_p + \|Y\|_q) \| |X + Y|^{p-1} \|_{\frac{p}{p-1}} \\ &= (\|X\|_p + \|Y\|_q) \|X + Y\|_p^{p-1}. \end{aligned}$$

C'est le résultat si $\|X + Y\|_p \neq 0$. L'inégalité est triviale si $\|X + Y\|_p = 0$. □

REMARQUE: De l'inégalité de Minkowski on déduit que $\|\cdot\|_p$ est une semi-norme (pas norme) sur \mathcal{L}^p (en effet $\|X\|_p = 0$ n'implique pas $X = 0$, mais seulement $X = 0$ p.s.). □

On introduit sur \mathcal{L}^0 la relation suivante:

$$X \sim Y \Leftrightarrow X = Y \text{ p.s.} \Leftrightarrow \{X \neq Y\} \text{ est négligeable.}$$

On peut montrer que cette relation est une relation d'équivalence.

On notera $L^p = L^p(\Omega, \mathcal{A}, P)$ l'espace quotient de \mathcal{L}^p par cette relation d'équivalence. Ainsi un élément X de L^p s'identifie à un représentant de la classe de tous les éléments $Y \in \mathcal{L}^p$ tels que $X = Y$ p.s.

Proposition 2.15 *Pour tout $p \geq 1$, l'espace L^p est un espace vectoriel normé complet (de Banach).*

Nous démontrerons ce résultat au §4.3.

REMARQUE: Pour p et q conjugués et $1 \leq p < \infty$ le dual de L^p est L^q . Autrement dit, les formes linéaires continues sur L^p sont les variables aléatoires de la forme $L^p \ni X \mapsto E(XY) \in \mathbb{R}$, pour $Y \in L^q$. De plus $\|X\|_p = \sup\{E(XY) : \|Y\|_q \leq 1\}$. On prendra garde au fait que L^1 n'est pas en général le dual de L^∞ . \square

On pourra utiliser des arguments géométriques dans les espaces L^2 par le suivant :

Proposition 2.16 *L'espace L^2 est un espace de Hilbert pour le produit scalaire $\langle X, Y \rangle = E(XY)$.*

2.5 Fonctions caractéristiques

On a déjà vu que la fonction de répartition d'une variable aléatoire X caractérise sa loi. Autrement dit, sur \mathbb{R} par exemple, la donnée de

$$F_X(t) = E(\mathbb{1}_{]-\infty, t]}(X)), t \in \mathbb{R}$$

détermine la loi de X . Mais comme les indicatrices sont des fonctions boréliennes bornées, la donnée de $E(\phi(X))$ pour toute fonction borélienne bornée ϕ caractérise la loi P_X . Par exemple $\mathbb{1}_{]-\infty, t]}$ peut être approchée par

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

Il s'ensuit (par convergence dominée) que la donnée de $E(\phi(X))$ pour toute fonction continue bornée ϕ sur \mathbb{R} caractérise P_X . On peut approcher les fonctions indicatrices par des fonctions C^∞ bornées; donc la donnée de $E(\phi(X))$ pour toute fonction infiniment dérivable bornée caractérise P_X . Dans ce paragraphe on prendra la famille des fonctions sinus et cosinus.

Définition 2.8 *Soit X un vecteur aléatoire sur (Ω, \mathcal{A}, P) à valeurs dans \mathbb{R}^d . On appelle **fonction caractéristique** de X ou **transformée de Fourier** de sa loi, et on note φ_X , la fonction à valeurs complexes*

$$\begin{aligned} \mathbb{R}^d \ni t \mapsto \varphi_X(t) &= E(e^{i\langle t, X \rangle}) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} P_X(dx) \\ &= E(\cos \langle t, X \rangle) + iE(\sin \langle t, X \rangle). \end{aligned}$$

REMARQUE: La fonction caractéristique est bien définie car $x \mapsto e^{i\langle t, x \rangle}$ est bornée donc intégrable par rapport à P_X . \square

REMARQUE: Si X est à densité f par rapport à la mesure de Lebesgue sur \mathbb{R}^d , alors

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

est aussi appelée la transformée de Fourier de f . \square

Exemple: (loi gaussienne)

Soit $X \sim \mathcal{N}(0,1)$. Alors

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} \frac{e^{itx-x^2/2}}{\sqrt{2\pi}} dx = e^{-t^2/2}.$$

En effet, soit $z = a + ib \in \mathbb{C}$. Comme $|xe^{zx-x^2/2}| \leq |x|e^{ax-x^2/2}$ avec $\int_{\mathbb{R}} |x|e^{ax-x^2/2} dx < \infty$, la fonction complexe

$$\mathbb{C} \ni z \mapsto \int_{\mathbb{R}} \frac{e^{zx-x^2/2}}{\sqrt{2\pi}} dx \in \mathbb{C}$$

est holomorphe (on peut dériver sous le signe intégrale). En calculant cette intégrale pour z réel, par exemple $z = a$ on trouve

$$\int_{\mathbb{R}} \frac{e^{ax-x^2/2}}{\sqrt{2\pi}} dx = e^{a^2/2}.$$

Comme la fonction complexe

$$\mathbb{C} \ni z \mapsto e^{z^2/2} \in \mathbb{C}$$

est aussi holomorphe et coïncidant sur \mathbb{R} à celle introduite ci-dessus, elles coïncident sur tout \mathbb{C} :

$$\int_{\mathbb{R}} \frac{e^{zx-x^2/2}}{\sqrt{2\pi}} dx = e^{z^2/2}, \forall z \in \mathbb{C}.$$

En particulier, pour $z = it$ on trouve la fonction caractéristique de la loi gaussienne standard. Si $Y \sim \mathcal{N}(m, \sigma^2)$, alors $Y = \sigma X + m$, où $X \sim \mathcal{N}(0,1)$, et on trouve

$$\varphi_Y(t) = \mathbb{E}(e^{itY}) = e^{itm} \mathbb{E}(e^{it\sigma X}) = e^{itm - \frac{t^2\sigma^2}{2}}.$$

Théorème 2.5 *La fonction caractéristique caractérise la loi, c'est-à-dire, $\varphi_X = \varphi_Y$ si et seulement si $P_X = P_Y$.*

Preuve: Supposons, pour simplifier, que X et Y sont deux variables aléatoires réelles. On va noter, pour $\sigma > 0$,

$$f_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \quad \hat{f}_\sigma(t) = e^{-\frac{t^2\sigma^2}{2}}.$$

D'après l'exemple précédent

$$\int_{\mathbb{R}} f_\sigma(x) e^{itx} dx = \hat{f}_\sigma(t),$$

donc

$$\begin{aligned} f_\sigma(t-s) &= \frac{1}{\sqrt{2\pi\sigma^2}} \hat{f}_\sigma\left(\frac{t-s}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f_\sigma(x) e^{i(t-s)x/\sigma^2} dx. \end{aligned}$$

Supposons maintenant que $\varphi_X = \varphi_Y$. En admettant qu'on puisse échanger l'intégration en dx et l'espérance, on peut écrire

$$\begin{aligned} \mathbb{E}[f_\sigma(X-s)] &= \mathbb{E}\left[\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f_\sigma(x) e^{i(X-s)x/\sigma^2} dx\right] \\ &= \int_{\mathbb{R}} f_\sigma(x) \frac{1}{\sqrt{2\pi\sigma^2}} \varphi_X\left(\frac{x}{\sigma^2}\right) e^{-isx/\sigma^2} dx. \end{aligned}$$

La même égalité est vraie pour Y . Donc

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)],$$

pour toute fonction de la forme $f(t) = f_\sigma(t-s)$ pour $s \in \mathbb{R}$ et $\sigma > 0$ arbitraires. Par conséquent aussi pour toute fonction f dans l'espace vectoriel engendré par ces fonctions. D'après le théorème Stone-Weierstrass, cet espace est dense dans l'espace C_0 des fonctions continues sur \mathbb{R} et ayant une limite nulle à l'infini, pour la topologie de la convergence uniforme. Ainsi

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)], \forall f \in C_0.$$

Mais si O est un ouvert, $\mathbb{1}_O$ est limite croissante de fonctions de C_0 . On en déduit que

$$\mathbb{P}_X(O) = \mathbb{E}[\mathbb{1}_O(X)] = \mathbb{E}[\mathbb{1}_O(Y)] = \mathbb{P}_Y(O), \forall O \text{ ouvert}$$

ce qui implique $\mathbb{P}_X = \mathbb{P}_Y$. □

Une autre preuve: On suppose toujours que X et Y sont deux variables aléatoires réelles. Par calcul direct (voir aussi la loi de Laplace)

$$\int_{\mathbb{R}} e^{itx} e^{-\lambda|x|} dx = \frac{1}{\lambda-it} + \frac{1}{\lambda+it} = \frac{2\lambda}{\lambda^2+t^2}.$$

On peut écrire

$$\begin{aligned} \mathbb{E}\left[\frac{2\lambda}{\lambda^2+(X-s)^2}\right] &= \int_{\mathbb{R}} \frac{2\lambda}{\lambda^2+(y-s)^2} \mathbb{P}_X(dy) \\ &= \int_{\mathbb{R}} \mathbb{P}_X(dy) \int_{\mathbb{R}} dx e^{i(t-s)x} e^{-\lambda|x|} = \int_{\mathbb{R}} e^{-isx} e^{-\lambda|x|} \varphi_X(x) dx \end{aligned}$$

où dans la dernière égalité on a utilisé le théorème de Fubini. Soit f une fonction continue à support compact. En appliquant à nouveau le théorème de Fubini et un changement de variable, on en déduit

$$\begin{aligned} &\int_{\mathbb{R}} ds f(s) \int_{\mathbb{R}} \mathbb{P}_X(dy) \frac{2\lambda}{\lambda^2+(y-s)^2} \\ &= \int_{\mathbb{R}} \mathbb{P}_X(dy) \int_{\mathbb{R}} du f(y-\lambda u) \frac{2}{1+u^2} \end{aligned}$$

Pour $\lambda \downarrow 0$, l'intégrant $\int_{\mathbb{R}} f(y - \lambda u) \frac{2}{1+u^2} du$ converge vers $2\pi f(y)$ en restant dominé par $2\pi \|f\|_{\infty}$. Ainsi, par convergence dominée

$$\begin{aligned} \mathbb{E}[f(X)] &= \int_{\mathbb{R}} f(y) P_X(dy) \\ &= \lim_{\lambda \downarrow 0} \int_{\mathbb{R}} ds f(s) \int_{\mathbb{R}} e^{-isx} e^{-\lambda|x|} \varphi_X(x) \frac{dx}{2\pi}. \end{aligned}$$

Le terme de droite est une fonction de f et de φ_X , et donc φ_X caractérise bien la loi P_X . \square

Exemples: i) Si $X = a$ p.s., $a \in \mathbb{R}^d$, alors $P_X = \delta_a$, donc $\varphi_X(t) = e^{i\langle t, a \rangle}$.
ii) Si $X \sim \gamma(p, \lambda)$ alors $\varphi_X(t) = \left(\frac{\lambda}{\lambda - it}\right)^p$.

Proposition 2.17 Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d . Alors :

- i) $|\varphi_X(t)| \leq 1$.
- ii) $\varphi_X(-t) = \overline{\varphi_X(t)}$.
- iii) $\varphi_X(0) = 1$.
- iv) φ_X est uniformément continue.
- v) φ_X est "de type positif":

$$\forall n \in \mathbb{N}^*, \forall t_1, \dots, t_n \in \mathbb{R}^d, \forall z_1, \dots, z_n \in \mathbb{C},$$

$$\sum_{j,k=1}^n \varphi_X(t_k - t_j) z_k \bar{z}_j \geq 0.$$

Preuve: On a $|\varphi_X(t)| \leq \mathbb{E}(|e^{i\langle t, X \rangle}|) = \mathbb{E}(1) = 1$ et $\varphi_X(0) = \mathbb{E}(1) = 1$. Pour prouver (ii) on écrit

$$\begin{aligned} \bar{\varphi}_X(t) &= \overline{\mathbb{E}(e^{i\langle t, X \rangle})} = \mathbb{E}(\overline{e^{i\langle t, X \rangle}}) = \mathbb{E}(e^{-i\langle t, X \rangle}) \\ &= \mathbb{E}(e^{i\langle -t, X \rangle}) = \varphi_X(-t). \end{aligned}$$

Pour montrer (iv) on calcule, pour $t, h \in \mathbb{R}^d$

$$\begin{aligned} \varphi_X(t+h) - \varphi_X(t) &= \mathbb{E}(e^{i\langle t+h, X \rangle} - e^{i\langle t, X \rangle}) \\ &= \mathbb{E}\left[e^{i\langle t, X \rangle} e^{\frac{i}{2}\langle h, X \rangle} \left(e^{\frac{i}{2}\langle h, X \rangle} - e^{-\frac{i}{2}\langle h, X \rangle}\right)\right] \\ &= \mathbb{E}\left[e^{i\langle t, X \rangle} e^{\frac{i}{2}\langle h, X \rangle} 2i \sin \frac{\langle h, X \rangle}{2}\right], \end{aligned}$$

d'où

$$|\varphi_X(t+h) - \varphi_X(t)| \leq \mathbb{E}\left[2 \left|\sin \frac{\langle h, X \rangle}{2}\right|\right] \leq [(|h| |X|) \wedge 2].$$

Le membre de droite de cette dernière inégalité tend vers zéro quand $|h|$ tend vers zéro indépendamment de t .

Enfin, pour (v),

$$\begin{aligned} \sum_{j,k=1}^n \varphi_X(t_k - t_j) z_k \bar{z}_j &= \sum_{j,k=1}^n \mathbb{E} \left(e^{i\langle t_k - t_j, X \rangle} \right) z_k \bar{z}_j \\ &= \mathbb{E} \left[\left(\sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k \right) \left(\sum_{j=1}^n e^{i\langle -t_j, X \rangle} \bar{z}_j \right) \right] \\ &= \mathbb{E} \left[\left(\sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k \right) \overline{\left(\sum_{j=1}^n e^{i\langle t_j, X \rangle} z_j \right)} \right] \\ &= \mathbb{E} \left(\left| \sum_{k=1}^n e^{i\langle t_k, X \rangle} z_k \right|^2 \right) \geq 0. \end{aligned}$$

□

REMARQUE: L'intérêt de la proposition précédente réside surtout dans le fait qu'elle admet la réciproque suivante (théorème de Bochner) : *une fonction φ vérifiant les conditions iii)-v) est la fonction caractéristique d'un vecteur aléatoire.* □

REMARQUE: Il existe plusieurs formules dites d'inversion permettant d'obtenir effectivement la loi à partir de la fonction caractéristique. En voici une possible (théorème d'inversion de Fourier) : *Soit φ une fonction caractéristique intégrable par rapport à la mesure de Lebesgue sur \mathbb{R}^d . Alors φ est la fonction caractéristique d'un vecteur aléatoire X admettant la densité continue bornée donnée par:*

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

□

Proposition 2.18 *Soit X une variable aléatoire réelle de fonction caractéristique φ_X .*

i) Si X admet un moment d'ordre $k \in \mathbb{N}^$, alors, φ_X est k fois dérivable, de dérivée ℓ -ième ($\ell \leq k$)*

$$\varphi_X^{(\ell)}(t) = i^\ell \mathbb{E} \left(X^\ell e^{itX} \right).$$

En particulier, $\varphi_X^{(\ell)}(0) = i^\ell \mathbb{E}(X^\ell)$.

ii) Si φ_X est k fois dérivable en 0, $k \in \mathbb{N}^$, alors X admet des moments d'ordres plus petits ou égaux à $2\ell \leq k$.*

Preuve: i) On peut voir que, pour tout $t \in \mathbb{R}$,

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

En effet, $f_1(u) = i \int_0^u e^{ix} dx$ est de module plus petit que $|u|$ et par récurrence, $f_n(u) = i \int_0^u f_{n-1}(x) dx$ est de module plus petit que $|u|^n/(n!)$. On suppose que $E(|X|) < \infty$ et on prouve que φ_X est dérivable en tout point $t \in \mathbb{R}$. Pour tout $h \neq 0$

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} P_X(dx).$$

D'après l'inégalité ci-dessus pour $n = 1$,

$$\left| e^{itx} \frac{e^{ihx} - 1}{h} \right| \leq |x|$$

fonction qui est intégrable par rapport à P_X indépendamment de h . Par le théorème de convergence dominée

$$\begin{aligned} \varphi'_X(t) &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} P_X(dx) \\ &= \int_{\mathbb{R}} x e^{itx} P_X(dx) = E(X e^{itX}). \end{aligned}$$

Les dérivées d'ordres supérieurs se calculent de la même façon.

ii) D'après le fait que $p \mapsto \|\cdot\|_p$ est croissante, il suffit de prouver l'existence des moments d'ordre pair inférieur à 2ℓ pour assurer celle des moments d'ordre inférieur à 2ℓ . Supposons montrée l'existence du moment d'ordre $2r$, ce qui est vrai pour $r = 0$. D'après la première partie de la proposition,

$$\varphi_X^{(2r)}(0) = (\operatorname{Re} \varphi_X)^{(2r)}(0) = (-1)^r E(X^{2r}),$$

puisque la partie imaginaire $\operatorname{Im} \varphi_X$ est impaire. On en déduit

$$\begin{aligned} (-1)^{r+1} \varphi_X^{(2r+2)}(0) &= \lim_{h \rightarrow 0} \frac{(\operatorname{Re} \varphi_X)^{(2r)}(h) - (\operatorname{Re} \varphi_X)^{(2r)}(0)}{h} \\ &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} x^{2r} \frac{1 - \cos hx}{2h^2} P_X(dx). \end{aligned}$$

En effet, par calcul direct on voit que, par exemple pour $k = 1$

$$\begin{aligned} \varphi_X''(0) &= \lim_{h \rightarrow 0} \frac{1}{2} \left[\frac{\varphi_X'(2h) - \varphi_X'(0)}{2h} + \frac{\varphi_X'(0) - \varphi_X'(-2h)}{2h} \right] \\ &= \lim_{h \rightarrow 0} \frac{2\varphi_X'(2h) - 2\varphi_X'(-2h)}{8h} \\ &= \lim_{h \rightarrow 0} \frac{1}{4h^2} [\varphi_X(2h) - 2\varphi_X(0) + \varphi_X(-2h)] \\ &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} \left(\frac{e^{ihx} - e^{-ihx}}{2h} \right)^2 P_X(dx) \\ &= - \lim_{h \rightarrow 0} \int_{\mathbb{R}} x^2 \left(\frac{\sin hx}{hx} \right)^2 P_X(dx). \end{aligned}$$

Par le lemme de Fatou on conclut que

$$\int_{\mathbb{R}} x^{2r+2} P_X(dx) \leq (-1)^{r+1} \varphi_X^{(2r+2)}(0) < \infty.$$

□

REMARQUE: En général, une loi n'est pas caractérisée par ses moments. □

Définition 2.11 Soit X un vecteur aléatoire sur (Ω, \mathcal{A}, P) à valeurs dans \mathbb{R}^d . On appelle **fonction génératrice des moments** de X ou **transformée de Laplace** de sa loi, et on note

$$\ell_X(s) = \mathbb{E}(e^{-\langle s, X \rangle}),$$

la fonction définie pour les valeurs de s pour lesquelles $e^{-\langle s, X \rangle}$ est intégrable.

REMARQUE: La transformée de Laplace, si elle est finie sur un voisinage de 0, caractérise la loi, comme la transformée de Fourier. □

Proposition 2.19 Soit X une variable aléatoire réelle telle que e^{-sX} est intégrable pour s dans un intervalle contenant 0. Alors la transformée de Laplace ℓ_X est définie sur un intervalle contenant 0. De plus elle est analytique sur un voisinage de 0 et

$$\ell_X(s) = \sum_{n \geq 0} \frac{(-1)^n s^n}{n!} \mathbb{E}(X^n),$$

pour tout s dans ce voisinage. En particulier,

$$\mu_n(X) = \mathbb{E}(X^n) = (-1)^n (\ell_X)^{(n)}(0).$$

2.6 Retour sur les lois de probabilités usuelles

1. Loi de Bernoulli.

Si $X \sim \mathcal{B}(1, p)$, alors

$$\mathbb{E}(X) = p, \text{Var}(X) = p(1 - p)$$

et

$$\varphi_X(t) = 1 - p + pe^{it}$$

2. Loi binomiale.

Si $X \sim \mathcal{B}(n, p)$, alors

$$\mathbb{E}(X) = np, \text{Var}(X) = np(1 - p)$$

et

$$\varphi_X(t) = (1 - p + pe^{it})^n$$

3. Loi uniforme.

Si $X \sim \mathcal{U}(1, \dots, r)$, alors

$$\mathbb{E}(X) = \frac{r+1}{2}, \text{Var}(X) = \frac{r^2-1}{12}$$

et

$$\varphi_X(t) = \frac{1}{r} e^{it} \frac{1 - e^{irt}}{1 - e^{it}}$$

4. Loi géométrique.

Si $X \sim \mathcal{G}(p)$, alors

$$E(X) = \frac{1}{p}, \text{Var}(X) = \frac{1-p}{p^2}$$

et

$$\varphi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}$$

5. Loi de Poisson.

Si $X \sim \mathcal{P}(\lambda)$, alors

$$E(X) = \lambda, \text{Var}(X) = \lambda$$

et

$$\varphi_X(t) = \exp(\lambda(e^{it} - 1))$$

6. Loi hypergéométrique.

Si $X \sim \mathcal{H}(N, n, p)$, alors

$$E(X) = np, \text{Var}(X) = \frac{N-n}{N-1} np(1-p).$$

7. Loi multinomiale.

Si $(X_1, \dots, X_d) \sim \mathcal{M}(n, p_1, \dots, p_n)$, $n \in \mathbb{N}$, alors

$$E(X_1, \dots, X_d) = (np_1, \dots, np_d),$$

$$\text{Var}(X_j) = np_j(1-p_j), \text{Cov}(X_j, X_k) = -np_j p_k$$

et

$$\varphi_X(t_1, \dots, t_d) = \left(\sum_{j=1}^d p_j e^{it_j p_j} \right)^n.$$

8. Loi uniforme.

Si $X \sim \mathcal{U}_{[a,b]}$, alors

$$E(X) = \frac{a+b}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}$$

et

$$\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

9. Loi exponentielle.

Si $X \sim \mathcal{E}(\lambda)$, alors

$$E(X) = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$$

et

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

10. Loi gamma.

Si $X \sim \gamma(p, \lambda)$, alors

$$E(X) = \frac{p}{\lambda}, \text{Var}(X) = \frac{p}{\lambda^2}$$

et

$$\varphi_X(t) = \left(\frac{\lambda}{\lambda - it} \right)^p$$

10. Loi de chi-deux.

Si $X \sim \chi^2(n)$, alors

$$E(X) = n, \text{Var}(X) = 2n$$

et

$$\varphi_X(t) = \frac{1}{(1 - 2it)^{n/2}}$$

11. Loi beta.

Si $X \sim \beta(p, q)$, alors

$$E(X) = \frac{B(p+1, q)}{B(p, q)}, \text{Var}(X) = \frac{pq}{(p+q)^2(p+q+1)}.$$

11. Loi de Laplace.

Si X suit une loi de Laplace, alors

$$E(X) = 0, \text{Var}(X) = 1$$

et

$$\varphi_X(t) = \frac{1}{1 + t^2}$$

12. Loi de Cauchy.

Si X suit une loi de Cauchy, alors

$$\varphi_X(t) = e^{-|t|}.$$

13. Loi de Gauss (ou normale).

Si $X \sim \mathcal{N}(m, \sigma^2)$, alors

$$E(X) = m, \text{Var}(X) = \sigma^2$$

et

$$\varphi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}$$

13. Loi de Gauss (ou normale) multidimensionnelle.

Si $(X_1, \dots, X_d) \sim \mathcal{N}(m, \Sigma)$, alors

$$E(X_1, \dots, X_d) = m, K_X = \Sigma$$

et

$$\varphi_X(t_1, \dots, t_d) = e^{i\langle m, t \rangle - \frac{1}{2} t^* \Sigma t}.$$

2.7 Exercices

2.1. Calculer l'espérance, la variance, la fonction caractéristique et éventuellement la fonction génératrice et la fonction génératrice des moments pour les lois de probabilités usuelles.

2.2. Deux joueurs A et B jouent à un jeu d'argent. A gagne avec probabilité p , $0 < p < 1$, et il mise une somme de s , tandis que B gagne avec une probabilité $1 - p$, en misant une somme s' . Le vainqueur empoche le total des enjeux. Trouver une condition portant sur p, s, s' pour que le jeu soit équitable.

2.3. Calculer l'espérance, la variance et la fonction caractéristique des variables aléatoires décrites dans les exercices **1.20**, **1.21**, **1.23**, **1.24**, **1.25**, **1.28**, **1.33**, **1.34**.

2.4. Une urne contient r boules dont 2 rouges et $r - 2$ noires. On effectue r tirages sans remise et on note X le rang du premier tirage d'une boule rouge et Y le rang du second tirage.

(i) Trouver la loi de (X, Y) et de ses marginales.

(ii) Calculer $\text{Cov}(X, Y)$.

2.5. (i) Une personne a n clés dans sa poche et veut ouvrir sa porte dans l'obscurité. Elle prend au hasard les clés les unes après les autres et les essaye. On note X le nombre de clés qu'elle essaye avant de trouver la bonne. En supposant qu'une clé une fois essayée est ensuite mise de côté, quelle est la probabilité que cette personne tire la bonne clé à la k -ème tentative? Quelle est la loi de X , son espérance, sa variance?

(ii) Une urne contient N boules numérotées de 1 à N . On extrait ces boules de l'urne les unes après les autres, au hasard et sans remise, et on note R_k le numéro porté par la k -ème boule tirée de l'urne. Donner la loi de R_k , son espérance et sa variance.

(iii) On note $W_N = R_1 + \dots + R_N$. Calculer l'espérance de W_N . Trouver la loi du couple (R_j, R_k) , $j \neq k$, et montrer qu'elle ne dépend pas de (j, k) . Calculer $E(W_N^2)$, $E(R_j R_k)$, puis la variance de W_N .

2.6. Calculer l'espérance de la variable aléatoire X décrite dans l'exercice **1.23 (i)** en remarquant que

$$P(X = k) = \frac{1}{4}P(\xi = k) + \frac{3}{4}P(\eta = k), \forall k \in \mathbb{N},$$

où $\eta = \zeta + 1$ et où ξ et ζ sont deux variables de loi $\mathcal{P}(2)$.

2.7. Soit Y une variable aléatoire uniforme sur $\{a, a + 1, \dots, b\}$. A partir des expressions de l'espérance et de la variance de la loi $\mathcal{U}(1, 2, \dots, r)$ déduire $E(Y)$ et $\text{Var}(Y)$.

2.8. Soit $X \sim \mathcal{G}(p)$ et soit $n \in \mathbb{N}^*$.

(i) On note $Y = \inf\{X, n\}$. Calculer $E(Y)$.

(ii) Une urne contient 2 boules rouges et 3 boules noires. On tire successivement et avec remise à chaque fois une boule de l'urne et on cesse les tirages dès qu'une boule rouge est sortie. On désigne par Z le nombre de boules noires obtenues pendant l'expérience. Calculer $E(Z)$ et $\text{Var}(Z)$.

2.9. Soit X une variable aléatoire positive de fonction de répartition F_X . Soit ϕ une fonction positive, croissante de classe C^1 sur $[0, \infty[$, nulle en 0.

(i) Montrer que

$$E(\phi(X)) = \int_0^\infty \phi'(t)(1 - F_X(t))dt.$$

(ii) On suppose que $\phi(X)$ est intégrable. Montrer que

$$\lim_{t \uparrow \infty} \phi(t)P(X > t) = 0.$$

Cas particulier $\phi(t) = t^n$, $n \in \mathbb{N}^*$.

2.10. Soit X une variable aléatoire réelle de densité

$$f_X(x) = c [x\mathbb{1}_{[0,1]}(x) + (2-x)\mathbb{1}_{[1,2]}(x)].$$

Calculer c , $\mu_n(X)$, $n \in \mathbb{N}^*$ et $\text{Var}(X)$.

2.11. (i) Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^d d'espérance $E(X)$ et de matrice de covariance K_X . Soit A une matrice $p \times d$ et $u \in \mathbb{R}^p$. Montrer que

$$E(u + AX) = u + AE(X), K_{u+AX} = AK_A^*.$$

(ii) Soit le couple aléatoire (X, Y) d'espérance $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ et de matrice de covariance

$$K = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

Vérifier que la matrice K est semi-définie positive et ensuite calculer l'espérance et la matrice de covariance de $(X+1, X-Y, X+Y)$.

2.12. On a étudié la glycémie des individus et on a observé qu'elle suit la loi normale. Sur un échantillon de 300 individus on a constaté que 20% des glycémies sont inférieures à 0,82g/l et que 30% des glycémies sont supérieures à 0,98g/l. Déterminer la glycémie moyenne et l'écart-type par rapport à cette moyenne de la glycémie d'un individu.

2.13. Soit $X \sim \mathcal{N}(0,1)$ et on note $Y = X^2$. Trouver la loi de Y et calculer son espérance.

2.14. On pose

$$f(x,y) = \frac{c}{2\pi} \exp \left[-\frac{1}{2}(3x^2 - 3xy + y^2) \right].$$

(i) Déterminer c pour que f soit la densité d'un couple aléatoire (X, Y) .

(ii) Trouver les lois marginales et calculer le vecteur espérance et la matrice de covariance de (X, Y) .

2.15. (i) Soit $X \sim \gamma(p, \lambda)$. Montrer que $X \in L^r$, pour $1 \leq r < \infty$, mais que $X \notin L^\infty$.

Mêmes questions pour $X \sim \mathcal{N}(0,1)$.

(ii) Soit Y une variable aléatoire de densité

$$f_Y(x) = \frac{p}{x^{p+1}} \mathbb{1}_{[1, \infty[}(x), p \geq 1.$$

Montrer que $Y \in L^r$, pour $r < p$ et que $Y \notin L^p$.

(iii) Construire une variable Z telle que $Z \in L^p$ et $Z \notin L^r$, pour tout $r > p$.

2.16. Soit X une variable aléatoire réelle telle que pour tout $t \geq 1$,

$$P(|X| \geq t) \leq \frac{c}{t^\alpha},$$

où $c > 0$ et $\alpha > 1$. Montrer que X admet un moment d'ordre k , pour tout $k \in \mathbb{N}^*$, $k < \alpha$.

2.17. Soient $p, q, r \in]0, \infty[$ tels que $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Montrer que si $X \in L^p$ et $Y \in L^q$, alors $XY \in L^r$ et on a

$$\|XY\|_r \leq \|X\|_p \|Y\|_q.$$

2.18. Montrer que la fonction caractéristique est à valeurs réelles si et seulement si X et $-X$ ont la même loi.

2.19. Montrer que la conjuguée et la partie réelle d'une fonction caractéristique sont aussi des fonctions caractéristiques.

2.20. Soient φ et ψ deux fonctions caractéristiques et $p \in [0, 1]$. Montrer que $p\varphi + (1-p)\psi$ est encore une fonction caractéristique.

2.21. Soit la fonction

$$f(x, y) = \frac{1 + xy(x^2 - y^2)}{4} \mathbb{1}_{[-1, 1] \times [-1, 1]}(x, y).$$

(i) Montrer qu'il existe un couple aléatoire (X, Y) à valeurs p.s. dans $[-1, 1]$ de densité f .

(ii) Trouver les lois et les fonctions caractéristiques de X et Y .

2.22. Soit X une variable aléatoire telle que $E(e^{\lambda X}) < \infty$, pour $\lambda > 0$ (ou seulement $0 < \lambda < \lambda_0$). Montrer que, pour tout $t \in \mathbb{R}$,

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

où $I(t) := \sup_\lambda (\lambda t - \ln E(e^{\lambda X}))$.

2.23. Soit $X \sim \mathcal{N}(0, 1)$.

(i) Trouver la densité f_Y de $Y = e^X$ et calculer $\mu_n(Y)$, $n \in \mathbb{N}^*$.

(ii) On note Y_a , $|a| \leq 1$, la variable aléatoire de densité

$$f_{Y_a}(x) = f_Y(x) (1 + a \sin 2\pi \ln x).$$

Calculer $\mu_n(Y_a)$, $n \in \mathbb{N}^*$.

(iii) En déduire que les moments ne caractérisent pas la loi de probabilité.

Chapitre 3

Indépendance

3.1 Indépendance

Définition 3.1 Sur un espace de probabilité (Ω, \mathcal{A}, P) , deux événements A, B sont dits **indépendants** si

$$P(A \cap B) = P(A)P(B).$$

Exemple. On jette deux dés, un rouge et un noir. Les événements

$$A = \{ \text{le dé rouge montre un nombre strictement supérieur à 4} \}$$

et

$$B = \{ \text{le dé noir montre un 4} \}.$$

sont indépendants, car les deux dé le sont. Pour modéliser on considère

$$(\Omega, \mathcal{A}, P) = (\{(i, j) : i, j = 1, 2, 3, 4, 5, 6\}, \mathcal{P}(\Omega), P), P \text{ la probabilité uniforme.}$$

Alors,

$$A = \{(5, 1), \dots, (5, 6), (6, 1), \dots, (6, 6)\}, P(A) = \frac{1}{3}$$

$$B = \{(1, 4), \dots, (6, 4)\}, P(B) = \frac{1}{6}$$

donc

$$A \cap B = \{(5, 4), (6, 4)\}, P(A \cap B) = \frac{1}{18} = P(A)P(B).$$

REMARQUE. Si A, B sont deux événements indépendants, alors A, B^c , A^c, B et A^c, B^c sont des paires d'événements indépendants. Ainsi on peut dire que les tribus $\sigma(\{A\})$ et $\sigma(\{B\})$ sont indépendantes.

Définition 3.2 Soit (Ω, \mathcal{A}, P) un espace de probabilité.

a) Une famille quelconque d'événements $A_i \in \mathcal{A}$, $i \in I$, est **mutuellement indépendante** si pour tout $J \subset I$ fini

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

b) Une famille quelconque de sous-tribus $\mathcal{A}_i \subset \mathcal{A}$, $i \in I$, est **mutuellement indépendante** si toute famille d'événements $A_i \in \mathcal{A}_i$, $i \in I$, est mutuellement indépendante.

Exemples: i) Soit $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}([0,1]), \lambda)$ et soit, pour tout $n \in \mathbb{N}^*$

$$A_n = \bigcup_{1 \leq j \leq 2^{n-1}} \left] \frac{2(j-1)}{2^n}, \frac{2j-1}{2^n} \right].$$

La famille $\{A_n : n \in \mathbb{N}^*\}$ est mutuellement indépendante.

ii) On jette deux dés, un rouge et un noir. Les événements

$$A_1 = \{ \text{le dé rouge montre un nombre impair} \},$$

$$A_2 = \{ \text{le dé noir montre un nombre impair} \},$$

$$A_3 = \{ \text{la somme des deux dés est un nombre impair} \}$$

sont trois événements indépendants deux à deux (c'est-à-dire, A_1, A_2 sont indépendants, A_2, A_3 sont indépendants et A_1, A_3 sont indépendants), mais ne sont pas indépendants mutuellement: pour $i, j \in \{1, 2, 3\}$, $i \neq j$,

$$P(A_i) = \frac{1}{2}, P(A_i \cap A_j) = \frac{1}{4}, P(A_1 \cap A_2 \cap A_3) = 0.$$

Par la suite par le mot "indépendance" on désigne "indépendance mutuelle" (et toute autre forme d'indépendance, plus faible, sera précisée explicitement).

Proposition 3.1 Si $\mathcal{E}_1, \mathcal{E}_2$ sont deux familles d'ensembles stables par intersection finie (ou deux algèbres) indépendantes, dans l'espace de probabilité (Ω, \mathcal{A}, P) , alors les tribus $\sigma(\mathcal{E}_1)$, $\sigma(\mathcal{E}_2)$ sont indépendantes.

Preuve. Soit $A_1 \in \mathcal{E}_1$. La classe monotone

$$\mathcal{M}_1 := \{A_2 \in \sigma(\mathcal{E}_2) : P(A_1 \cap A_2) = P(A_1)P(A_2)\}$$

des événements indépendants de A_1 contient \mathcal{E}_2 . Elle contient donc la classe monotone engendrée par \mathcal{E}_2 et par le théorème de classe monotone, elle contient $\sigma(\mathcal{E}_2)$. Soit maintenant $A_2 \in \sigma(\mathcal{E}_2)$. La classe monotone

$$\mathcal{M}_2 := \{A_1 \in \sigma(\mathcal{E}_1) : P(A_1 \cap A_2) = P(A_1)P(A_2)\}$$

des événements indépendants de A_2 contient \mathcal{E}_1 d'après ce qui précède et donc $\sigma(\mathcal{E}_1)$. \square .

Définition 3.3 Une famille quelconque de variables aléatoires X_i , $i \in I$, sur (Ω, \mathcal{A}, P) à valeurs dans (E_i, \mathcal{B}_i) est (mutuellement) **indépendante** si la famille des tribus engendrées par les X_i est (mutuellement) indépendante, c'est-à-dire, pour tout $J \subset I$ fini et tous les ensembles mesurables $B_j \in \mathcal{B}_j$, $j \in J$,

$$P(X_j \in B_j : j \in J) = P\left(\bigcap_{j \in J} \{X_j \in B_j\}\right) = \prod_{j \in J} P(X_j \in B_j).$$

Exemples: i) On reprend l'exemple du jet de dés rouge et noir. Soient

$$X_1(i,j) = i, X_2(i,j) = j,$$

les projections de Ω sur la première et la seconde composante. Alors

$$P(A \cap B) = P(X_1^{-1}(\{5,6\}) \cap X_2^{-1}(\{4\})) = P(\{(5,4), (6,4)\}) = \frac{1}{18} = P(A)P(B),$$

donc A et B sont indépendants. On vérifie que X_1 et X_2 le sont aussi. La tribu $\sigma(X_1)$ est formée des ensembles $A_1 \times \{1,2,3,4,5,6\}$ avec $A_1 \subset \{1,2,3,4,5,6\}$ et la tribu $\sigma(X_2)$ est formée des ensembles $\{1,2,3,4,5,6\} \times B_2$ avec $B_2 \subset \{1,2,3,4,5,6\}$. Donc si $A \in \sigma(X_1)$ et $B \in \sigma(X_2)$ sont non vides, alors

$$A = A_1 \times \{1,2,3,4,5,6\}, B = \{1,2,3,4,5,6\} \times B_2$$

et

$$P(A \cap B) = P(A_1 \times B_2) = \sum_{(i,j) \in A_1 \times B_2} \frac{1}{36} = \sum_{i \in A_1} \frac{1}{6} \sum_{j \in B_2} \frac{1}{6} = P(A_1)P(B_2) = P(A)P(B).$$

ii) La famille de variables aléatoires

$$X_n = \mathbb{1}_{A_n} = \sum_{1 \leq j \leq 2^{n-1}} \mathbb{1}_{\left] \frac{2(j-1)}{2^n}, \frac{2j-1}{2^n} \right]}, n \in \mathbb{N}$$

de $[0,1]$ dans $\{0,1\}$ est indépendante. La loi commune des X_n est une loi de Bernoulli de paramètre $\frac{1}{2}$.

REMARQUE: Lorsque les tribus $\mathcal{B}_i, i \in I$ sont engendrées par des familles $\mathcal{E}_i, i \in I$ stables par intersection finie il suffit de prendre les ensembles $B_j, j \in J$ dans chaque famille de générateurs $\mathcal{E}_j, j \in J$.

REMARQUE: Soit $X_i, i \in I$ une famille de variables aléatoires indépendante à valeurs dans (E_i, \mathcal{B}_i) et soient $g_i, i \in I$ fonctions mesurables de (E_i, \mathcal{B}_i) dans (F_i, \mathcal{C}_i) . Alors la famille de variables aléatoires $Y_i = g_i(X_i), i \in I$ est une famille de variables aléatoires indépendantes. En effet, en remplaçant les tribus indépendantes $\mathcal{A}_i, i \in I$ par des tribus plus petites $\mathcal{A}'_i \subset \mathcal{A}_i, i \in I$, nous obtenons une nouvelle famille de tribus indépendante. Aors la remarque résulte du fait que $\sigma(Y_i) = \sigma(g_i(X_i)) \subset \sigma(X_i)$ pour tout $i \in I$.

REMARQUE: On a équivalence entre $A_i, i \in I$ événements indépendants, $\sigma(A_i), i \in I$ tribus indépendantes et $\mathbb{1}_{A_i}, i \in I$ variables aléatoires indépendantes.

Il est utile de reformuler l'indépendance des variables aléatoires en terme de lois de ces variables. On a vu que lorsqu'on donne un vecteur aléatoire $X = (X_1, \dots, X_d)$, sa loi détermine toutes les lois marginales, c'est-à-dire la loi de chacune des X_j . On a aussi vu que la réciproque est fautive en général; néanmoins si les coordonnées sont (mutuellement) indépendantes, les lois marginales déterminent la loi du vecteur.

Proposition 3.2 Soit (X_1, \dots, X_d) une famille finie de variables aléatoires indépendantes sur (Ω, \mathcal{A}, P) à valeurs dans \mathbb{R}^r . La loi $P_{(X_1, \dots, X_d)}$ du vecteur aléatoire sur $(\mathbb{R}^{rd}, \mathcal{B}(\mathbb{R}^{rd}))$ est

égale au produit des lois marginales $P_{X_1} \otimes \dots \otimes P_{X_d}$. Réciproquement, si la loi du vecteur est égale au produit des marginales, alors les variables sont indépendantes.

Preuve. Pour simplifier on va supposer $d = 2$ et $r = 1$. Par définition $P_{X_1} \otimes P_{X_2}$ est l'unique probabilité sur $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ telle que

$$P_{X_1} \otimes P_{X_2}(B_1 \times B_2) = P_{X_1}(B_1)P_{X_2}(B_2), \forall B_1, B_2 \in \mathcal{B}(\mathbb{R}).$$

Supposons X_1, X_2 indépendantes. Alors

$$\begin{aligned} P_{(X_1, X_2)}(B_1 \times B_2) &= P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) = P(X_1^{-1}(B_1))P(X_2^{-1}(B_2)) \\ &= P_{X_1}(B_1)P_{X_2}(B_2) = P_{X_1} \otimes P_{X_2}(B_1 \times B_2) \end{aligned}$$

L'égalité s'étend à la famille des réunions finies disjointes de pavés (fermée à l'intersection finie) laquelle engendre $\mathcal{B}(\mathbb{R}^2)$. La réciproque se déduit de l'identité précédente et de la définition de la loi, car

$$P((X_1, X_2) \in B_1 \times B_2) = P_{X_1}(B_1)P_{X_2}(B_2)$$

équivalent à

$$P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) = P(X_1^{-1}(B_1))P(X_2^{-1}(B_2)). \quad \square$$

REMARQUE : Supposons que les densités marginales de X_1 et X_2 sont à densité f_{X_i} alors la probabilité produit a la densité $f_{X_1}f_{X_2}$. La proposition précédente montre que si les variables aléatoires X_1, X_2 sont indépendantes alors le couple a une densité donnée par

$$f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2), \forall x_1, x_2 \in \mathbb{R}.$$

Réciproquement, si la densité du couple (X_1, X_2) admet une décomposition de type

$$f_{(X_1, X_2)}(x, y) = g(x)h(y), \forall x, y \in \mathbb{R},$$

où g, h sont deux fonctions intégrables positives, alors g et h sont, à des facteurs positifs près, les densités marginales de X_1 et X_2 et X_1, X_2 sont indépendantes. Cette propriété se généralise au cas de plusieurs variables aléatoires.

REMARQUE : Pour deux variables réelle discrètes, X_1, X_2 sont indépendantes si et seulement si

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2), \forall x_1, x_2 \in \mathbb{R}.$$

Proposition 3.3 Une famille quelconque de variables aléatoires $\{X_i : i \in I\}$ sur (Ω, \mathcal{A}, P) à valeurs dans $(BBr^d, \mathcal{B}(\mathbb{R}^d))$ est indépendante si et seulement si pour toute famille finie $J \subset I$ et toute famille finie de fonctions réelles boréliennes bornées ou positives $\phi_j \ j \in J$,

$$E \left(\prod_{j \in J} \phi_j(X_j) \right) = \prod_{j \in J} E(\phi_j(X_j)).$$

Preuve. Supposons la famille $\{X_i : i \in I\}$ indépendante et soit $J \subset I$ une partie finie, par exemple $J = \{1, \dots, n\}$. Par le théorème de Fubini et la définition de la mesure produit

$$\begin{aligned} \mathbb{E} \left(\prod_{j=1}^n \phi_j(X_j) \right) &= \int \prod_{j=1}^n \phi_j(x_j) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int \prod_{j=1}^n \phi_j(x_j) d\mathbb{P}_{X_1}(x_1) \otimes \dots \otimes d\mathbb{P}_{X_n}(x_n) = \prod_{j=1}^n \int \phi_j(x_j) d\mathbb{P}_{X_j}(x_j) = \prod_{j=1}^n \mathbb{E}(\phi_j(X_j)). \end{aligned}$$

La réciproque s'obtient en considérant pour les fonctions $\phi_j = \mathbb{1}_{B_j}$, B_j boréliens. \square

Proposition 3.4 *La famille (X_1, \dots, X_d) de variables aléatoires à valeurs dans respectivement $\mathbb{R}^{r_1}, \dots, \mathbb{R}^{r_d}$ est indépendante si et seulement si pour tout $(t_1, \dots, t_d) \in \mathbb{R}^{r_1} \times \dots \times \mathbb{R}^{r_d}$,*

$$\varphi_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = \varphi_{X_1}(t_1) \dots \varphi_{X_d}(t_d),$$

ou, pour $r_1 = \dots = r_d = 1$,

$$F_{(X_1, \dots, X_d)}(t_1, \dots, t_d) = F_{X_1}(t_1) \dots F_{X_d}(t_d).$$

Preuve. Le produit $\varphi_{X_1} \dots \varphi_{X_d}$ est la fonction caractéristique de la loi produit $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}$ et le résultat s'ensuit car la fonction caractéristique caractérise la loi. Une autre preuve serait de prendre les fonctions bornées $e^{i\langle t_j, x_j \rangle}$ dans la version complexe de la Proposition 3.3. Pour les fonctions de répartition on applique la Proposition 3.3 avec les fonctions bornées $\mathbb{1}_{]-\infty, t_j]}(x_j)$. \square

REMARQUE. Un cas particulier de la Proposition 3.3 est que si X_1, \dots, X_d sont indépendantes et intégrables alors

$$\mathbb{E}(X_1 \dots X_d) = \mathbb{E}(X_1) \dots \mathbb{E}(X_d).$$

Cette propriété ne caractérise pas l'indépendance. Elle décrit une propriété plus faible:

Définition 3.4 *Deux variables aléatoires réelles $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ sont dites **non corrélées** si*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \Leftrightarrow \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = 0.$$

On dit aussi que les variables centrées $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ sont orthogonales dans L^2 (pour son produit scalaire).

Exemples. i) Deux variable aléatoires indépendantes de carré intégrable sont non corrélées.
ii) Soit $X \sim \mathcal{N}(0,1)$ et on note $Y = X^2$. Alors X et Y sont non corrélées. En effet, elles sont de carré intégrable et

$$\mathbb{E}(XY) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X)\mathbb{E}(Y),$$

par les calculs des moments des variables gaussiennes centrées. Toutefois X, Y ne sont pas indépendantes car

$$\mathbb{P}(X \geq 1, Y \geq 1) = \mathbb{P}(X \geq 1) \neq \mathbb{P}(X \geq 1)\mathbb{P}(Y \geq 1),$$

puisque $P(Y \geq 1) \leq 1 = P(Y > -\infty)$.

Proposition 3.5 (*identité de Bienaymé et inégalité de Bienaymé-Tchebychev*).

Si X_1, \dots, X_d sont deux-à-deux non corrélées, alors

$$\text{Var} \left(\sum_{j=1}^d X_j \right) = \sum_{j=1}^d \text{Var}(X_j).$$

On peut en déduire l'inégalité

$$P \left(\left| \sum_{j=1}^d (X_j - E(X_j)) \right| \geq t \right) \leq \frac{1}{t^2} \sum_{j=1}^d \text{Var}(X_j), t > 0.$$

Preuve. Les variables $X_j - E(X_j)$ et $X_k - E(X_k)$, $j \neq k$, sont orthogonales dans L^2 , donc

$$\begin{aligned} \text{Var} \left(\sum_{j=1}^d X_j \right) &= E \left(\left(\sum_{j=1}^d (X_j - E(X_j)) \right)^2 \right) \\ &= \sum_{1 \leq j, k \leq d} E((X_j - E(X_j))(X_k - E(X_k))) = \sum_{j=1}^d E((X_j - E(X_j))^2) = \sum_{j=1}^d \text{Var}(X_j). \end{aligned}$$

L'inégalité Bienaymé-Tchebychev est une conséquence de l'inégalité de Tchebychev. \square

3.2 Sommes de variables aléatoires indépendantes

On a vu que pour des variable aléatoires X_1, \dots, X_n indépendantes de même loi et de carré intégrable, l'inégalité de Bienaymé-Tchebychev montre que si $t > 0$,

$$P \left(\left| \sum_{j=1}^n (X_j - E(X_1)) \right| \geq t\sqrt{n} \right) \leq \frac{\text{Var}(X_1)}{t^2}.$$

Ainsi l'ordre de grandeur de la somme $\sum_{j=1}^n (X_j - E(X_1))$ est au plus \sqrt{n} (on peut dire que $\sum_{j=1}^n X_j$ ressemble à un terme déterministe $nE(X_1)$ de l'ordre n si $E(X_1) \neq 0$ plus un terme aléatoire de l'ordre au plus \sqrt{n}). On va évaluer la loi de $\sum_{j=1}^n X_j$. On précisera la loi du terme aléatoire au chapitre IV. La somme de variables aléatoires indépendantes joue un rôle essentiel en probabilités et donc on calculera sa loi.

Proposition 3.6 Soient X et Y deux variables aléatoires réelles indépendantes sur (Ω, \mathcal{A}, P) . La loi de la somme $X + Y$ est donnée par le **produit de convolution** $P_X \star P_Y$ des lois P_X et P_Y , défini pour toute fonction borélienne bornée $\phi : \mathbb{R} \rightarrow \mathbb{R}$, par

$$\int_{\mathbb{R}} \phi(t)(P_X \star P_Y)(dt) = \iint_{\mathbb{R}^2} \phi(x+y)P_X(dx)P_Y(dy).$$

Preuve. Soit $g(x,y) = x + y$ et on écrit le théorème de transport:

$$\begin{aligned} \int_{\mathbb{R}} \phi(t) P_{X+Y}(dt) &= E(\phi(X+Y)) = E(\phi \circ g(X,Y)) \\ &= \iint_{\mathbb{R}^2} \phi \circ g(x,y) dP_{(X,Y)}(x,y) = \iint_{\mathbb{R}^2} \phi \circ g(x,y) d(P_X \otimes P_Y)(x,y) \\ &= \iint_{\mathbb{R}^2} \phi(x+y) P_X(dx) P_Y(dy). \end{aligned}$$

□.

REMARQUE: Le produit de convolution vérifie un certain nombre de propriétés algébriques issues de sa description en terme de variables aléatoires, mais qui ne suffisent pas à le caractériser:

- i) (élément neutre) $P_X \star \delta_0 = P_X$ (car $X + 0 = X$);
- ii) (commutativité) $P_X \star P_Y = P_Y \star P_X$ (car $X + Y = Y + X$);
- iii) (associativité) $(P_X \star P_Y) \star P_Z = P_X \star (P_Y \star P_Z)$ (car $(X + Y) + Z = X + (Y + Z)$);
- iv) (distributivité) $P_X \star (\lambda P_Y + (1 - \lambda) P_Z) = \lambda(P_X \star P_Y) + (1 - \lambda)(P_X \star P_Z)$, pour tout $\lambda \in [0,1]$. □

REMARQUE: Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{N} . Alors la loi de $X + Y$ est donnée par

$$P(X + Y = n) = \sum_{j=0}^n P(X = j) P(Y = n - j).$$

En effet, $\{X + Y = n\} = \cup_{j=0}^n \{X = j, Y = n - j\}$ et il suffit de remarquer que les événements sont incompatibles et d'utiliser l'indépendance de X et Y . □

REMARQUE: Soient X et Y deux variables aléatoires réelles indépendantes ayant les densités par rapport à la mesure de Lebesgue f et g . Alors la somme $X + Y$ a une densité h par rapport à la mesure de Lebesgue, donnée par le **produit de convolution** des fonctions f, g

$$h(x) := (f \star g)(x) = \int_{\mathbb{R}} f(x - y) g(y) dy = \int_{\mathbb{R}} g(x - y) f(y) dy = (g \star f)(x), x \in \mathbb{R}.$$

En effet, si ϕ est une fonction borélienne bornée

$$\begin{aligned} \int_{\mathbb{R}} \phi(t) (P_X \star P_Y)(dt) &= \iint_{\mathbb{R}^2} \phi(x+y) P_X(dx) P_Y(dy) \\ &= \iint_{\mathbb{R}^2} \phi(x+y) f(x) g(y) dx dy = \iint_{\mathbb{R}^2} \phi(t) f(t-y) g(y) dt dy = \int_{\mathbb{R}} \phi(t) h(t) dt. \end{aligned}$$

Ce produit de convolution a un effet "régularisant" car la fonction h est continue bien que f et g sont seulement mesurables (et positives d'intégrale 1). □

Les fonctions caractéristiques fournissent un autre moyen de déterminer la loi de la somme de deux variables aléatoires indépendantes.

Proposition 3.7 *Si X et Y sont deux variables aléatoires réelles indépendantes sur $(\Omega, \mathcal{A}, \mathbb{P})$, la fonction caractéristique de leur somme est donnée par le produit des fonctions caractéristiques*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), t \in \mathbb{R}.$$

Preuve. Il s'agit d'une conséquence de la Proposition 3.3, car pour tout $t \in \mathbb{R}$,

$$\varphi_{X+Y}(t) = \mathbb{E}\left(e^{it(X+Y)}\right) = \mathbb{E}\left(e^{itX}e^{itY}\right) = \mathbb{E}\left(e^{itX}\right)\mathbb{E}\left(e^{itY}\right) = \varphi_X(t)\varphi_Y(t).$$

□

REMARQUE: Il ne faut pas confondre la fonction caractéristique d'un couple aléatoire dont les coordonnées sont indépendantes

$$\varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t), (s,t) \in \mathbb{R}^2$$

avec la fonction caractéristique de la somme

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), t \in \mathbb{R}.$$

Exemples. i) Si $X = a$ p.s. et $Y = b$ p.s., alors X et Y sont indépendantes, et

$$X + Y = a + b, \text{ p.s.} \Leftrightarrow \delta_a \star \delta_b = \delta_{a+b}.$$

ii) Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ sont indépendantes, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.

iii) Si $X \sim \mathcal{B}(n,p)$ et $Y \sim \mathcal{B}(m,p)$ sont indépendantes, alors $X + Y \sim \mathcal{B}(n + m,p)$.

iv) Si $X \sim \gamma(p,\lambda)$ et $Y \sim \gamma(q,\lambda)$ sont indépendantes, alors $X + Y \sim \gamma(p + q,\lambda)$.

v) Si $X \sim \mathcal{N}(m,\sigma^2)$ et $Y \sim \mathcal{N}(n,\tau^2)$ sont indépendantes, alors $X + Y \sim \mathcal{N}(m + n,\sigma^2 + \tau^2)$.

vi) On jette une pièce n fois et on veut savoir la loi du nombre de pile. On modélise les n jets par n variables aléatoires indépendantes X_1, \dots, X_n chacune de loi de Bernoulli $\mathcal{B}(1,p)$ (avec $p = \frac{1}{2}$ si la pièce n'est pas truquée), où $X_j = 1$ si au j -ème lancer on obtient pile et $= 0$ sinon. Le nombre de pile est donc $S_n = X_1 + \dots + X_n$ et par iii) on voit que $S_n \sim \mathcal{B}(n,p)$. En particulier $\frac{S_n}{n}$ est le nombre moyen de pile sur n lancers. De plus par l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \text{Var}\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{\varepsilon^2 n}.$$

Donc la probabilité que $\frac{S_n}{n}$ s'écarte de sa valeur moyenne tend vers 0 lorsque n tend vers l'infini. Cela signifie que si nous lançons un grand nombre de fois une pièce non truquée la proportion de pile sera avec un grande probabilité $\frac{1}{2}$.

3.3 Applications de l'indépendance

Étant données des probabilités P_j , par exemple sur \mathbb{R} peut-on trouver des variables aléatoires X_j sur un certain espace de probabilité (Ω, \mathcal{A}, P) qui soient indépendantes et telles que $P_{X_j} = P_j$? Lorsqu'il y a un nombre fini (P_1, \dots, P_d) cela ne pose pas de problème: on prend $\Omega = \mathbb{R}^d$ muni de sa tribu borélienne, $P = P_1 \otimes \dots \otimes P_d$ et on considère $X_j : \mathbb{R}^d \rightarrow \mathbb{R}$ les applications coordonnées.

On va décrire à présent le cas dénombrable $\{P_j : j \in \mathbb{N}\}$. On cherche à construire un espace de probabilité (Ω, \mathcal{A}, P) et une famille de variables aléatoires indépendantes $\{X_j : j \in \mathbb{N}\}$ sur (Ω, \mathcal{A}, P) à valeurs dans (E_j, \mathcal{B}_j) , $j \in \mathbb{N}$, telles que $P_{X_j} = P_j$ (en fait en pratique $E_j = \mathbb{R}$ ou \mathbb{R}^d).

Soit $\Omega = \prod_{j \in \mathbb{N}} E_j$ et on pose X_j la projection sur la j -ème coordonnée. Soit \mathcal{A} la tribu produit des \mathcal{B}_j , $j \in \mathbb{N}$ ou équivalent, la tribu engendrée par les X_j . Cette tribu est aussi engendrée par la famille \mathcal{C} des **cylindres** qui sont les ensembles A de la forme

$$A = C_n \times E_{n+1} \times E_{n+2} \times \dots, C_n \in \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n$$

(C_n est la **base** de A). Sur \mathcal{C} on peut définir la fonction $Q : \mathcal{C} \rightarrow [0, 1]$, par

$$Q(A) := P_1 \otimes \dots \otimes P_n(C_n), A \in \mathcal{C} \text{ (de base } C_n).$$

Théorème 3.1 (de Kolmogorov).

La fonction d'ensemble Q se prolonge en une unique probabilité P sur (Ω, \mathcal{A}) . Sous P , les variables aléatoires X_j sont indépendantes et de lois P_j .

De ce théorème (admis) on peut parler librement d'une suite $\{X_n : n \in \mathbb{N}\}$ de variables aléatoires réelles indépendantes sur un espace de probabilité (Ω, \mathcal{A}, P) .

REMARQUE. Voici un exemple de construction de suite de variables aléatoires indépendantes: on prend l'espace de probabilité $[0, 1], \mathcal{B}([0, 1]), \lambda$ et on écrit $\omega \in \Omega$ sous sa forme dyadique, $\omega = \sum_{n=1}^{\infty} X_n(\omega) \frac{1}{2^n}$ (le cas où tous les $X_n(\omega)$ valent 1 à partir d'un certain rang est exclue). La suite $\{X_n : n \geq 1\}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $\frac{1}{2}$. En effet, si pour tout entier m on donne x_1, \dots, x_m de 0 et 1 et si on pose $y = \sum_{j=1}^m x_j \frac{1}{2^j}$, alors

$$\{X_1 = x_1, \dots, X_m = x_m\} = [y, y + \frac{1}{2^m}[\Rightarrow P(X_1 = x_1, \dots, X_m = x_m) = \frac{1}{2^m} = \prod_{j=1}^m P(X_j = x_j).$$

Définition 3.5 Soit $\{\mathcal{T}_n : n \in \mathbb{N}\}$ une famille indépendante de tribus sur (Ω, \mathcal{A}, P) (par exemple, on peut prendre $\mathcal{T}_n = \sigma(X_n)$ avec $\{X_n : n \in \mathbb{N}\}$ variables aléatoires indépendantes). On désigne par \mathcal{A}_n la tribu engendrée par $\mathcal{T}_n, \mathcal{T}_{n+1}, \dots$ (dans l'exemple la tribu est engendrée par X_n, X_{n+1}, \dots) et on note la **tribu asymptotique** ou **terminale**:

$$\mathcal{A}_{\infty} := \bigcap_{n \in \mathbb{N}} \mathcal{A}_n$$

Théorème 3.2 (loi du tout ou rien ou loi de 0-1).

Pour tout $A \in \mathcal{A}_{\infty}$, $P(A) = 0$ ou $= 1$ (\mathcal{A}_{∞} est presque sûrement triviale, c'est-à-dire est presque sûrement égale à $\{\emptyset, \Omega\}$).

Preuve. Soit $A \in \mathcal{A}_\infty$ fixé. On considère la classe monotone des événements indépendants de A ,

$$\mathcal{M} := \{B \in \mathcal{A} : P(A \cap B) = P(A)P(B)\}.$$

Si on montre que $\mathcal{M} \supset \mathcal{A}_\infty$, alors $A \in \mathcal{M}$ et donc $P(A) = P(A)^2$, c'est-à-dire $P(A) = 0$ ou $= 1$.

Soit, pour tout n , $\mathcal{B}_n = \sigma(\mathcal{T}_0, \dots, \mathcal{T}_n)$ et on pose $\mathcal{B}_\infty = \cup_n \mathcal{B}_n$. En tant que réunion croissante, \mathcal{B}_∞ est une algèbre. On peut voir que les tribus \mathcal{B}_n et \mathcal{A}_{n+1} sont indépendantes. Ainsi, tout élément de \mathcal{B}_n est indépendant de $A \in \mathcal{A}_\infty \subset \mathcal{A}_{n+1}$. Ainsi $\mathcal{B}_\infty \subset \mathcal{M}$. Par le théorème de classe monotone, $\sigma(\mathcal{B}_\infty) = \mathcal{M}(\mathcal{B}_\infty) \subset \mathcal{M}$. Il reste à prouver que $\sigma(\mathcal{B}_\infty) \supset \mathcal{A}_\infty$. 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. \square

Exemple. Soit $\{A_n : n \in \mathbb{N}\}$ une suite d'événements indépendants de (Ω, \mathcal{A}, P) . Alors

$$A = \limsup_n A_n = \{A_n \text{ se réalise une infinité de fois}\}$$

est un événement terminal pour la suite de tribus $\mathcal{T}_n = \sigma(A_n) = \{\emptyset, A_n, A_n^c, \Omega\}$. Donc $P(A) = 0$ ou 1 .

On abrège " A_n se réalise une infinité de fois" par " A_n infiniment souvent" ou " A_n i.s.". Notons que $P(A_n \text{ i.s.}) = 0$ signifie que p.s. seulement un nombre fini d'événements A_n se réalisent: pour presque tous $\omega \in \Omega$, il existe $n(\omega) \in \mathbb{N}$ (dépend de ω) tel que si $n \geq n(\omega)$, alors $\omega \notin A_n$, c'est-à-dire A_n ne se réalise pas.

Théorème 3.3 (*Lemme de Borel-Cantelli*).

Soit $\{A_n : n \in \mathbb{N}\}$ une suite d'événements de (Ω, \mathcal{A}, P) .

a) $\sum_{n \in \mathbb{N}} P(A_n) < \infty \implies P(A_n \text{ i.s.}) = 0$.

b) Si la suite $\{A_n : n \in \mathbb{N}\}$ est indépendante, alors la réciproque est vraie:

$$\sum_{n \in \mathbb{N}} P(A_n) = \infty \implies P(A_n \text{ i.s.}) = 1.$$

Preuve. La partie a) est évidente, car 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, lorsque $n \rightarrow \infty$ si la série converge.

Pour prouver b) on remarque d'abord que pour tout n et tout $N \geq n$,

$$P\left(\bigcup_{n \leq m \leq N} A_m\right) = 1 - P\left(\bigcap_{n \leq m \leq N} A_m^c\right) = 1 - \prod_{n \leq m \leq N} (1 - P(A_m)).$$

Comme $1 - x \leq e^{-x}$, pour tout $x \geq 0$, on obtient

$$P\left(\bigcup_{n \leq m \leq N} A_m\right) \geq 1 - \exp\left(-\sum_{n \leq m \leq N} P(A_m)\right).$$

Par l'hypothèse, lorsque $N \rightarrow \infty$, $\sum_{n \leq m \leq N} P(A_m) \rightarrow \infty$, pour tout n . Donc $P(\cup_{m \geq n} A_m) = 1$ et il suffit de remarquer que $P(A) = \lim_{n \rightarrow \infty} P(\cup_{m \geq n} A_m)$. \square

Exemple. On lance une pièce équilibrée une infinité de fois. Quelle est la probabilité d'obtenir une infinité de fois deux pile consécutifs? Si $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires indépendantes de Bernoulli de paramètre $\frac{1}{2}$ et si on pose $A_n = \{X_n = X_{n+1} = 1\}$ on s'intéresse à $P(A_n \text{ i.s.})$. $\{A_n : n \in \mathbb{N}\}$ n'est pas une suite indépendante car X_{n+1} détermine à la fois A_n et A_{n+1} , mais $\{A_{2n} : n \in \mathbb{N}\}$ est indépendante, avec $P(A_{2n}) = \frac{1}{4}$, pour tout n . On peut appliquer la deuxième partie du lemme de Borel-Cantelli, donc $P(A_{2n} \text{ i.s.}) = 1$. Comme $\{A_{2n} \text{ i.s.}\} \subset \{A_n \text{ i.s.}\}$, on conclut que $P(A_n \text{ i.s.}) = 1$.

3.4 Vecteurs gaussiens et indépendance

$X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ si sa densité par rapport à la mesure de Lebesgue sur \mathbb{R} est

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

ou, équivalent sa fonction caractéristique est

$$\varphi_X(t) = \exp\left(imt - \frac{1}{2}\sigma^2 t^2\right), t \in \mathbb{R}.$$

Si $Y \sim \mathcal{N}(0,1)$, alors $X = m + \sigma Y \sim \mathcal{N}(m, \sigma^2)$. Les paramètres d'espérance m et de variance σ^2 caractérisent la loi.

Définition 3.6 Le vecteur aléatoire $X = (X_1, \dots, X_d) : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ suit une loi gaussienne si pour tout $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$, la variable aléatoire réelle

$$\langle \alpha, X \rangle = \sum_{j=1}^d \alpha_j X_j$$

est gaussienne.

Dans cette définition la variable $\langle \alpha, X \rangle$ est caractérisée par son espérance

$$E\left(\sum_{j=1}^d \alpha_j X_j\right) = \sum_{j=1}^d \alpha_j E(X_j)$$

et par sa variance

$$\text{Var}\left(\sum_{j=1}^d \alpha_j X_j\right) = \sum_{1 \leq j, k \leq d} \alpha_j \alpha_k E((X_j - E(X_j))(X_k - E(X_k))).$$

Ainsi le vecteur gaussien $X = (X_1, \dots, X_d)$ est caractérisé par son vecteur espérance

$$m = E(X) = (E(X_1), \dots, E(X_d))^*$$

et sa matrice covariance

$$\Sigma = (\mathbb{E}((X_j - \mathbb{E}(X_j))(X_k - \mathbb{E}(X_k))))_{1 \leq j, k \leq d}.$$

Un vecteur gaussien centré a la matrice de covariance $(\mathbb{E}(X_j X_k))_{1 \leq j, k \leq d}$. En termes de fonction caractéristique:

$$\varphi_X(t) = \exp\left(i \langle u, m \rangle - \frac{1}{2} t^* \Sigma t\right), t \in \mathbb{R}^d.$$

Notons que si X est un vecteur gaussien ses coordonnées sont des variables aléatoires gaussiennes mais la réciproque est fautive en général.

Un exemple fondamental est le vecteur $G = (G_1, \dots, G_d)$ dont les composantes sont indépendantes de même loi $\mathcal{N}(0,1)$. Ce vecteur est centré et sa matrice de covariance est la matrice identité. Ceci est un vecteur gaussien et sa densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d est $(2\pi)^{-d/2} \exp(-\|x\|^2/2)$ où $\|x\|^2 = x_1^2 + \dots + x_d^2$. On notera $\mathcal{N}_d(0, I_d)$ la loi de G .

Toute matrice de covariance est symétrique et semi-définie positive (voir Remarque après Définition 2.9). Donc elle peut s'écrire AA^* , A matrice carrée.

Proposition 3.8 (facultatif) *Soit $X = (X_1, \dots, X_d)$ un vecteur gaussien centré de matrice de covariance $\Sigma = AA^*$, où A est une matrice carrée. Alors X a la même loi que AG , où $G \sim \mathcal{N}_d(0, I_d)$.*

Preuve. Si A est inversible, pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} \mathbb{P}(X \in B) &= \mathbb{P}(G \in A^{-1}(B)) = \frac{1}{(2\pi)^{d/2} |\det A|} \int_B \exp\left(-\frac{1}{2} \langle A^{-1}x, A^{-1}x \rangle\right) dx \\ &= \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \int_B \exp\left(-\frac{1}{2} x^* \Sigma^{-1} x\right) dx. \end{aligned}$$

□

REMARQUE. Pour simuler un vecteur gaussien général il suffit de simuler des variables gaussiennes standard indépendantes (voir aussi l'exercice 3.16 (ii)). On peut être encore plus précis: on peut toujours écrire $\Sigma = Q\Delta Q^*$, où Q est une matrice orthogonale ($Q^{-1} = Q^*$) et Δ est une matrice diagonale positive (avec éventuellement des zéros sur la diagonale). Par changement de base on voit que le vecteur gaussien Q^*X a pour matrice de covariance la matrice diagonale Δ , donc la même loi que $\sqrt{\Delta}G$. La diagonalisation de la matrice de covariance de X nous a donc permis de trouver une nouvelle base dans laquelle les composantes de X sont orthogonales:

Proposition 3.9 (facultatif) *Soit X un vecteur gaussien centré de matrice de covariance Σ . Si les composantes de X sont deux à deux non corrélées (c'est-à-dire Σ est diagonale), alors la famille (X_1, \dots, X_d) est indépendante.*

REMARQUE. Si $G \sim \mathcal{N}_d(0, I_d)$ et si Q est une matrice orthogonale quelconque, alors $QG \sim \mathcal{N}_d(0, I_d)$.

3.5 Probabilité (et espérance) conditionnelle

Reprenons l'exemple du lancer d'un dé. Cette expérience aléatoire peut être modélisée en donnant $\Omega = \{1,2,3,4,5,6\}$ muni de la tribu de ses parties $\mathcal{P}(\Omega)$ et la probabilité uniforme $P(\{\omega\}) = \frac{1}{6}$, pour tout $\omega \in \Omega$. La variable aléatoire résultat du lancer est l'identité sur Ω . Supposons maintenant que le lancer s'effectue sans regarder le dé et qu'un spectateur nous dise que nous avons obtenu un chiffre impair. Avec cette information on peut recalculer les chances d'avoir obtenu un certain résultat. Par exemple il est clair que la chance d'obtenir un résultat pair est nulle et celle d'obtenir un résultat parmi 1, 3 et 5 est $\frac{1}{3}$. On note $\Omega_{\text{impair}} = \{1,3,5\}$. Nous calculons la probabilité de $\{\omega\}$ sachant que $\omega \in \Omega_{\text{impair}}$ en calculant

$$\frac{P(\{\omega\} \cap \Omega_{\text{impair}})}{P(\Omega_{\text{impair}})}$$

Définition 3.7 Soit (Ω, \mathcal{A}, P) un espace de probabilité et soit $B \in \mathcal{A}$ un événement tel que $P(B) > 0$.

i) On appelle **probabilité conditionnelle de A sachant B** le nombre

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

ii) On appelle **probabilité conditionnelle sachant B** la probabilité notée $P(\cdot | B)$

$$\mathcal{A} \ni A \mapsto \frac{P(A \cap B)}{P(B)}.$$

REMARQUE: Si A et B sont indépendants $P(A | B) = P(A)$, c'est-à-dire que dans cette situation, la connaissance de B n'apporte aucune information quant à la réalisation de A .

REMARQUE: Les égalités suivantes sont vraies pour tous les événements A et B (de probabilité positive ou pas):

$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A).$$

REMARQUE: Soit X une variable aléatoire réelle intégrable. On peut calculer son espérance par rapport à la probabilité (conditionnelle) $P(\cdot | B)$, nommée **espérance conditionnelle sachant B**:

$$E(X | B) := \int_{\Omega} X(\omega)P(\cdot | B)(d\omega) = \frac{1}{P(B)} \int_B X(\omega)P(d\omega) = \frac{1}{P(B)}E(X\mathbf{1}_B).$$

De la même façon on peut parler de fonction de répartition conditionnelle ou de fonction caractéristique conditionnelle sachant B :

$$\mathbb{R} \ni t \mapsto P(X \leq t | B), \mathbb{R} \ni t \mapsto E(e^{itX} | B)$$

lesquelles caractérisent la loi de X sachant B .

Exemples. i) Soit $X \sim \mathcal{E}(\lambda)$. Par calcul direct on voit que, pour tout $s, t > 0$ $P(X > s + t | X > s) = P(X > t)$. On dit que la loi exponentielle n'a pas de mémoire. On peut

vérifier la même chose pour la loi géométrique.

ii) Soient U_1, \dots, U_n variables aléatoires indépendantes de même loi uniforme sur $[0,1]$. On note $M_n = \max_{1 \leq j \leq n} U_j$ et $m_n = \min_{1 \leq j \leq n} U_j$. Par calcul direct on peut vérifier que pour tous $u_1, \dots, u_n \in [0,1]$ et tous $0 \leq a < b \leq 1$,

$$P(U_1 \leq u_1, \dots, U_n \leq u_n \mid a \leq m_n \leq M_n \leq b) = P(V_1 \leq u_1, \dots, V_n \leq u_n),$$

où V_1, \dots, V_n sont variables aléatoires indépendantes de même loi uniforme sur $[a,b]$.

Définition 3.8 Soit (Ω, \mathcal{A}, P) un espace de probabilité. Une famille d'événements $(B_j)_{j \in J}$, $J \subset \mathbb{N}$, est **système complet d'événements** si les B_j sont disjoints et $P(\cup_{j \in J} B_j) = 1$. Ainsi $(B_j)_{j \in J}$ forme une partition de Ω quitte à ajouter un événement négligeable.

Proposition 3.10 Soit $(B_j)_{j \in J}$ un système complet d'événements sur (Ω, \mathcal{A}, P) .

1. Pour tout $A \in \mathcal{A}$ la formule des probabilités totales a lieu:

$$P(A) = \sum_{j \in J} P(A \mid B_j)P(B_j) = \sum_{j \in J^*} P(A \mid B_j)P(B_j),$$

où $J^* = \{j \in J : P(B_j) > 0\}$.

2. Si de plus $P(A) > 0$, pour tout $k \in J^*$ la formule de Bayes a lieu:

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{j \in J^*} P(A \mid B_j)P(B_j)}.$$

Preuve. C'est immédiat car $P(A) = \sum_{j \in J^*} P(A \cap B_j)$ et pour tout k ,

$$P(B_k \mid A)P(A) = P(B_k \cap A) = P(A \mid B_k)P(B_k).$$

□

REMARQUE: (facultatif)

On observe que la tribu engendrée par une partition $(B_j)_{j \in J}$ est décrite comme la collection de toutes les unions possibles d'événements B_j et de leurs complémentaires. Ainsi tout événement peut être fractionné sur l'ensembles élémentaires B_j .

Définition 3.9 (facultatif) Un événement B d'une tribu \mathcal{B} est un **atome** si pour tout événement $C \in \mathcal{B}$ tel que $C \subset B$, soit $C = \emptyset$ soit $C = B$.

REMARQUE: (facultatif)

Ainsi si $(B_j)_{j \in J}$ est une partition mesurable de Ω , alors B_j sont les atomes de la tribu engendrée par les B_j , $\mathcal{B} = \sigma(B_j : j \in J)$.

Définition 3.10 (facultatif) Soit $\mathcal{B} = \sigma(B_j : j \in J)$, où $(B_j)_{j \in J}$ est une partition mesurable sur (Ω, \mathcal{A}, P) et soit $J^* = \{j \in J : P(B_j) > 0\}$. On appelle **probabilité conditionnelle de $A \in \mathcal{A}$ sachant \mathcal{B}** , la variable aléatoire

$$P(A \mid \mathcal{B}) = \sum_{j \in J^*} P(A \mid B_j) \mathbb{1}_{B_j}.$$

REMARQUE: (facultatif)

La probabilité conditionnelle de A sachant \mathcal{B} est donc une variable aléatoire constante sur les atomes de \mathcal{B} , donc mesurable par rapport à \mathcal{B} . Pour tout $\omega \in \Omega$ l'application $\mathcal{A} \ni A \mapsto P(A | \mathcal{B})(\omega)$ est une probabilité telle que $P(B_j | \mathcal{B})(\omega) = 1$ si $\omega \in B_j$ et $P(A | \mathcal{B})(\omega) = 0$, si $\omega \in B_j$ et $P(A \cap B_j) = 0$. On peut aussi remarquer que, pour tout $B \in \mathcal{B}$ tel que $P(B) > 0$,

$$E(P(A | \mathcal{B})\mathbb{1}_B) = P(A \cap B) = E(\mathbb{1}_A \mathbb{1}_B).$$

On pourrait aussi écrire

$$P(A | \mathcal{B}) = \sum_{j \in J^*} \left(\frac{1}{P(B_j)} E(\mathbb{1}_A \mathbb{1}_{B_j}) \right) \mathbb{1}_{B_j}.$$

Définition 3.11 (facultatif) Soit $\mathcal{B} = \sigma(B_j : j \in J)$, où $(B_j)_{j \in J}$ est une partition mesurable sur (Ω, \mathcal{A}, P) et soit $J^* = \{j \in J : P(B_j) > 0\}$. Soit X une variable aléatoire intégrable. On appelle **espérance conditionnelle de X sachant \mathcal{B}** la variable aléatoire \mathcal{B} -mesurable, définie P-p.s.

$$E(X | \mathcal{B}) = \sum_{j \in J^*} \left(\frac{1}{P(B_j)} E(X \mathbb{1}_{B_j}) \right) \mathbb{1}_{B_j}.$$

Si \mathcal{B} est engendrée par une variable aléatoire discrète Z on note $E(X | \mathcal{B}) = E(X | Z)$.

3.6 Exercices

3.1. Montrer les équivalences suivantes:

(i) pour A événement

$$P(A) = 0 \text{ ou } = 1 \Leftrightarrow A, B \text{ indépendants pour tout } B \in \mathcal{A}$$

(ii) pour A, B événements

$$A, B \text{ indépendants} \Leftrightarrow A, B^c \text{ indépendants} \Leftrightarrow A^c, B \text{ indépendants} \Leftrightarrow A^c, B^c \text{ indépendants}$$

(iii) pour A, B, C événements

$$\begin{aligned} A, B, C \text{ indépendants} &\Leftrightarrow A, B, C^c \text{ indépendants} \Leftrightarrow A, B^c, C^c \text{ indépendants} \\ &\Leftrightarrow A^c, B^c, C^c \text{ indépendants} \end{aligned}$$

Montrer que sous une de ces conditions A est indépendant de $B \cup C$, de $B \setminus C$.

(iv) On lance deux fois de suite un dé et on considère

$$A = \{ \text{le premier résultat est un nombre pair} \},$$

$$B = \{ \text{le deuxième résultat est un nombre pair} \}$$

et

$$C = \{ \text{la somme des deux résultats est un nombre pair} \}.$$

Étudier l'indépendance de A, B, C .

3.2. On dispose de deux pièces de monnaie. La première a une probabilité $\frac{1}{4}$ de donner pile et la deuxième a une probabilité $\frac{1}{2}$ de donner pile. On choisit au hasard une pièce et on effectue une série de lancers avec la règle suivante: si au n -ème lancer la pièce donne pile (respectivement face) on effectue le lancer suivant avec la même pièce (respectivement avec l'autre pièce). On note α_n la probabilité de lancer la première pièce au n -ème jet. Montrer que α_n satisfait à une relation du type $\alpha_{n+1} = a\alpha_n + b$. En déduire la valeur de α_n .

3.3. Étudier l'indépendance des coordonnées des couples aléatoires (X,Y) des exercices **1.32**, **1.33**, **1.34**, **1.35**, **2.4**, **2.11**, **2.14**, **2.21**.

3.4. Soit les variables aléatoires indépendantes X,Y . Calculer la loi de la somme $X + Y$ lorsque:

- (i) $X \sim \mathcal{B}(n,p)$, $Y \sim \mathcal{B}(m,p)$;
- (ii) $X \sim \mathcal{P}(\lambda)$, $Y \sim \mathcal{P}(\mu)$;
- (iii) $X \sim \gamma(p,\lambda)$, $Y \sim \gamma(q,\lambda)$;
- (iv) $X \sim \mathcal{N}(m,\sigma^2)$, $Y \sim \mathcal{N}(n,\tau^2)$;
- (v) Quelle est la loi de $X^2 + Y^2$ lorsque $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$?

3.5. Soit les variables aléatoires indépendantes X,Y de même loi $\mathcal{U}(1,2, \dots, n)$.

- (i) Calculer $P(X = Y)$.
- (ii) Trouver la loi de $X + Y$.
- (iii) On pose $U = \min\{X,Y\}$ et $V = \max\{X,Y\}$. Trouver les lois de (U,V) , U et V .
- (iv) U et V sont-elles indépendantes?

3.6. Soit les variables aléatoires indépendantes X,Y de même loi $\mathcal{G}(p)$, $0 < p < 1$. On pose $U = \min\{X,Y\}$ et $V = \max\{X,Y\} - U$. Trouver les lois de U et V et étudier leur indépendance.

3.7. Deux personnes ont rendez-vous à 14h00 mais elles sont peu ponctuelles: les instants X et Y de leur arrivées sont deux variables aléatoires indépendantes uniformément répartis dans $[14,15]$. Calculer la loi de la variable T durée d'attente du premier arrivé.

3.8. Soient X et Y deux variables aléatoires centrées ayant seulement les valeurs ± 1 . Trouver leurs lois et donner une condition nécessaire et suffisante pour qu'elles soient non corrélées.

3.9. Soit F une fonction de répartition de classe C^1 sur \mathbb{R} .

- (i) Si X a la fonction de répartition F , montrer que X admet une densité.
- (ii) Soit Y une variable aléatoire indépendante de X et de même loi. Montrer que $P(X = Y) = 0$ et que $P(X < Y) = \frac{1}{2}$. Montrer que (X,Y) et (Y,X) ont la même loi.
- (iii) Soit Z encore une variable aléatoire indépendante de (X,Y) et de même loi que X . Calculer $P(X < Y < Z)$ et montrer que la valeur ne dépend pas de F . Montrer que (X,Y,Z) et (Y,Z,X) ont la même loi.
- (iv) Généraliser au cas de n variables aléatoires indépendantes de même loi.

3.10. Soient X_1, \dots, X_n variables aléatoires réelles indépendantes de même loi. On note $U = \min_{1 \leq j \leq n} X_j$ et $V = \max_{1 \leq j \leq n} X_j$.

- (i) Calculer les fonction de répartition de U et V à l'aide de la fonction de répartition commune

des X_j .

(ii) Si la loi commune des X_j admet une densité, montrer que U et V admettent des densités et les calculer.

3.11. Soient U et $X = (X_1, \dots, X_n)$ deux variables aléatoires indépendantes, U à valeurs dans $\{1, 2, \dots, n\}$ et X ayant une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^n . On note $X_U(\omega) = X_{U(\omega)}(\omega)$. Montrer que X_U est une variable aléatoire ayant une densité. Étudier ensuite le cas où $U \sim \mathcal{U}(1, 2, \dots, n)$ et les X_j sont de même loi.

3.12. Une urne contient b boules blanches et r rouges; une boule étant tirée, on la remet et avec elle encore c boules de la couleur tirée. On pose $p = \frac{b}{b+r}$, $q = 1 - p$, $\gamma = \frac{c}{b+r}$. On note

$$X_n = \mathbb{1}_{\{\text{la } n\text{-ème boule tirée est blanche}\}}$$

et $S_n = X_1 + \dots + X_n$.

(i) Trouver les lois de X_1 et X_2 en fonction de p, q, γ .

(ii) Exprimer le coefficient de corrélation de X_1 et X_2 en fonction de p, q, γ .

(iii) Trouver la loi conditionnelle de X_n sachant $S_{n-1} = k$.

(iv) Exprimer la loi de X_n en fonction de p, q, γ et $E(S_{n-1})$. En déduire la loi de X_n par récurrence.

3.13. Soit $X \sim \mathcal{U}_{[0,1]}$ indépendante de $Y \sim \mathcal{U}(0, 1, \dots, n)$. Trouver la loi de $X + Y$ et calculer de deux façons l'espérance de $X + Y$.

3.14. Soient X, Y deux variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$.

(i) Trouver la loi de $X + Y$, son espérance et sa variance.

(ii) Soit Z une variable aléatoire de même loi que $X + Y$. Calculer la loi de $1 - Y$ et déduire que $X - Y$ a la même loi que $Z - 1$. Quelle est la densité de $X - Y$?

(iii) Les variables $X + Y$ et $X - Y$ sont-elles indépendantes?

3.15. Soient X, Y deux variables aléatoires indépendantes dont les densités sont

$$f_X(x) = \frac{2}{\pi\sqrt{1-x^2}} \mathbb{1}_{]0,1[}(x), f_Y(x) = xe^{-\frac{x^2}{2}} \mathbb{1}_{]0,\infty[}.$$

Trouver la loi de XY .

3.16. (i) Soient X, Y deux variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$. On note $U = \cos \theta X - \sin \theta Y$ et $V = \sin \theta X + \cos \theta Y$. Trouver la loi de (U, V) et de ses marginales. U, V sont-elles indépendantes?

(ii) Soient R et Θ deux variables aléatoires indépendantes telles que $\Theta \sim \mathcal{U}_{[0, 2\pi]}$ et $R > 0$ avec $R^2 \sim \mathcal{E}(1/2)$. On note $X = R \cos \Theta$ et $Y = R \sin \Theta$. Trouver leurs lois et étudier leur indépendance.

3.17. Soit X une variable aléatoire réelle gaussienne centrée réduite et ε une autre variable aléatoire indépendante de X , prenant seulement les valeurs ± 1 .

(i) Montrer que $Y := \varepsilon X$ est une variable aléatoire réelle gaussienne centrée réduite.

(ii) Calculer $\text{Cov}(X, Y)$.

(iii) La variable aléatoire $X + Y$ est-elle gaussienne?

3.18. Montrer que le carré et le carré du module d'une fonction caractéristique sont aussi des fonctions caractéristiques.

3.19. Soient X et Y deux variables aléatoires indépendantes de même loi. (i) Montrer que si X et Y sont deux variables gaussiennes centrées réduites, alors $X + Y$ et $X - Y$ sont indépendantes.

(ii) *Théorème de Bernstein.* Réciproquement, on suppose que X et Y sont de carré intégrable et que $X + Y$ et $X - Y$ sont indépendantes. On veut montrer que X et Y sont deux variables gaussiennes. Pour cela :

(a) Montrer qu'on peut supposer que X et Y sont centrées, de variance 1.

(b) Montrer que φ , la fonction caractéristique commune de X et de Y , satisfait l'égalité $\varphi(2t) = \varphi(t)^3\varphi(-t)$. En déduire que φ ne s'annule nulle part.

(c) On pose $\psi(t) := \varphi(t)/\varphi(-t)$. Montrer que $\psi(2t) = \psi(t)^2$ et que $\psi(t) = 1 + o(t^2)$, lorsque $t \downarrow 0$. En déduire que, pour tout t , $\psi(t) = 1$ et que $\varphi(t) = \varphi(t/2)^4$. Conclure.

3.20. Un examen consiste en 20 questions auxquelles il faut répondre par oui ou par non; chaque réponse juste est notée 1 point et chaque réponse fausse, 0 points. Un étudiant répond entièrement au hasard et sa note finale est une variable aléatoire X .

(i) Soit X_j la note qu'il obtient à la j -ème question. Trouver la loi de X_j , son espérance, sa variance.

(ii) Exprimer X en fonction des X_j , donner sa loi, son espérance et sa variance.

(iii) Calculer la probabilité d'avoir une note inférieure à 5. Donner un majorant de cette probabilité à l'aide de l'inégalité de Bienaymé-Thébychev.

(iv) Un étudiant sérieux estime qu'il donnera une réponse exacte à chaque question avec une probabilité de 0,8. Quelle est la loi de sa note, l'espérance et la variance. Donner une minoration pour la probabilité qu'il ait plus de 12.

3.21. Soit (X_1, \dots, X_d) un vecteur aléatoire de loi

$$P((X_1, \dots, X_d) = (x_1, \dots, x_d)) = \prod_{j=1}^d p_j^{x_j} (1 - p_j)^{1-x_j},$$

où $p_1, \dots, p_d \in [0,1]$ et $x_1, \dots, x_d \in \{0,1\}$.

(i) Trouver les lois des X_j et étudier leur indépendance.

(ii) Soient $u_1, \dots, u_d \in \mathbb{R}^d$ des vecteurs de norme au plus 1. On note $Y = \sum_{j=1}^d X_j u_j$ et $w = \sum_{j=1}^d p_j u_j$. Calculer $E(\|Y - w\|^2)$ et montrer qu'on peut majorer cette quantité par $d/4$.

(iii) Endéduire qu'il existe une valeur de (X_1, \dots, X_d) pour laquelle $\|Y - w\|^2 \leq d/4$.

(iv) Montrer que dans l'espace euclidien \mathbb{R}^d , toute combinaison linéaire des vecteurs u_1, \dots, u_d de norme au plus 1, à coefficients dans $[0,1]$ peut-être approximée à $\sqrt{d}/2$ près par une combinaison linéaire à coefficients dans $\{0,1\}$.

Chapitre 4

Convergence des suites de variables aléatoires

Les différentes notions de convergence de variables aléatoires sont essentielles pour les applications. Voici un exemple : la fréquence observée du nombre de faces pile obtenu au cours d'un jeu de pile ou face, après n lancers est "proche" de la probabilité p d'obtenir pile, pourvu que n soit grand. Donc, si p est inconnue (on ne sait pas si la pièce est truquée), nous avons un moyen de l'approximer.

Dans tout ce chapitre les suites de variables aléatoires sont supposées construites sur un espace de probabilité (Ω, \mathcal{A}, P) . Pour simplifier on ne considère que des variables aléatoires réelles, mais les énoncés et les résultats sont vrais pour des vecteurs aléatoires à valeurs dans \mathbb{R}^d .

4.1 Convergence presque sûre

Définition 4.1 Une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement (p.s.) vers la variable aléatoire réelle X si

$$P\left(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

Dans ce cas on note $\lim_{n \rightarrow \infty} X_n = X$ p.s. ou $X_n \xrightarrow{p.s.} X$, lorsque $n \rightarrow \infty$.

REMARQUE: On voit que l'ensemble $\{\lim_{n \rightarrow \infty} X_n = X\}$ est bien un événement, car

$$\{\lim_{n \rightarrow \infty} X_n = X\} = \bigcap_{r \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \left\{ |X_n - X| < \frac{1}{r} \right\}.$$

On voit que pour une suite d'événements $(A_r)_{r \geq 1}$, $P(\bigcap_{r \geq 1} A_r) = 1$ si et seulement si $P(A_r) = 1$ pour tout r . Donc X_n converge vers X p.s. si et seulement si

$$\forall \varepsilon > 0, P\left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{|X_n - X| < \varepsilon\}\right) = 1,$$

ou, par passage au complémentaire,

$$\forall \varepsilon > 0, \mathbb{P} \left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - X| \geq \varepsilon\} \right) = 0,$$

ou, équivalent à

$$(*) \quad \forall \varepsilon > 0, \mathbb{P} (|X_n - X| \geq \varepsilon \text{ i.s.}) = 0.$$

Par convergence monotone, c'est encore équivalent à

$$(**) \quad \forall \varepsilon > 0, \lim_{m \rightarrow \infty} \mathbb{P} \left(\sup_{n \geq m} |X_n - X| \geq \varepsilon \right) = 0.$$

On peut encore caractériser la convergence presque sûre à l'aide du critère de Cauchy: X_n converge p.s. si et seulement si

$$\forall \varepsilon > 0, \mathbb{P} \left(\bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{|X_n - X_m| > \varepsilon\} \right) = 1.$$

Enfin on peut aussi dire que X_n converge vers X p.s. si X_n converge ponctuellement vers X , quitte à enlever un ensemble de mesure nulle (celui pour lequel $X_n(\omega)$ ne converge pas vers $X(\omega)$).

REMARQUE. Il est clair qu'on peut prendre des variables à valeurs dans \mathbb{R}^d . Il suffit alors de remplacer $|\cdot|$ par une norme quelconque $\|\cdot\|$ sur \mathbb{R}^d . De plus il est facile de voir que

$$(X_n^1, \dots, X_n^d) \xrightarrow{\text{p.s.}} (X^1, \dots, X^d) \Leftrightarrow X_n^j \xrightarrow{\text{p.s.}} X^j, j = 1, \dots, d.$$

Proposition 4.1 (lemme de Borel-Cantelli de convergence p.s.)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

- 1) Si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$, alors $X_n \xrightarrow{\text{p.s.}} X$.
- 2) Si les variables $(X_n)_{n \in \mathbb{N}}$ sont mutuellement indépendantes, alors $X_n \xrightarrow{\text{p.s.}} x$, $x \in \mathbb{R}$, si et seulement si, pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - x| \geq \varepsilon) < \infty$.

Preuve. Pour la première partie on considère $\varepsilon > 0$ et les événements

$$A_n = \{|X_n - X| \geq \varepsilon\}, n \in \mathbb{N}.$$

On applique le lemme de Borel-Cantelli 3.3 aux A_n . On déduit $\mathbb{P}(A_n \text{ i.s.}) = 0$ ce qui fournit le résultat d'après (*). La deuxième partie se démontre de façon analogue à partir de la partie indépendantes du lemme de Borel-Cantelli 3.3. Remarquer qu'il convient de supposer X nulle ou constante x , sans quoi les événements A_n ne sont pas nécessairement indépendants. \square

Exemple: Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi de

Bernoulli $\mathcal{B}(1,p)$. On montre que $Y_n = \sum_{j=1}^n 2^{-j} X_j$ converge presque sûrement en vérifiant le critère de Cauchy: pour, $n \geq m$, $|Y_n - Y_m| \leq \sum_{j=m+1}^n 2^{-j} \leq 2^{-m}$. Donc

$$\begin{aligned} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{\omega \in \Omega : |Y_n(\omega) - Y_m(\omega)| < \varepsilon\} &\supset \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{\omega \in \Omega : 2^{-m} < \varepsilon\} \\ &= \bigcup_{m \in \mathbb{N}} \{\omega \in \Omega : 2^{-m} < \varepsilon\} = \Omega \end{aligned}$$

Proposition 4.2 Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires réelles telles que X_n converge p.s. vers une variable aléatoire X et Y_n converge p.s. vers la variable aléatoire Y . Alors on peut voir que:

- 1) si $g : \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors $g(X_n)$ converge p.s. vers $g(X)$.
- 2) pour tous $\alpha, \beta \in \mathbb{R}$, $\alpha X_n + \beta Y_n$ converge p.s. vers $\alpha X + \beta Y$.
- 3) $X_n Y_n$ converge p.s. vers XY .

Preuve. Pour prouver 1), soit $C = \{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$. Alors, pour tout $\omega \in C$, $\lim_{n \rightarrow \infty} g(X_n(\omega)) = g(X(\omega))$ par la continuité de g . Ainsi $\{\omega \in \Omega : \lim_{n \rightarrow \infty} g(X_n(\omega)) = g(X(\omega))\} \supset C$, donc c'est un événement de probabilité 1. Pour les autres parties on peut utiliser la convergence presque sûre du vecteur (X_n, Y_n) et les fonctions continues $(x, y) \mapsto (\alpha x + \beta y, xy)$ et les projections. \square

4.2 Convergence en probabilité

Définition 4.2 Une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle X si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Dans ce cas on note $\lim_{n \rightarrow \infty} X_n = X$ en probabilité ou $X_n \xrightarrow{prob} X$, lorsque $n \rightarrow \infty$.

REMARQUE: Autrement dit, $\lim_{n \rightarrow \infty} X_n = X$ en probabilité si

$$\forall \varepsilon > 0, \forall \delta > 0 \exists n_0(\varepsilon, \delta) \text{ tel que pour } n \geq n_0, \mathbb{P}(|X_n - X| \geq \varepsilon) < \delta,$$

ou équivalent

$$\forall \varepsilon > 0, \exists n_0(\varepsilon) \text{ tel que pour } n \geq n_0, \mathbb{P}(|X_n - X| \geq \varepsilon) < \varepsilon.$$

Il est clair que la première implique la deuxième; pour voir l'autre implication, si $\varepsilon \leq \delta$ il suffit de prendre $n_0(\varepsilon, \delta) = n_0(\varepsilon)$; si $\delta < \varepsilon$, on prend $n_0(\varepsilon, \delta) = n_0(\delta)$ et par la deuxième on a alors $\mathbb{P}(|X_n - X| \geq \delta) \leq \delta$ et il suffit de voir que $\{|X_n - X| \geq \varepsilon\} \subset \{|X_n - X| \geq \delta\}$.

REMARQUE: A nouveau on peut prendre des vecteurs aléatoires dans \mathbb{R}^d avec une norme quelconque à la place de $|\cdot|$ et une fois de plus il y a équivalence avec la convergence des coordonnées grâce aux inégalités suivantes:

$$\mathbb{P}\left(\max_{1 \leq j \leq d} |X_n^j - X^j| \geq \varepsilon\right) \leq \mathbb{P}\left(\bigcup_{1 \leq j \leq d} \{|X_n^j - X^j| \geq \varepsilon\}\right) \leq \sum_{1 \leq j \leq d} \mathbb{P}(|X_n^j - X^j| \geq \varepsilon).$$

Proposition 4.3 La convergence presque sûre implique la convergence en probabilité.

Preuve. On voit la différence par rapport à la convergence p.s. qui exige un supremum (**). Plus précisément, soit A l'événement de probabilité 1 qui apparaît dans la définition de la convergence presque sûre. On fixe $\varepsilon > 0$ et pour chaque entier n on considère

$$B_n = \{\omega \in A : \sup_{m \geq n} |X_m(\omega) - X(\omega)| \geq \varepsilon\}.$$

La suite (B_n) est décroissante et $\cap_n B_n = \emptyset$, car X_n converge vers X sur A . On a donc, $\lim_{n \rightarrow \infty} P(B_n) = 0$. Comme

$$\{|X_n - X| \geq \varepsilon\} \subset B_n \cup A^c,$$

on tire la convergence en probabilité.

Une autre manière de prouver: si $X_n \xrightarrow{\text{p.s.}} X$, pour tout $\varepsilon > 0$,

$$\{\lim_{n \rightarrow \infty} X_n = X\} \subset \liminf_n \{|X_n - X| < \varepsilon\},$$

donc par le lemme de Fatou (ou exercice 1.14)

$$0 \leq \limsup_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) \leq P\left(\limsup_n \{|X_n - X| \geq \varepsilon\}\right) \leq 1 - P(\lim_{n \rightarrow \infty} X_n = X) = 0.$$

□

REMARQUE: La convergence en probabilité n'implique pas la convergence presque sûre, comme le montre l'exemple suivant: soit l'espace de probabilité $([0,1], \mathcal{B}([0,1]), \lambda)$ et on définit

$$Y_{2^n, j} = \mathbb{1}_{\left] \frac{j}{2^n}, \frac{j+1}{2^n} \right]}(\omega), 0 \leq j \leq 2^n - 1, n \geq 1$$

et

$$X_{2^n + j} = Y_{2^n, j}, 0 \leq j \leq 2^n - 1, n \geq 1.$$

Alors, pour tout $\omega \in [0,1]$, $\liminf_{n \rightarrow \infty} X_n(\omega) = 0$ et $\limsup_{n \rightarrow \infty} X_n(\omega) = 1$, donc X_n ne converge pas presque sûrement. Cependant, pour tout $\varepsilon \in]0,1[$ et $m = 2^n + j$, $0 \leq j \leq 2^n - 1$, $P(|X_m| \geq \varepsilon) = 2^{-n}$. Donc X_n converge en probabilité vers 0.

Exemples: i) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles non-corrélées, centrées, telles que $\text{Var}(X_n) = \sigma^2$. Alors leurs moyennes partielles $\frac{1}{n} \sum_{1 \leq j \leq n} X_j$ convergent en probabilité vers 0. En effet, par l'inégalité de Tchebychev,

$$P\left(\left|\frac{1}{n} \sum_{1 \leq j \leq n} X_j\right| \geq \varepsilon\right) \leq \frac{1}{n^2 \varepsilon^2} \text{Var}\left(\sum_{1 \leq j \leq n} X_j\right) = \frac{\sigma^2}{n \varepsilon^2}$$

De point de vue modélisation, il suffit de lancer une cinquantaine de fois une pièce non truquée pour voir que la proportion de pile se stabilise vers 0,5.

ii) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes telle que $X_n \sim \mathcal{B}(n, p_n)$. Alors

$$X_n \xrightarrow{\text{prob}} 0 \iff \lim_{n \rightarrow \infty} p_n = 0.$$

Mais par le lemme de Borel-Cantelli (Proposition 4.1) on a

$$X_n \xrightarrow{\text{p.s.}} 0 \iff \sum_n P(|X_n| \geq \varepsilon) < \infty, \forall \varepsilon > 0,$$

ce qui est équivalent à $\sum_n p_n < \infty$, puisque $P(|X_n| \geq \varepsilon) = p_n$ si $0 < \varepsilon < 1$.

REMARQUE: Il est possible de définir une métrique sur $L^0(\Omega, \mathcal{A}, P)$ définissant la convergence en probabilité des variables aléatoires:

$$d(X, Y) := E(|X - Y| \wedge 1), X, Y \in L^0(\Omega, \mathcal{A}, P).$$

En effet, $d(X, Y) = 0$ si et seulement si $X = Y$ p.s. et on vérifie que $d(X, Y) \geq 0$ et que $d(X, Z) \leq d(X, Y) + d(Y, Z)$.

Proposition 4.4 *La suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si et seulement si*

$$\lim_{n \rightarrow \infty} d(X_n, X) = 0.$$

Preuve: Par l'inégalité de Markov, pour tout $\varepsilon \in]0, 1]$,

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X| \wedge 1 \geq \varepsilon) \leq \frac{d(X_n, X)}{\varepsilon}.$$

Par ailleurs,

$$\begin{aligned} d(X_n, X) &= E[(|X_n - X| \wedge 1) \mathbf{1}_{\{|X_n - X| \geq \varepsilon\}}] + E[(|X_n - X| \wedge 1) \mathbf{1}_{\{|X_n - X| < \varepsilon\}}] \\ &\leq P(|X_n - X| \geq \varepsilon) + \varepsilon. \end{aligned}$$

Si $X_n \xrightarrow{\text{prob}} X$, 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 $n \geq n_0$, d'où la conclusion. \square

REMARQUE: Dans tout ce qui précède on pourrait aussi utiliser la distance

$$d'(X, Y) = E\left(\frac{|X - Y|}{1 + |X - Y|}\right).$$

REMARQUE: Par unicité de la limite dans tout espace métrique on déduit que la limite en probabilité est unique p.s.

Proposition 4.5 *Soit X et $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires réelles. Alors $X_n \xrightarrow{\text{prob}} X$ si et seulement si de toute suite croissante d'entiers (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k} \xrightarrow{\text{p.s.}} X$.*

Preuve:

“ \implies ”: par la convergence en probabilité, pour tout $k \in \mathbb{N}^*$, soit n'_k le plus petit entier tel que

$$P(|X_{n'_k} - X| \geq \frac{1}{k}) < \frac{1}{2^k}.$$

Alors,

$$\sum_{k \in \mathbb{N}^*} P(|X_{n'_k} - X| \geq \frac{1}{k}) < \infty.$$

En particulier, pour tout $\varepsilon > 0$, il existe un k tel que $P(|X_{n'_k} - X| \geq \varepsilon) \leq P(|X_{n'_k} - X| \geq \frac{1}{k})$, d'où

$$\sum_{k \in \mathbb{N}^*} P(|X_{n'_k} - X| \geq \varepsilon) < \infty,$$

et donc, par le lemme de Borel-Cantelli (Proposition 4.1), $\lim_{k \rightarrow \infty} X_{n'_k} = X$ p.s.

“ \Leftarrow ” : supposons le contraire, c'est-à-dire que X_n ne converge pas vers X en probabilité. Alors il existe une suite croissante (n') , il existe $\delta > 0$ et $\varepsilon > 0$ tels que $P(|X_{n'} - X| \geq \varepsilon) \geq \delta$. Mais, par l'hypothèse, on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers X , et par la Proposition 4.3 en probabilité. Mais ceci est absurde car $P(|X_{n'_k} - X| \geq \varepsilon) \geq \delta$. \square

REMARQUE: De cette proposition on voit que la convergence presque sûre n'est pas une convergence métrique, car si c'était le cas, elle coïnciderait avec la convergence en probabilité.

Proposition 4.6 Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires réelles telles que X_n converge en probabilité vers une variable aléatoire X et Y_n converge en probabilité vers la variable aléatoire Y .

- 1) Si $g : \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors $g(X_n)$ converge en probabilité vers $g(X)$.
- 2) Pour tous $\alpha, \beta \in \mathbb{R}$, $\alpha X_n + \beta Y_n$ converge en probabilité vers $\alpha X + \beta Y$.
- 3) De plus $X_n Y_n$ converge en probabilité vers XY .

Preuve: Prouvons 3): soit (n') une suite extraite croissante et on extrait une sous-suite (n'') telle que $X_{n''} \rightarrow X$ p.s. De (n'') on peut extraire une sous-suite (n''') telle que $Y_{n'''} \rightarrow Y$ p.s. Alors $X_{n'''} Y_{n'''} \rightarrow XY$ p.s. On conclut à l'aide des Proposition 4.2 et 4.5. \square

Proposition 4.7 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles telle qu'elle vérifie le critère de Cauchy en probabilité, c'est-à-dire que

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 \quad P(|X_n - X_m| \geq \varepsilon) < \varepsilon,$$

ou de façon équivalente, que

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq m \geq n_0 \quad d(X_n, X_m) < \varepsilon.$$

Alors X_n converge en probabilité si et seulement si X_n est de Cauchy en probabilité. Autrement dit, l'espace $L^0(\Omega, \mathcal{A}, P)$ est complet pour la distance d métrisant la convergence en probabilité.

Preuve: Supposons que X_n est convergente en probabilité vers X . Ainsi, pour tout $\varepsilon > 0$, il existe n_0 tel que dès que $n \geq n_0$ on ait $P(|X_n - X| \geq \varepsilon) < \varepsilon$. Par l'inégalité triangulaire on peut écrire, pour tout $n \geq m \geq n_0$

$$\{|X_n - X_m| \geq \varepsilon\} \subset \{|X_n - X| \geq \varepsilon\} \cup \{|X_m - X| \geq \varepsilon\}.$$

D'ici on déduit le critère de Cauchy en probabilité. Réciproquement, soit $\varepsilon = 2^{-k}$ dans la condition de Cauchy en probabilité. On pose $n_1 = 1$ et soit n_k le plus petit entier tel que $P(|X_n - X_m| \geq 2^{-k}) < 2^{-k}$:

$$n_k = \min\{N > n_{k-1} : P(|X_n - X_m| \geq 2^{-k}) < 2^{-k} \forall n, m \geq N\}$$

Ainsi $n_k \leq n_{k+1}$ et

$$P(|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}) < 2^{-k}.$$

Par le lemme de Borel-Cantelli 3.3, pour presque tout ω , il existe un entier $k(\omega) < \infty$ tel que si $m \geq k(\omega)$, $|X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leq 2^{-m}$. Alors la suite de réels $X_{n_m}(\omega)$ est de Cauchy. En effet, soit $\varepsilon > 0$ et $p > l$; on a

$$|X_{n_p}(\omega) - X_{n_l}(\omega)| \leq \sum_{l \leq m \leq p-1} |X_{n_{m+1}}(\omega) - X_{n_m}(\omega)| \leq \sum_{l \leq m \leq p-1} 2^{-m} \leq \sum_{m \geq l} 2^{-m} \leq 2^{-l+1}.$$

Puisque pour tout $p > l > l_0$, $2^{-l+1} \leq \varepsilon$, donc $|X_{n_p}(\omega) - X_{n_l}(\omega)| \leq \varepsilon$. Ainsi X_{n_k} converge p.s. vers une limite X et en particulier cette sous-suite converge en probabilité vers X . Pour conclure on écrit :

$$P(|X_n - X| \geq \varepsilon) \leq P\left(|X_n - X_{n_k}| \geq \frac{\varepsilon}{2}\right) + P\left(|X_{n_k} - X| \geq \frac{\varepsilon}{2}\right)$$

et on fait petit chaque terme en utilisant la condition de Cauchy et la précédente convergence en proba de la sous-suite X_{n_k} . \square

4.3 Convergence dans L^p

Définition 4.3 Une suite de variables aléatoires réelles p -intégrables, $0 < p < \infty$, $(X_n)_{n \in \mathbb{N}}$ converge dans L^p vers une variable p -intégrable X si

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0 \Leftrightarrow \lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0.$$

Dans ce cas on note $X_n \xrightarrow{L^p} X$, lorsque $n \rightarrow \infty$.

Proposition 4.8 La convergence dans L^p implique la convergence en probabilité.

Preuve. Par l'inégalité de Markov, pour tout $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon) \leq \frac{E(|X_n - X|^p)}{\varepsilon^p}, p > 0,$$

d'où la conclusion. \square

REMARQUE: La Proposition 4.4 justifie la terminologie pour la convergence L^0 pour la convergence en probabilité.

REMARQUE: La convergence en probabilité n'implique pas la convergence L^p comme le montre l'exemple suivant: soit l'espace de probabilité $(]0,1[, \mathcal{B}(]0,1]), \lambda)$, soit $\alpha > 0$ et on considère

$$X_n(\omega) = \omega^{-\alpha} \mathbb{1}_{]-\infty, \frac{1}{n}]}(\omega), n \geq 1.$$

Pour tout $\varepsilon \in]0,1[$, $P(|X_n| \geq \varepsilon) = \frac{1}{n}$, d'où $X_n \rightarrow 0$ en probabilité. Or, $X_n \notin L^p$, dès que $\alpha p \geq 1$ car

$$E(X_n^p) = \int_0^{\frac{1}{n}} \omega^{-\alpha p} d\omega.$$

REMARQUE: La convergence en presque sûre n'implique pas la convergence L^p comme le montre l'exemple suivant: soit l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, avec \mathbb{P} une probabilité, et soit X_n de loi $(1 - n^{-p})\delta_0 + n^{-p}\delta_n$, $p > 1$. Soit $\varepsilon > 0$; pour $n \geq \varepsilon^{-p}$, $\mathbb{P}(|X_n| \geq \varepsilon) = n^{-p}$. Par le lemme de Borel-Cantelli (Proposition 4.1) X_n converge presque sûrement vers 0, puisque $p > 1$. Toutefois, $\mathbb{E}(|X_n|^p) = 1$, pour tout n .

Définition 4.4 Une suite de variables aléatoires intégrables $(X_n)_{n \in \mathbb{N}}$ est dite **uniformément intégrable** si

$$\lim_{c \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > c\}}) = 0.$$

REMARQUE: Une famille de variables aléatoires avec un seul élément est uniformément intégrable par convergence dominée.

Une suite de variables aléatoires dominée par une variable Y intégrable, c'est-à-dire, si $|X_n| \leq Y$ pour tout n , alors la suite est uniformément intégrable:

$$\lim_{c \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > c\}}) \leq \lim_{c \rightarrow \infty} \mathbb{E}(Y \mathbf{1}_{\{Y > c\}}) = 0.$$

Enfin, toute suite finie de variables aléatoires intégrables est uniformément intégrable car elle est dominée par la variable intégrable $\sum_{1 \leq j \leq n} |X_j|$.

Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables intégrables telles que $|X_n| \leq |Y_n|$. Alors si la suite Y_n est uniformément intégrable, alors la suite X_n est uniformément intégrable.

Si les variables X_n sont p -intégrables alors la suite $(|X_n|^p)_{n \in \mathbb{N}}$ est uniformément intégrable si cette suite est **bornée dans $L^{p+\delta}$** :

$$\sup_n \mathbb{E}(|X_n|^{p+\delta}) < \infty.$$

En effet

$$\begin{aligned} \sup_n \mathbb{E}(|X_n|^p \mathbf{1}_{\{|X_n|^p > c\}}) &= \sup_n \mathbb{E}\left(|X_n|^p \mathbf{1}_{\left\{\left|\frac{X_n}{c^{1/p}}\right| > 1\right\}}\right) = \sup_n \mathbb{E}\left(|X_n|^p \cdot 1 \cdot \mathbf{1}_{\left\{\frac{|X_n|^\delta}{c^{\delta/p}} > 1\right\}}\right) \\ &\leq \sup_n \mathbb{E}\left(|X_n|^p \frac{|X_n|^\delta}{c^{\delta/p}}\right) \leq c^{-\delta/p} \sup_n \mathbb{E}(|X_n|^{p+\delta}). \end{aligned}$$

REMARQUE: Montrons que:

$$X \text{ intégrable} \implies \forall \varepsilon > 0, \exists \eta > 0, \text{ t.q. si } A \in \mathcal{A} \text{ avec } \mathbb{P}(A) < \eta \text{ alors } \mathbb{E}(|X| \mathbf{1}_A) \leq \varepsilon.$$

D'une part, par le théorème de convergence dominée, pour c suffisamment grand,

$$\mathbb{E}(|X| \mathbf{1}_{\{|X| > c\}}) \leq \frac{\varepsilon}{2}.$$

D'autre part, pour $\eta \leq \frac{\varepsilon}{2c}$,

$$\mathbb{E}(|X| \mathbf{1}_A) = \mathbb{E}(|X| \mathbf{1}_{A \cap \{|X| > c\}}) + \mathbb{E}(|X| \mathbf{1}_{A \cap \{|X| \leq c\}}) \leq \frac{\varepsilon}{2} + c\mathbb{P}(A) \leq \frac{\varepsilon}{2} + c\eta \leq \varepsilon.$$

Proposition 4.9 La suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles intégrables est uniformément intégrable si et seulement si:

$$a) \forall \varepsilon > 0, \exists \eta > 0, \text{ t.q. si } A \in \mathcal{A} \text{ avec } \mathbb{P}(A) < \eta \text{ alors } \forall n \in \mathbb{N} \mathbb{E}(|X_n| \mathbf{1}_A) \leq \varepsilon$$

et

$$b) \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|) < \infty.$$

Preuve: Supposons l'uniforme intégrabilité:

$$\forall \varepsilon > 0, \exists c > 0 \text{ t.q. } \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > c\}}) \leq \frac{\varepsilon}{2}.$$

Pour $A \in \mathcal{A}$, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(|X_n|) = \mathbb{E}(|X_n| \mathbf{1}_{A \cap \{|X_n| > c\}}) + \mathbb{E}(|X_n| \mathbf{1}_{A \cap \{|X_n| \leq c\}}) \leq \frac{\varepsilon}{2} + cP(A).$$

On obtient $a)$ en prenant $\eta = \frac{\varepsilon}{2c}$ et $b)$ en prenant $A = \Omega$. Réciproquement, on fixe $\varepsilon > 0$ et $\eta > 0$ fournis par $a)$, on note $M = \sup_{n \in \mathbb{N}} \mathbb{E}(|X_n|)$ qui est fini par $b)$ et on note $c_0 = \frac{M}{\eta}$. Par l'inégalité de Markov, pour tout $c \geq c_0$ et tout n , $P(|X_n| > c) \leq \eta$. On applique $a)$ à $A = \{|X_n| > c\}$ pour chaque n et on trouve $\mathbb{E}(|X_n| \mathbf{1}_{\{|X_n| > c\}}) \leq \varepsilon$, d'où la conclusion. \square

Proposition 4.10 *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires intégrables. Il y a équivalence entre:*

- 1) $(X_n)_{n \in \mathbb{N}}$ converge dans L^1 ;
- 2) $(X_n)_{n \in \mathbb{N}}$ est une suite de Cauchy dans L^1 , c'est-à-dire $\mathbb{E}(|X_n - X_m|) \rightarrow 0$, lorsque $n, m \rightarrow \infty$;
- 3) $(X_n)_{n \in \mathbb{N}}$ est une suite uniformément intégrable et $(X_n)_{n \in \mathbb{N}}$ converge en probabilité.

REMARQUE: Par ce résultat on voit que l'uniforme intégrabilité et la convergence en probabilité (ou la convergence p.s.) impliquent la convergence dans L^1 . En particulier comme $|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}(X_n - X)$, on déduit la convergence de la suite premier moment. On a ainsi, pour une suite dominée, un résultat de convergence dominée pour la convergence en probabilité. Enfin cette proposition montre que L^1 est un espace complet (voir Proposition 2.15).

Preuve:

1) \implies 2):

La convergence dans L^1 implique la condition de Cauchy dans L^1 par l'inégalité triangulaire.

2) \implies 3):

Par la condition de Cauchy, pour tout $\varepsilon > 0$, il existe $n_0(\varepsilon)$ tel que si $n \geq m \geq n_0$, $\mathbb{E}(|X_n - X_m|) \leq \frac{\varepsilon}{2}$. Alors, pour tout $A \in \mathcal{A}$ et tout $n \geq n_0$

$$\mathbb{E}(|X_n| \mathbf{1}_A) = \mathbb{E}(|X_n - X_{n_0} + X_{n_0}| \mathbf{1}_A) \leq \mathbb{E}(|X_{n_0}| \mathbf{1}_A) + \mathbb{E}(|X_n - X_{n_0}|) \leq \mathbb{E}(|X_{n_0}| \mathbf{1}_A) + \frac{\varepsilon}{2}.$$

Donc

$$\sup_{n \geq n_0} \mathbb{E}(|X_n| \mathbf{1}_A) \leq \mathbb{E}(|X_{n_0}| \mathbf{1}_A) + \frac{\varepsilon}{2}, \text{ d'où } \sup_{n \geq n_0} \mathbb{E}(|X_n| \mathbf{1}_A) \leq \sup_{j \leq n_0} \mathbb{E}(|X_j| \mathbf{1}_A) + \frac{\varepsilon}{2}.$$

Si $A = \Omega$,

$$\sup_{n \geq n_0} \mathbb{E}(|X_n|) \leq \sup_{j \leq n_0} \mathbb{E}(|X_j|) + \frac{\varepsilon}{2} < \infty,$$

donc la partie *b*) de la Proposition 4.9 est vérifiée. Par ailleurs la famille finie $\{X_j : j \leq n_0\}$ est uniformément intégrable, donc, pour tout $\varepsilon > 0$, il existe un $\eta > 0$ tel que si $P(A) < \eta$, on ait $\sup_{j \leq n_0} E(|X_j| \mathbb{1}_A) \leq \frac{\varepsilon}{2}$. On déduit alors

$$\sup_n E(|X_n| \mathbb{1}_A) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

c'est-à-dire la partie *a*) de la Proposition 4.9. Donc la suite est uniformément intégrable. Enfin, par l'inégalité de Markov

$$P(|X_n - X_m| \geq \varepsilon) \leq \frac{E(|X_n - X_m|)}{\varepsilon} \rightarrow 0, n, m \rightarrow \infty,$$

donc la suite est de Cauchy en probabilité. Par la Proposition 4.7 la suite est convergente en probabilité.

3) \implies 1):

Comme X_n converge en probabilité vers une variable X , de toute suite croissante (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers X (Proposition 4.5). Alors, par le lemme de Fatou

$$E(|X|) = E(\liminf_{k \rightarrow \infty} |X_{n'_k}|) \leq \liminf_{k \rightarrow \infty} E(|X_{n'_k}|) \leq \sup_{n \in \mathbb{N}} E(|X_n|) < \infty,$$

par le point *b*) de la Proposition 4.9, donc X est intégrable.

Par ailleurs, pour tout $\varepsilon > 0$,

$$\begin{aligned} E(|X_n - X|) &\leq E(|X_n - X| \mathbb{1}_{\{|X_n - X| < \varepsilon\}}) + E(|X_n| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}) + E(|X| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}) \\ &\leq \varepsilon + E(|X_n| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}) + E(|X| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}). \end{aligned}$$

La suite de variables aléatoires intégrables $\{X, X_1, X_2, \dots, X_n, \dots\}$ est encore uniformément intégrable. On fixe $\eta > 0$ et par la convergence en probabilité de la suite X_n , pour n suffisamment grand on a $P(|X_n - X| \geq \varepsilon) \leq \eta$. En prenant dans le *a*) de la Proposition 4.9 $A = \{|X_n - X| \geq \varepsilon\}$, on voit que pour tout n suffisamment grand

$$E(|X_n| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}) \leq \varepsilon \text{ et } E(|X| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}) \leq \varepsilon.$$

On a obtenu $E(|X_n - X|) \leq 3\varepsilon$, d'où la convergence dans L^1 . \square

Proposition 4.11 *Soit $p \geq 1$ et soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires dans L^p . Il y a équivalence entre:*

- 1) $(X_n)_{n \in \mathbb{N}}$ converge dans L^p ;
- 2) $(X_n)_{n \in \mathbb{N}}$ est une suite de Cauchy dans L^p , c'est-à-dire $\|X_n - X_m\|_p \rightarrow 0$, lorsque $n, m \rightarrow \infty$;
- 3) $(|X_n|^p)_{n \in \mathbb{N}}$ est une suite uniformément intégrable et $(X_n)_{n \in \mathbb{N}}$ converge en probabilité.

REMARQUE: Par ce résultat la preuve de la Proposition 2.15 est complète. Par ailleurs, pour assurer l'uniforme intégrabilité de $(|X_n|^p)_{n \in \mathbb{N}}$ il suffit de vérifier que la suite $(X_n)_{n \in \mathbb{N}}$ est bornée dans $L^{p+\delta}$, $\delta > 0$.

Preuve. Cette preuve est similaire à celle de la Proposition 4.10. On donne les idées.

1) \implies 2):

La condition de Cauchy dans L^p s'obtient de la convergence dans L^p par l'inégalité de Minkowski (Proposition 2.14).

2) \implies 3):

Si $(X_n)_{n \in \mathbb{N}}$ est de Cauchy dans L^p elle est de Cauchy en probabilité (par l'inégalité de Markov), donc convergente en probabilité vers une variable aléatoire X . A nouveau par l'inégalité de Minkowski

$$\| \|X_n\|_p - \|X_m\|_p \| \leq \|X_n - X_m\|_p \rightarrow 0, n, m \rightarrow \infty.$$

Donc la suite réelle $(\|X_n\|_p)_{n \in \mathbb{N}}$ est de Cauchy, donc convergente, donc $\sup_n \|X_n\|_p < \infty$ (condition *b*) dans la Proposition 4.9). De plus par le lemme de Fatou on peut déduire aussi que la limite en probabilité satisfait aussi $\|X\|_p < \infty$, donc $X \in L^p$. Enfin

$$E(|X_n|^p \mathbf{1}_A) = E(|X_n - X_m + X_m|^p \mathbf{1}_A) \leq 2^p \|X_n - X_m\|_p^p + 2^p E(|X_m|^p \mathbf{1}_A).$$

Par la condition de Cauchy on fixe n_0 et on prend $n \geq n_0$ et par l'uniforme intégrabilité de $\{|X_{n_0}|\}$ on peut faire petite la quantité $2^p E(|X_{n_0}|^p \mathbf{1}_A)$, dès que $P(A)$ est petite. On trouve ainsi la condition *a*) dans la Proposition 4.9 et donc l'uniforme intégrabilité.

3) \implies 1):

De toute suite croissante (n') on peut extraire une sous-suite (n'_k) telle que $X_{n'_k}$ converge p.s. vers une variable X . Par l'uniforme intégrabilité et le lemme de Fatou

$$E(|X|^p) \leq \liminf_{k \rightarrow \infty} E(|X_{n'_k}|^p) \sup_n E(|X_n|^p) < \infty,$$

donc $X \in L^p$. Ensuite on fait comme dans la proposition précédente. □

4.4 Convergence en loi

On a vu que deux variables aléatoires X, Y ont la même loi lorsque $P_X = P_Y$, ou si et seulement si $F_X = F_Y$, ou si et seulement si, pour toute fonction continue bornée $g : \mathbb{R} \rightarrow \mathbb{R}$, $E[g(X)] = E[g(Y)]$, ou si et seulement si $\varphi_X = \varphi_Y$. Ces diverses égalités donnent lieu à diverses définitions de la convergence en loi.

Définition 4.5 (et théorème)

Une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ **converge en loi** vers X ou que la suite de lois $(P_{X_n})_{n \in \mathbb{N}}$ **converge étroitement** vers P_X si, lorsque $n \rightarrow \infty$:

- 1) $E[g(X_n)] \rightarrow E[g(X)] \iff \int_{\mathbb{R}} g(x) P_{X_n}(dx) \rightarrow \int_{\mathbb{R}} g(x) P_X(dx)$ pour toute fonction continue bornée $g : \mathbb{R} \rightarrow \mathbb{R}$;
- 2) $F_{X_n}(t) \rightarrow F_X(t)$ en tout point de continuité t de F_X ;
- 3) $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ pour tout $t \in \mathbb{R}$.

On note alors

$$X_n \xrightarrow{\text{loi}} X \iff P_{X_n} \xrightarrow{e} P_X$$

Proposition 4.12 La convergence presque sûre implique la convergence en loi.

Preuve. Ceci est une conséquence du théorème de convergence dominée et de la partie 1) de la définition. \square

Proposition 4.13 *La convergence en probabilité implique la convergence en loi.*

Preuve. Soit g continue et bornée. Alors si $g(X_n)$ converge en proba vers $g(X)$ si X_n converge en proba vers X . De plus la suite $(g(X_n))_{n \in \mathbb{N}}$ est bornée donc u.i. D'ici on déduit la convergence dans L^1 de $g(X_n)$ vers $g(X)$ et enfin on obtient 1). \square

Proposition 4.14 *La convergence dans L^p implique la convergence en loi.*

Preuve. On rappelle que la convergence dans L^p , entraîne la convergence en probabilité. Donc si X_n converge dans L^p , $p > 0$, vers X , alors X_n converge en loi vers X . \square

REMARQUE: La convergence en loi n'implique ni la convergence presque sûre, ni la convergence en probabilité, comme le montre l'exemple suivant: soit $X \sim \mathcal{N}(0,1)$ et $X_n = (-1)^n X$. Alors, par la symétrie de la loi gaussienne standard, X_n à même loi que X . Donc X_n converge en loi vers X , mais ne converge ni presque sûrement vers X , ni en probabilité vers X .

Proposition 4.15 *Si la suite X_n converge en loi vers une constante c si et seulement si X_n converge en probabilité vers c .*

Preuve: Par 2) on a que, lorsque $n \rightarrow \infty$,

$$F_{X_n}(t) \rightarrow \begin{cases} 0 & \text{si } t < c \\ 1 & \text{si } t \geq c, \end{cases} \quad \forall t \neq c.$$

Pour tout $\varepsilon > 0$, on a

$$P(|X_n - c| \leq \varepsilon) = P(c - \varepsilon \leq X_n \leq c + \varepsilon)$$

$$\geq P(c - \varepsilon < X_n \leq c + \varepsilon) = F_{X_n}(c + \varepsilon) - F_{X_n}(c - \varepsilon) \rightarrow 1 - 0 = 1$$

car $c + \varepsilon, c - \varepsilon \neq c$, d'où la conclusion. \square

REMARQUE: Si X_n et Y_n convergent en loi vers X et Y , on ne peut rien dire sur la convergence du couple (X_n, Y_n) . Dans l'exemple ci-dessus avec les lois gaussiennes, le couple (X, X_n) , ne converge pas en loi. Toutefois on peut montrer que si Y_n converge en loi vers une constante c , alors (X_n, Y_n) converge en loi vers (X, c) et en particulier $X_n + Y_n$ converge en loi vers $X + c$ et $X_n Y_n$ converge en loi vers cX .

REMARQUE: (**convergence en loi des variables aléatoires à valeurs entières**)

Soient $X_n, n \in \mathbb{N}$, et X à valeurs entières. Alors, lorsque $n \rightarrow \infty$

$$X_n \xrightarrow{\text{loi}} X \stackrel{(i)}{\iff} P(X_n = k) \longrightarrow P(X = k), \forall k \in \mathbb{N} \stackrel{(ii)}{\iff} G_{X_n}(s) \longrightarrow G_X(s) \text{ pour } s \in [0,1].$$

$\stackrel{(i)}{\iff}$: en effet, pour $k \in \mathbb{N}$, soient s, t des points de continuité de F_X , tels que $k - 1 < s < k < t < k + 1$, d'où, lorsque $n \rightarrow \infty$,

$$P(X_n = k) = F_{X_n}(t) - F_{X_n}(s) \longrightarrow F_X(t) - F_X(s) = P(X = k).$$

$\xleftarrow{(i)}$: soit g une fonction continue à support compact dans $[0, N]$. On a

$$\mathbb{E}[g(X_n)] = \sum_{k=0}^N g(k)P(X_n = k) \rightarrow \sum_{k=0}^N g(k)P(X = k) = \mathbb{E}[g(X)]$$

$\xrightarrow{(ii)}$: on utilise le même argument que ci-dessus en ajoutant la convergence dominée.

$\xleftarrow{(ii)}$: on a besoin d'un résultat classique sur la convergence des fonctions holomorphes:

Soit une suite des fonctions holomorphes sur le disque unité $(f_n)_{n \in \mathbb{N}}$, ayant, pour chaque n , tous les coefficients de chaque série entière positifs, et qui converge simplement sur $[0, 1]$; alors la suite converge simplement sur tout le disque unité et la limite f est une fonction holomorphe sur le disque unité; de plus pour tout entier k , la k -ième dérivée de f_n converge simplement vers la k -ième dérivée de f (sur le disque ouvert).

On prend ici $f_n = G_{X_n}$ dont (la limite de) la k -ième dérivée au point 1 est $(k!)P(X_n = k)$. On obtient la conclusion.

Preuve du Théorème 4.5:

1) \implies 2):

Pour tout $\varepsilon > 0$ il existe une fonction continue, $g_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ telle que $g_\varepsilon(x) = 1$ si $x \leq t$ et $g_\varepsilon(x) = 0$ si $x \geq t + \varepsilon$. On a

$$\lim_{n \rightarrow \infty} \mathbb{E}[g_\varepsilon(X_n)] = \mathbb{E}[g_\varepsilon(X)] \leq \mathbb{E}[\mathbb{1}_{]-\infty, t+\varepsilon]}(X) = P(X \leq t + \varepsilon) = F_X(t + \varepsilon)$$

et

$$\mathbb{E}[g_\varepsilon(X_n)] \geq \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X_n) = P(X_n \leq t) = F_{X_n}(t),$$

donc $\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t + \varepsilon)$ et par continuité à droite, $\limsup_{n \rightarrow \infty} F_{X_n}(t) \leq F_X(t)$ en prenant ε arbitrairement petit. Soit ensuite $h_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ telle que $h_\varepsilon(x) = 1$ si $x \leq t - \varepsilon$ et $h_\varepsilon(x) = 0$ si $x \geq t$. On a

$$\lim_{n \rightarrow \infty} \mathbb{E}[h_\varepsilon(X_n)] = \mathbb{E}[h_\varepsilon(X)] \geq \mathbb{E}[\mathbb{1}_{]-\infty, t-\varepsilon]}(X) = P(X \leq t - \varepsilon) = F_X(t - \varepsilon)$$

et

$$\mathbb{E}[h_\varepsilon(X_n)] \leq \mathbb{E}[\mathbb{1}_{]-\infty, t]}(X_n) = P(X_n \leq t) = F_{X_n}(t),$$

donc $\liminf_{n \rightarrow \infty} F_{X_n}(t) \geq F_X(t - \varepsilon)$. Par hypothèse $F_X(t-) = F_X(t)$ et pour ε arbitrairement petit on trouve $\liminf_{n \rightarrow \infty} F_{X_n}(t) \geq F_X(t)$.

2) \implies 1):

Pas a)

Soit d'abord $g : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^1 à support compact; en particulier g' est bornée à support compact. Par le théorème de Fubini, on a

$$\mathbb{E}[g(X_n)] = \int_{\mathbb{R}} P_{X_n}(dx) \int_{-\infty}^x g'(y)dy = \int_{\mathbb{R}} (1 - F_{X_n}(y))g'(y)dy.$$

On sait que l'ensemble des points de discontinuité de F_X est au plus dénombrable, donc de mesure de Lebesgue nulle; hors de ces points $1 - F_{X_n}$ converge vers $1 - F_X$ tout en restant bornée par 1, car toute fonction de répartition est à valeurs dans $[0, 1]$. La mesure positive $|g'(y)|dy$ a sa masse totale finie, c'est-à-dire:

$$\int_{\mathbb{R}} |g'(y)|dy \leq \sup_{y \in \text{supp}(g')} |g'(y)| \times \lambda_{Leb}(\text{supp}(g')) < \infty,$$

donc par convergence dominée,

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(X_n)] = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} (1 - F_{X_n}(y))g'(y)dy = \int_{\mathbb{R}} (1 - F_X(y))g'(y)dy = \mathbb{E}[g(X)].$$

Pas b)

Soit ensuite g continue à support compact, mais pas nécessairement dérivable. Pour tout $\varepsilon > 0$, il existe une fonction h de classe C^1 à support compact telle que $|g(x) - h(x)| \leq \frac{\varepsilon}{4}$, pour tout $t \in \mathbb{R}$ (par exemple la convolée de g par l'approximation de l'unité de classe C^1 à support compact). Par ce qu'on vient de voir, on sait trouver un entier n_0 , tel que $|\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| \leq \frac{\varepsilon}{2}$ dès que $n \geq n_0$. Donc, pour $n \geq n_0$,

$$|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| \leq \mathbb{E}(|g(X_n) - h(X_n)|) + |\mathbb{E}[h(X_n)] - \mathbb{E}[h(X)]| + \mathbb{E}(|g(X) - h(X)|) \leq \varepsilon.$$

Pas c)

Enfin soit g continue bornée et $\varepsilon > 0$. On peut supposer que $|g| \leq 1$. Il existe une fonction \check{g} continue à support compact $0 \leq \check{g} \leq 1$, telle que $\mathbb{E}[\check{g}(X)] > 1 - \frac{\varepsilon}{5}$. On déduit qu'il existe un entier n_1 tel que $\mathbb{E}[\check{g}(X_n)] > 1 - \frac{\varepsilon}{4}$ dès que $n \geq n_1$. La fonction $g\check{g}$ est continue à support compact, donc il existe un entier n_2 tel que $|\mathbb{E}[(g\check{g})(X_n)] - \mathbb{E}[(g\check{g})(X)]| \leq \frac{\varepsilon}{2}$ dès que $n \geq n_2$. On a alors, pour tout $n \geq \max\{n_1, n_2\}$,

$$\begin{aligned} |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| &\leq |\mathbb{E}[(g\check{g})(X_n)] - \mathbb{E}[(g\check{g})(X)]| + |\mathbb{E}[g(1 - \check{g})(X_n)]| + |\mathbb{E}[g(1 - \check{g})(X)]| \\ &\leq \frac{\varepsilon}{2} + \mathbb{E}[(1 - \check{g})(X_n)] + \mathbb{E}[(1 - \check{g})(X)] \leq \varepsilon. \end{aligned}$$

Ceci termine la preuve de cette implication.

1) \implies 3):

Il suffit de prendre $g_1(x) = \cos(tx)$ et $g_2(x) = \sin(tx)$:

$$\varphi_{X_n}(t) = \mathbb{E}[g_1(X_n)] + i\mathbb{E}[g_2(X_n)] \rightarrow \mathbb{E}[g_1(X)] + i\mathbb{E}[g_2(X)] = \varphi_X(t).$$

3) \implies 1):

Pas a)

Supposons d'abord que $g : \mathbb{R} \rightarrow \mathbb{R}$ est continue à support compact, avec sa transformée de Fourier $\hat{g}(t) = \int_{\mathbb{R}} e^{itx}g(x)dx$ est dans $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. On sait, par inversion de Fourier que $g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx}\hat{g}(t)dt$. Par le théorème de Fubini,

$$\mathbb{E}[g(X_n)] = \frac{1}{2\pi} \mathbb{E} \left[\int_{\mathbb{R}} e^{-itX_n}\hat{g}(t)dt \right] = \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E} [e^{-itX_n}] \hat{g}(t)dt = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_{X_n}(-t)\hat{g}(t)dt.$$

Mais, $\varphi_{X_n}(-t)$ converge vers $\varphi_X(-t)$ pour chaque t et de plus, on sait que $|\varphi_{X_n}(-t)| \leq 1$, pour tout $t \in \mathbb{R}$. Comme \hat{g} est intégrable, par convergence dominée on trouve

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(X_n)] = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(-t)\hat{g}(t)dt = \mathbb{E}[g(X)].$$

Pas b)

On peut voir que la condition $\hat{g} \in L^1(\mathbb{R})$ est satisfaite pour toute fonction g de classe C^2 à support compact. En effet, la transformée de Fourier \hat{g} est bornée et la transformée de Fourier de la fonction à support compact g'' est bornée et est égale à $t^2\hat{g}(t)$. D'où pour

$|t| \rightarrow \infty$, $\hat{g}(t) = O(\frac{1}{t^2})$. Comme \hat{g} est bornée, elle est donc intégrable sur \mathbb{R} . Ainsi l'affirmation est vérifiée pour g de classe C^2 à support compact.

Pas c)

Pour g juste continue à support compact et ensuite g continue bornée on peut répéter les mêmes arguments que dans la preuve 2) \implies 1) (les pas *b*) et *c*)).

La preuve du théorème est complète. \square

REMARQUE: (*facultatif*)

En fait l'implication 3) \implies 1) peut être énoncée aussi comme le **théorème de Lévy**:
Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires de fonctions caractéristiques φ_{X_n} . Supposons que la suite φ_{X_n} converge ponctuellement vers une fonction ψ continue en 0. Alors ψ est la fonction caractéristique d'une variable aléatoire X et la suite X_n converge en loi vers X .

Ce résultat repose sur une question intéressante : savoir si d'une suite de variables aléatoires on peut extraire une sous suite convergente en loi (on dira que la suite est **relativement compacte** pour la convergence étroite).

Ce n'est pas toujours le cas: soit la suite de variables aléatoires $X_n = n$. Pour toute fonction g continue tendant vers 0 à l'infini on a $E[g(X_n)] = g(n) \rightarrow 0$, donc on ne peut pas extraire une sous suite convergente en loi. Il s'agit du fait que "la masse est partie à l'infini". Lorsqu'on évite ce phénomène alors la suite est relativement compacte et réciproquement.

La formulation du phénomène de ne pas perdre de masse est la suivante:

Une suite de lois P_n sur \mathbb{R} est dite **tendue** si

$$\lim_{M \rightarrow \infty} \sup_{n \in \mathbb{N}} P_n(\{x \in \mathbb{R} : |x| \geq M\}) = 0,$$

ou encore, pour tout $\varepsilon > 0$, il existe un $M > 0$ tel que

$$P_n(\{x \in \mathbb{R} : |x| \geq M\}) \leq \varepsilon, \forall n \in \mathbb{N}.$$

Le résultat important est donné par le **théorème de Prohorov**:

Une suite de lois de probabilités est relativement compacte (pour la convergence étroite) si et seulement si elle est tendue. \square

4.5 Les lois des grands nombres et le théorème central limite

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et de même loi (i.i.d.) qu'une variable notée X . Pour tout entier $n \geq 1$, on $S_n = \sum_{j=1}^n X_j$ et on va s'intéresser aux propriétés asymptotiques de S_n , ou plus précisément de $\frac{S_n}{n}$, la **moyenne empirique** des X_j . De point de vue pratique, par exemple, si X_j modélise le nombre de pile montre au lancer d'une pièce, $\frac{S_n}{n}$ est la proportion de lancers montrant pile dans n essais. On peut voir sur cet exemple que $\frac{S_n}{n}$ converge vers $E(X) = P(X = 1)$ et que la loi de $(S_n - E(S_n))/\sqrt{n}$ ressemble à une loi normale lorsque n est assez grand.

Le premier résultat montre que l'intuition est correcte: si la pièce est non truquée la proportion de pile tend à se stabiliser vers 0,5; donc la théorie construite constitue une modélisation raisonnable du réel et qu'une certaine régularité apparaît dans les phénomènes

aléatoires.

Théorème 4.1 (*loi faible de grands nombres*)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires i.i.d. avec $E(|X|) < \infty$ et soit $S_n = X_1 + \dots + X_n$. Alors, lorsque $n \rightarrow \infty$,

$$\frac{S_n}{n} \xrightarrow{\text{prob}} E(X).$$

Preuve: Comme $X \in L^1$, la fonction caractéristique φ_X est dérivable et on sait que $\varphi_X(0) = 1$ et $\varphi'_X(0) = iE(X)$, d'où son développement,

$$\varphi_X(t) = 1 + itE(X) + o(t), \text{ lorsque } t \rightarrow 0.$$

Comme les variables X_1, \dots, X_n sont indépendantes et de même loi que X , la fonction caractéristique de S_n est $\varphi_{S_n} = \varphi_X^n$. Donc lorsque $n \rightarrow \infty$,

$$\varphi_{\frac{S_n}{n}}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \varphi_X\left(\frac{t}{n}\right)^n = \left(1 + itE(X)\frac{1}{n} + o\left(\frac{t}{n}\right)\right)^n.$$

Par convergence dominée, on voit que, lorsque $n \rightarrow \infty$,

$$n |\varphi_X\left(\frac{t}{n}\right) - (1 + i\frac{t}{n}E(X))| \rightarrow 0.$$

En effet on a¹

$$|\varphi_X\left(\frac{t}{n}\right) - (1 + i\frac{t}{n}E(X))| \leq E \left[\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n} \right],$$

avec

$$n \left(\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n} \right) \leq 2t|X| \in L^1 \text{ et } n \left(\frac{t^2|X|^2}{2n^2} \wedge \frac{2t|X|}{n} \right) \leq \frac{t^2|X|^2}{2n} \rightarrow 0.$$

1. Par la formule de Taylor au reste integral on a

$$e^{ix} = \sum_{k=0}^n \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds,$$

d'où

$$|e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!}| \leq \frac{x^{n+1}}{(n+1)!}.$$

Par ailleurs, un calcul simple montre que

$$e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} = \frac{i^n}{(n-1)!} \int_0^x (x-s)^{n-1} e^{is} ds - \frac{(ix)^n}{n!},$$

d'où

$$|e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!}| \leq \frac{|x|^n}{n!} + \frac{|x|^n}{n!} = \frac{2|x|^n}{n!}.$$

Ainsi

$$|e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!}| \leq \frac{x^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!}.$$

Enfin on déduit

$$|\varphi_X(t) - \sum_{k=0}^n \frac{(it)^k}{k!} E(X^k)| \leq E \left[\frac{|tX|^{n+1}}{(n+1)!} \wedge \frac{2|tX|^n}{n!} \right].$$

Il suffit de prendre $n=1$ et $n=2$ (pour la preuve du Théorème 4.3 ci-dessous).

Mais alors², lorsque $n \rightarrow \infty$,

$$|\varphi_X\left(\frac{t}{n}\right)^n - \left(1 + \frac{itE(X)}{n}\right)^n| \leq n|\varphi_X\left(\frac{t}{n}\right) - \left(1 + i\frac{t}{n}E(X)\right)| \rightarrow 0.$$

On en déduit

$$\lim_{n \rightarrow \infty} \varphi_{\frac{S_n}{n}}(t) = e^{itE(X)},$$

où le membre droit est la fonction caractéristique de la variable aléatoire constante $E(X)$. Par la Définition 4.5 on déduit que $\frac{S_n}{n}$ converge en loi vers la constante $E(X)$, et par la Proposition 4.14 on tire la conclusion. \square

Théorème 4.2 (*loi forte de grands nombres*)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires i.i.d. et soit $S_n = X_1 + \dots + X_n$. Les deux conditions suivantes sont équivalentes:

- 1) $E(|X|) < \infty$;
- 2) $\frac{S_n}{n} \xrightarrow{p.s.} E(X)$, lorsque $n \rightarrow \infty$.

Preuve:

2) \implies 1):

Si la suite $\frac{S_n}{n}$ converge p.s., alors la suite $\frac{X_n}{n}$ converge p.s. vers 0. Par le lemme de Borel-Cantelli (Proposition 4.1, 2)), puisque les X_j sont indépendantes et toutes de même loi que X , pour tout (ou seulement un) $\varepsilon > 0$,

$$\sum_{n \geq 1} P\left(\left|\frac{X_n}{n}\right| > \varepsilon\right) = \sum_{n \geq 1} P(|X_n| > \varepsilon n) = \sum_{n \geq 1} P(|X| > \varepsilon n) < \infty.$$

On conclut à l'aide de la Proposition 2.8.

1) \implies 2): Notons $E(X) = m$.

Pas a)

On prouve dans un premier temps le résultat sous l'hypothèse plus restrictive que X admet un moment d'ordre 2: $E(|X|^2) < \infty$. En considérant séparément la partie positive et sa partie négative X_n , on remarque qu'il suffit de démontrer le résultat pour des variables aléatoires positives, ce qu'on suppose désormais. On rappelle que $E(S_n) = nm$ et $\text{Var}(S_n) = n\text{Var}(X)$. D'où,

$$\text{Var}\left(\frac{S_n^4}{n^4} - m\right) = \frac{\text{Var}(X)}{n^4},$$

et donc, par l'inégalité de Tchebychev,

$$P\left(\left|\frac{S_n^4}{n^4} - m\right| \geq \frac{1}{n}\right) \leq \frac{\text{Var}(X)}{n^2}.$$

2. On a pour $i = 1, \dots, n$ et $a_i, b_i \in \mathbb{C}$, $|a_i|, |b_i| \leq 1$

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i - b_i|$$

(preuve par récurrence).

On déduit que la série $\sum_{n \geq 1} \mathbb{P}(|\frac{S_{n^4}}{n^4} - m| \geq \frac{1}{n})$ converge. Par le lemme de Borel-Cantelli on a presque sûrement,

$$|\frac{S_{n^4}}{n^4} - m| \leq \frac{1}{n}, \text{ sauf pour un nombre fini d'entiers } n.$$

On fixe un ω pour lequel l'événement ci-dessus est réalisé et on prend n suffisamment grand. Comme la suite S_n est croissante, pour tout entier k tel que $n^4 \leq k \leq (n+1)^4$, on a

$$\frac{S_k(\omega)}{k} \leq \frac{S_{(n+1)^4}(\omega)}{n^4} = \left(\frac{n+1}{n}\right)^4 \frac{S_{(n+1)^4}(\omega)}{(n+1)^4} \leq \left(\frac{n+1}{n}\right)^4 \left(m + \frac{1}{n+1}\right).$$

Ainsi,

$$\limsup_{k \rightarrow \infty} \frac{S_k(\omega)}{k} \leq m.$$

De même

$$\frac{S_k(\omega)}{k} \geq \frac{S_{n^4}(\omega)}{(n+1)^4} = \left(\frac{n}{n+1}\right)^4 \frac{S_{n^4}(\omega)}{n^4} \geq \left(\frac{n}{n+1}\right)^4 \left(m - \frac{1}{n}\right).$$

On déduit

$$\liminf_{k \rightarrow \infty} \frac{S_k(\omega)}{k} \geq m.$$

En conséquence, on a pour presque tout ω , $\lim_{k \rightarrow \infty} \frac{S_k(\omega)}{k} = m$.

Pas b)

Supposons maintenant X intégrable. Quitte à remplacer les variables X_j par $X_j - \mathbb{E}(X_j)$, on peut supposer que $m = \mathbb{E}(X) = 0$. On fixe $\varepsilon > 0$. Comme $\mathbb{E}(|X|) < \infty$, il existe des variables Y_j , bornées, centrées indépendantes et de même loi, telles que $\mathbb{E}(|X_j - Y_j|) \leq \varepsilon$.³

On note $T_n = Y_1 + \dots + Y_n$ et on a

$$\frac{|S_n|}{n} \leq \frac{1}{n} \sum_{j=1}^n |X_j - Y_j| + \frac{|T_n|}{n}.$$

D'après les pas a), $\frac{|T_n|}{n}$ converge presque sûrement vers 0. On va montrer que pour ε suffisamment petit,

$$(*) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n |X_j - Y_j| \leq 2\mathbb{E}(|X_1 - Y_1|),$$

d'où on tire, par le choix des variables Y_j ,

$$\limsup_{n \rightarrow \infty} \frac{|S_n|}{n} \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n |X_j - Y_j| + \limsup_{n \rightarrow \infty} \frac{|T_n|}{n} \leq 2\varepsilon$$

presque sûrement. Comme ε est arbitraire, on conclut. On note $Z_j = |X_j - Y_j|$ variables aléatoires positives, indépendantes, de même loi qu'une variable Z intégrable et on veut borner

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n Z_j.$$

3. Il existe $c > 0$ assez grand tel que $\mathbb{E}(|X| \mathbb{1}_{\{|X| > c\}}) \leq \varepsilon$. On prend $Y' = X \mathbb{1}_{\{|X| > c\}}$ et $Y = Y' - \mathbb{E}(Y')$ qui est centrée bornée par $2c$. Alors $\mathbb{E}(|X - Y|) \leq \mathbb{E}(|X - Y'|) + \mathbb{E}(Y') = \mathbb{E}(|X - Y'|) + |\mathbb{E}(X) - \mathbb{E}(Y')| \leq 2\mathbb{E}(|X - Y'|) \leq 2\varepsilon$. On prend $\varepsilon/2$ au lieu de ε et ensuite on fait des copies indépendantes de même loi de Y .

On a pour tout entier k et tout $\delta > 0$,

$$\begin{aligned}
& \mathbb{P} \left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{j=1}^n Z_j \geq 2\mathbb{E}(Z) + \delta \right) \\
& \leq \mathbb{P} \left(\exists j \in \{1, \dots, 2^{k+1}\} : Z_j > 2^k \right) + \mathbb{P} \left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{j=1}^n Z_j \mathbb{1}_{[0, 2^k]}(Z_j) \geq 2\mathbb{E}(Z) + \delta \right) \\
& \leq 2^{k+1} \mathbb{P}(X > 2^k) + \mathbb{P} \left(\sum_{j=1}^{2^{k+1}} Z_j \mathbb{1}_{[0, 2^k]}(Z_j) \geq 2^{k+1} \mathbb{E}(Z) + \delta 2^k \right) \\
& \leq 2^{k+1} \mathbb{P}(X > 2^k) + \mathbb{P} \left(\sum_{j=1}^{2^{k+1}} \left(Z_j \mathbb{1}_{[0, 2^k]}(Z_j) - \mathbb{E}(Z_j \mathbb{1}_{[0, 2^k]}(Z_j)) \right) \geq \delta 2^k \right).
\end{aligned}$$

On applique l'inégalité de Tchebychev au deuxième terme du majorant précédent et on trouve

$$\begin{aligned}
\mathbb{P} \left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{j=1}^n Z_j \geq 2\mathbb{E}(Z) + \delta \right) & \leq 2^{k+1} \mathbb{P}(X > 2^k) + \frac{1}{\delta^2 2^{2k}} 2^{k+1} \mathbb{E} \left(Z^2 \mathbb{1}_{[0, 2^k]}(Z) \right) \\
& \leq 2^{k+1} \mathbb{P}(X > 2^k) + \frac{2}{\delta^2 2^k} \mathbb{E} \left(Z^2 \mathbb{1}_{[0, 2^k]}(Z) \right)
\end{aligned}$$

Pour tout entier k ,

$$\int_{2^k}^{2^{k+1}} \mathbb{P}(X > t) dt \geq 2^k \mathbb{P}(X > 2^{k+1}),$$

donc la Proposition 2.8 implique

$$\sum_{k \geq 0} 2^{k+1} \mathbb{P}(Z > 2^k) \leq 4\mathbb{E}(Z) < \infty.$$

De plus, si $2^l < Z \leq 2^{l+1}$, alors

$$Z^2 \sum_{k \geq 1} 2^{-k} \mathbb{1}_{[0, 2^k]}(Z) \leq 2^{2l+2} \sum_{k \geq l+1} 2^{-k} \leq 4Z,$$

d'où

$$\sum_{k \geq 0} 2^{-k} \mathbb{E} \left(Z^2 \mathbb{1}_{[0, 2^k]}(Z) \right) = \mathbb{E} \left(Z^2 \sum_{k \geq 0} 2^{-k} \mathbb{1}_{[0, 2^k]}(Z) \right) \leq 4\mathbb{E}(Z).$$

Donc, finalement,

$$\sum_{k \geq 0} \mathbb{P} \left(\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{j=1}^n Z_j \geq 2\mathbb{E}(Z) + \delta \right) \leq 4 \left(1 + \frac{2}{\delta^2} \right) \mathbb{E}(Z) < \infty.$$

D'après le lemme de Borel-Cantelli (Théorème 3.3 a)), presque sûrement pour tout k assez grand

$$\max_{2^k < n \leq 2^{k+1}} \frac{1}{n} \sum_{j=1}^n Z_j < 2\mathbb{E}(Z) + \delta,$$

et comme δ est arbitraire on trouve l'inégalité (*). \square

REMARQUE: Une autre façon d'énoncer la loi forte des grands nombres est de dire que si $\mathbb{E}(|X|) < \infty$ alors $\frac{S_n}{n} = \mathbb{E}(X) + o(1)$ p.s. lorsque $n \rightarrow \infty$. Le théorème limite central donne, en un certain sens, un terme de plus dans le développement asymptotique, précisant le comportement limite en loi du terme $o(1)$, sous une hypothèse supplémentaire sur la loi commune des X_j . On a ainsi une estimation très précise de l'erreur commise en approchant l'espérance mathématique par la moyenne empirique. Le résultat permet d'approximer la loi de $\frac{S_n}{n}$ lorsque n est grand.

Théorème 4.3 (*central limite*)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. avec $\mathbb{E}(X^2) < \infty$ et soit $S_n = X_1 + \dots + X_n$. Alors, lorsque $n \rightarrow \infty$,

$$\frac{S_n - n\mathbb{E}(X)}{\sqrt{n\text{Var}(X)}} \xrightarrow{\text{loi}} \mathcal{N}(0,1).$$

Preuve: En remplaçant X_j par $(X_j - \mathbb{E}(X_j))/\sqrt{\text{Var}(X_j)}$ on peut supposer que $\mathbb{E}(X) = 0$ et $\text{Var}(X) = 1$. Comme $X \in L^2$ sa fonction caractéristique est deux fois dérivable et son développement est

$$\varphi_X(t) = 1 + t\varphi'_X(0) + \frac{t^2}{2}\varphi''_X(0) + o(t^2) = 1 - \frac{t^2}{2} + o(t^2), \text{ lorsque } t \rightarrow 0.$$

On rappelle que par indépendance et équidistribution on a

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_X(t/\sqrt{n})^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n, \text{ lorsque } n \rightarrow \infty.$$

Par convergence dominée, on voit que, lorsque $n \rightarrow \infty$,

$$\left|\varphi_X\left(\frac{t}{\sqrt{n}}\right) - \left(1 + i\frac{t}{\sqrt{n}}\mathbb{E}(X) - \frac{t^2}{n}\mathbb{E}(X^2)\right)\right| \rightarrow 0.$$

En effet on a

$$\left|\varphi_X\left(\frac{t}{\sqrt{n}}\right) - \left(1 + i\frac{t}{\sqrt{n}}\mathbb{E}(X) - \frac{t^2}{n}\mathbb{E}(X^2)\right)\right| \leq \mathbb{E}\left[\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right],$$

avec

$$n\left(\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right) \leq t^2|X|^2 \in L^1 \text{ et } n\left(\frac{t^3|X|^3}{6n^{3/2}} \wedge \frac{t^2|X|^2}{n}\right) \leq \frac{t^3|X|^3}{6\sqrt{n}} \rightarrow 0.$$

A nouveau, lorsque $n \rightarrow \infty$

$$\left|\varphi_X(t/\sqrt{n})^n - \left(1 - \frac{t^2}{2n}\right)^n\right| \leq n\left|\varphi_X\left(\frac{t}{\sqrt{n}}\right) - \left(1 + i\frac{t}{\sqrt{n}}\mathbb{E}(X) - \frac{t^2}{n}\mathbb{E}(X^2)\right)\right| \rightarrow 0.$$

On en déduit

$$\lim_{n \rightarrow \infty} \varphi_{S_n/\sqrt{n}}(t) = e^{-t^2/2}.$$

Le membre de droite est la fonction caractéristique de la loi $\mathcal{N}(0,1)$ et le théorème est établi. \square

Proposition 4.16 (*vitesse de convergence en théorème central limite*)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. avec $E(|X|^3) < \infty$. On note $E(X) = m$, $\text{Var}(X) = \sigma^2$, $S_n = X_1 + \dots + X_n$ et $T_n = (S_n - nm)/\sqrt{n\sigma^2}$ et $N \sim \mathcal{N}(0,1)$. Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction de classe C^3 dont la dérivée troisième est bornée. Alors, lorsque $n \rightarrow \infty$,

$$E[g(T_n)] - E[g(N)] = O\left(\frac{1}{\sqrt{n}}\right).$$

Preuve. Sans perte de généralité on peut supposer $m = 0$. On introduit une suite $(N_n)_{n \in \mathbb{N}}$ de variables aléatoires i.i.d. de loi $\mathcal{N}(0,1)$ et on pose $\bar{T}_n = (N_1 + \dots + N_n)/\sqrt{n}$ de sorte que pour chaque n , \bar{T}_n a encore pour loi $\mathcal{N}(0,1)$. Pour chaque $k \in \{1, \dots, n\}$, on considère

$$T_n^k = \frac{X_1 + \dots + X_k + N_{k+1} + \dots + N_n}{\sqrt{n}} \quad \text{et} \quad \hat{T}_n^k = \frac{X_1 + \dots + X_{k-1} + N_{k+1} + \dots + N_n}{\sqrt{n}}.$$

On cherche donc à estimer

$$\begin{aligned} |E[g(T_n)] - E[g(\bar{T}_n)]| &= \left| \sum_{k=0}^{n-1} E[g(T_n^k)] - E[g(T_n^{k+1})] \right| \\ &\leq \sum_{k=0}^{n-1} \left| E\left(g(T_n^k) - g(\hat{T}_n^k) + g(\hat{T}_n^k) - g(T_n^{k+1})\right) \right|. \end{aligned}$$

Pour chaque différence on écrit le développement de Taylor de g au point \hat{T}_n^k . On voit que $T_n^k = \hat{T}_n^k + X_k/\sqrt{n}$ et $T_n^{k+1} = \hat{T}_n^k + N_k/\sqrt{n}$:

$$g(T_n^k) - g(\hat{T}_n^k) = g'(\hat{T}_n^k) \frac{X_k}{\sqrt{n}} + \frac{1}{2} g''(\hat{T}_n^k) \left(\frac{X_k}{\sqrt{n}}\right)^2 + R_n^k$$

et

$$g(T_n^{k+1}) - g(\hat{T}_n^k) = g'(\hat{T}_n^k) \frac{N_k}{\sqrt{n}} + \frac{1}{2} g''(\hat{T}_n^k) \left(\frac{N_k}{\sqrt{n}}\right)^2 + \hat{R}_n^k.$$

On sait que \hat{T}_n^k est indépendante de N_k et aussi de X_k et que ces deux dernières variables sont centrées et ont la même variance. Quand on prend la somme des deux développements et l'espérance, tout s'annule sauf les restes, $E(R_n^k)$ et $E(\hat{R}_n^k)$. Or

$$|R_n^k| \leq (\sup_{x \in \mathbb{R}} |g'''(x)|) \frac{|X_k|^3}{n^{3/2}} \quad \text{et} \quad |\hat{R}_n^k| \leq (\sup_{x \in \mathbb{R}} |g'''(x)|) \frac{|N_k|^3}{n^{3/2}}.$$

On trouve

$$|E[g(T_n)] - E[g(\bar{T}_n)]| \leq c \sum_{k=0}^{n-1} \frac{1}{n^{3/2}} = c \frac{1}{n^{1/2}}.$$

\square

4.6 Exercices

4.1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes. Montrer que le rayon de convergence de la série entière $\sum_{n \in \mathbb{N}} X_n z^n$ à coefficients aléatoires est p.s. constant.

4.2. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de loi donnée par $P(X_n = 1) = p$, $P(X_n = -1) = 1 - p$, $0 < p < 1$, $p \neq \frac{1}{2}$. On pose $S_0 = 0$ et pour $n \geq 1$, $S_n = X_1 + \dots + X_n$. L'événement $A_n = \{S_n = 0\}$ s'appelle un retour à 0.

(i) Que représente l'événement $\limsup_n A_n$?

(ii) Montrer que $P(\limsup_n A_n) = 0$.

4.3. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes.

(i) Montrer que la série $\sum_{n \in \mathbb{N}} X_n$ converge ou diverge presque sûrement.

(ii) Montrer que si la série $\sum_{n \in \mathbb{N}} X_n$ converge presque sûrement, alors pour tout $a > 0$, $\sum_{n \in \mathbb{N}} P(|X_n| > a) < \infty$.

4.4. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes. Montrer que:

(i) si pour un $M \in \mathbb{R}$, $\sum_{n \in \mathbb{N}} (1 - F_{X_n}(M)) < \infty$, alors $\limsup_{n \uparrow \infty} X_n \leq M$ p.s.

(ii) si pour un $M \in \mathbb{R}$, $\sum_{n \in \mathbb{N}} F_{X_n}(M) < \infty$, alors $\liminf_{n \uparrow \infty} X_n \geq M$ p.s.

4.5. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes. La loi de X_n est:

$$P_{X_n} = \frac{1}{n^2}(\delta_{n^4} + \delta_{n^{-4}}) + \left(1 - \frac{2}{n^2}\right) \delta_0.$$

Montrer que la série $\sum_{n \in \mathbb{N}^*} X_n$ converge presque sûrement.

4.6.* Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi exponentielle de paramètre 1. On note $M_n = \max_{1 \leq j \leq n} X_j$.

(i) Montrer que pour tout $\varepsilon > 0$

$$P(M_n \leq (1 - \varepsilon) \ln n) = \exp(n \ln(1 - n^{-\varepsilon-1})).$$

(ii) Par le lemme de Borel-Cantelli, en déduire

$$\frac{M_n}{\ln n} \geq 1 - \varepsilon \text{ p.s. pour } n \text{ assez grand.}$$

(iii) Montrer que pour tout $\varepsilon > 0$

$$P(M_n \geq (1 + \varepsilon) \ln n) = 1 - \exp(n \ln(1 - n^{-\varepsilon-1})).$$

(iv) Soit la sous-suite $n_k = [(k + 1)^\delta]$, $k \in \mathbb{N}$, $\varepsilon \delta > 1$, où $[\cdot]$ note la partie entière. Par le lemme de Borel-Cantelli, en déduire

$$\limsup_{k \rightarrow \infty} \frac{M_{n_k}}{\ln n_k} \leq 1, \text{ p.s.}$$

Quelle est la limite presque sûre de $\frac{M_{n_k}}{\ln n_k}$?

(v) Pour tout entier $n > 0$ il existe $k \in \mathbb{N}$ tel que $n_k \leq n < n_{k+1}$. Montrer que

$$\frac{\ln n_k}{\ln n_{k+1}} \cdot \frac{M_{n_k}}{\ln n_k} \leq \frac{M_n}{\ln n} \leq \frac{M_{n_{k+1}}}{\ln n_{k+1}} \cdot \frac{\ln n_{k+1}}{\ln n_k}.$$

En déduire que, lorsque $n \rightarrow \infty$

$$\frac{M_n}{\ln n} \xrightarrow{\text{p.s.}} 1.$$

(vi) On a donc que $M_n = \ln n + o(\ln n)$ p.s. et on veut préciser le contenu du terme $o(\ln n)$. On note $Z_n = M_n - \ln n$. Montrer, en utilisant les fonctions de répartition que Z_n converge en loi vers une variable dont on précisera la densité.

4.7. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes centrées uniformément bornées.

(i) Montrer que $E[(X_1 + \dots + X_n)^4] \leq \text{cste} \cdot n^2$.

(ii) En déduire que, lorsque $n \rightarrow \infty$

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{p.s.}} 0.$$

4.8. Montrer que l'espace L^0 peut être muni d'une structure d'espace métrique complet. En déduire l'unicité de la limite en probabilité.

4.9. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles admettant un moment d'ordre deux et telles que $\lim_{n \rightarrow \infty} E(X_n) = c$ et $\lim_{n \rightarrow \infty} \text{Var}(X_n) = 0$. Montrer que X_n converge en probabilité vers c .

4.10. Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires telles que X_n converge vers une variable X en probabilité et Y_n converge vers une variable Y en probabilité. Soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction uniformément continue.

(i) Montrer que $g(X_n, Y_n)$ converge en probabilité vers $g(X, Y)$.

(ii) Montrer que $X_n + Y_n$ converge en probabilité vers $X + Y$.

4.11. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de loi de Bernoulli telles que $E(X_n) = \frac{1}{n}$. Étudier la convergence en probabilité et presque sûre de cette suite.

4.12. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles de même loi admettant un moment d'ordre deux, indépendantes. Montrer que, lorsque $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \max_{0 \leq j \leq n} |X_j| \xrightarrow{\text{prob}} 0.$$

4.13. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes telles que X_n suit la loi uniforme sur $\{\pm n^\alpha\}$, $\alpha > 0$. Montrer que si $0 < \alpha < \frac{1}{2}$ alors la suite satisfait la loi faible des grands nombres. Est-ce que la réciproque de cette affirmation est vraie?

4.14. Montrer que si X_n converge vers une variable X dans L^1 , alors $\lim_{n \rightarrow \infty} E(X_n) = E(X)$.

4.15. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et il existe un réel $M > 0$ tel que, pour tout entier n , $|X_n| \leq M$ p.s. Montrer que, lorsque $n \rightarrow \infty$

$$X_n \xrightarrow{\text{proba}} 0 \iff X_n \xrightarrow{L^1} 0.$$

4.16. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles qui converge en probabilité vers une constante c . Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue bornée. Montrer que $g(X_n) \xrightarrow{L^1} g(c)$.

4.17. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles deux-à-deux non corrélées telles que $\sup_{n \in \mathbb{N}^*} \mathbb{E}(|X_n^2|) < \infty$. On note $S_n = X_1 + \dots + X_n$.

(i) On suppose que les X_n ont toutes la même espérance m . Montrer que, lorsque $n \rightarrow \infty$

$$\frac{S_n}{n} \xrightarrow{L^2} m.$$

(ii) On suppose que $\lim_{n \rightarrow \infty} \frac{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)}{n} = m$. Montrer que, lorsque $n \rightarrow \infty$

$$\frac{S_n}{n} \xrightarrow{\text{prob}} m.$$

(iii) On suppose que les X_n sont toutes centrées et que de plus $\sum_{n \in \mathbb{N}^*} \text{Var}(X_n) < \infty$. Montrer que S_n converge dans L^2 .

4.18. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi binomiale $X_n \sim \mathcal{B}(n, \frac{\lambda}{n})$ ($0 < \lambda < 1$). Montrer que X_n converge en loi vers une variable aléatoire de loi de Poisson de paramètre λ .

4.19. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi uniforme $X_n \sim \mathcal{U}(1, \dots, n)$. Montrer que $\frac{X_n}{n}$ converge en loi vers une variable aléatoire de loi $\mathcal{U}_{[0,1]}$.

4.20. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi géométrique $X_n \sim \mathcal{G}(\frac{p}{n})$ ($0 < p < 1$). Montrer que $\frac{X_n}{n}$ converge en loi (à l'aide des fonctions caractéristiques et des fonctions de répartition).

4.21. Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires réelles telles que pour tout n X_n et Y_n sont indépendantes. On suppose que X_n converge en loi vers une variable X et Y_n converge en loi vers une variable Y . Montrer que la suite des couples (X_n, Y_n) converge en loi.

4.22. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi de Bernoulli de paramètre $\frac{1}{2}$ et soit $U \sim \mathcal{U}_{[0,1]}$. On pose

$$Y_n = \sum_{k=1}^n \frac{X_k}{2^k}.$$

(i) Montrer que la fonction caractéristique de Y_n peut se mettre sous les deux formes suivantes:

$$\varphi_{Y_n}(t) = e^{\frac{it}{2}} e^{-\frac{it}{2^{n+1}}} \cos\left(\frac{t}{4}\right) \cos\left(\frac{t}{8}\right) \cos\left(\frac{t}{16}\right) \dots \cos\left(\frac{t}{2^{n+1}}\right) = e^{\frac{it}{2}} e^{-\frac{it}{2^{n+1}}} \left(\frac{1}{2}\right)^n \frac{\sin\left(\frac{t}{2}\right)}{\sin\left(\frac{t}{2^{n+1}}\right)}.$$

(ii) En déduire que Y_n converge en loi vers U .

4.23. Soit $\psi : [0,1] \rightarrow [0,1]$ et soit $(U_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de loi uniforme sur $[0,1]$. On pose

$$I = \int_0^1 \psi(x) dx, Y_k = \begin{cases} 1 & \text{si } \psi(U_{2k}) \geq U_{2k+1} \\ 0 & \text{sinon} \end{cases}$$

- (i) Montrer que les variables aléatoires Y_k sont indépendantes et trouver leurs lois.
 (ii) Montrer que, lorsque $n \rightarrow \infty$

$$\bar{Y}_n = \frac{1}{n}(Y_1 + \dots + Y_n) \xrightarrow{\text{p.s.}} I.$$

(iii) (*méthode de Monte-Carlo simple*)

On veut estimer I à l'aide de \bar{Y}_n et l'erreur relative est

$$\varepsilon_n = \frac{|\bar{Y}_n - I|}{I}.$$

Donner un majorant de $P(\varepsilon_n > \alpha)$ en termes de α, I, n .

(iv) Supposons que des estimations ont permis par ailleurs de voir que $I > 0,5$. À partir de quelle valeur de n s'assure-t-on 19 chances sur 20 de faire une erreur relative inférieure à 1% en estimant I ?

4.24. Soit $f : \mathbb{R} \rightarrow [0, \infty[$ une densité de probabilité et soit $\psi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction intégrable par rapport à la mesure de Lebesgue. On pose

$$I = \int_{\mathbb{R}} \psi(x) f(x) dx.$$

(i) (*méthode de Monte-Carlo simple*)

Montrer que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles indépendantes ayant la même densité f alors, lorsque $n \rightarrow \infty$

$$\frac{1}{n}(\psi(X_1) + \dots + \psi(X_n)) \xrightarrow{\text{p.s.}} I.$$

(ii) (*méthode de Monte-Carlo avec rejet*)

Soit $(U_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de loi uniforme sur $[0,1]$ indépendante de $(X_n)_{n \in \mathbb{N}}$ et soit $g : \mathbb{R} \rightarrow [0, \infty[$ une densité de probabilité telle que pour tout $x \in \mathbb{R}$, $\frac{g(x)}{f(x)} \leq c$ pour une certaine constante c . Montrer que, lorsque $n \rightarrow \infty$

$$\frac{c}{n} \sum_{j=1}^n \psi(X_j) \mathbb{1}_{g(X_j) < cU_j f(X_j)} \xrightarrow{\text{p.s.}} \int_{\mathbb{R}} \psi(x) g(x) dx.$$

(iii) (*facultatif*) On pose

$$Y_n = \begin{cases} 1 & \text{si } cU_n < \frac{g(X_n)}{f(X_n)} \\ 0 & \text{sinon} \end{cases}$$

Montrer que la suite $(Y_n)_{n \in \mathbb{N}}$ est une suite de Bernoulli indépendante. On pose

$$N(n) = \min\{k \in \mathbb{N} : Y_1 + \dots + Y_k = n\}.$$

Donner un équivalent presque sûr de $N(n)$. On considère les variables aléatoires

$$Z_1 = X_{N(1)}, \dots, Z_n = X_{N(n)}, \dots$$

Montrer que la suite $(Z_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes de densité g . En déduire que, lorsque $n \rightarrow \infty$

$$\frac{1}{n}(\psi(Z_1) + \dots + \psi(Z_n)) \xrightarrow{\text{P.S.}} \int_{\mathbb{R}} \psi(x)g(x)dx.$$

4.25. (formule d'inversion de transformée de Laplace de Post-Widder)

Soit $f : [0, \infty[\rightarrow [0, \infty[$ une fonction (densité de probabilité) continue bornée. On rappelle que la transformée de Laplace de f est la fonction

$$g(s) = \int_0^\infty e^{-sx} f(x) dx.$$

(i) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes de même loi exponentielle $\mathcal{E}(\frac{1}{x})$. Quelles sont les densités de $S_n = X_1 + \dots + X_n$ et de $\frac{S_n}{n}$.

(ii) Montrer que g est indéfiniment dérivable sur $] \varepsilon, \infty[$, pour tout $\varepsilon > 0$, de dérivée n -ième donnée sur $]0, \infty[$ par

$$g^{(n)}(s) = (-1)^n \int_0^\infty x^n e^{-sx} f(x) dx.$$

(iii) Montrer que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] = f(x).$$

(iv) En déduire la formule d'inversion de Post-Widder:

$$\lim_{n \rightarrow \infty} \frac{(-1)^{n-1} n^n g^{(n-1)}(\frac{n}{x})}{x^n (n-1)!} = f(x),$$

et l'injectivité de la transformée de Laplace vue comme application de l'espace des fonctions continues bornées dans l'espace des fonctions continues.

4.26. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de loi de Poisson $X_n \sim \mathcal{P}(a_n)$ ($a_n > 0$). On note $s_n = a_1 + \dots + a_n$ et $S_n = X_1 + \dots + X_n$. On suppose que $\lim_{n \rightarrow \infty} s_n = \infty$.

(i) Montrer que, lorsque $n \rightarrow \infty$

$$\frac{S_n - s_n}{\sqrt{s_n}} \xrightarrow{\text{loi}} G \sim \mathcal{N}(0,1).$$

(ii)* On suppose que tous les $a_n = 1$ et on pose $T_n = \frac{S_n - n}{\sqrt{n}}$. Pourquoi pour tout $x > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(x \leq T_n) = \mathbb{P}(x \leq G)$? À l'aide de l'inégalité Bienaymé-Tchebychev montrer que on peut dominer $\mathbb{P}(x \leq T_n)$ par une fonction intégrable sur $]0, \infty[$. En déduire,

$$\lim_{n \rightarrow \infty} \mathbb{E}(T_n^+) = \mathbb{E}(G^+) = \frac{1}{\sqrt{2\pi}}.$$

Prouver la formule de Stirling: lorsque $n \rightarrow \infty$,

$$\frac{e^{-n} n^n \sqrt{n}}{n!} \sim \frac{1}{\sqrt{2\pi}}.$$

4.27 Une usine fabrique des pièces dont 3% ont des défauts.

(i) On prélève 1000 pièces au hasard. Quelle est la probabilité d'avoir plus de 50 pièces défectueuses? d'avoir entre 20 et 40 pièces défectueuses?

(ii) On veut 1950 pièces sans défaut. Par prudence on prélève 2000 au hasard. Quelle est la probabilité d'avoir suffisamment de pièces en bon état?

4.28. Deux sondages effectués l'un auprès de 800 Françaises, l'autre auprès de 800 Français donnet 51% de fumeuses et 49% de fumeurs. Est-il raisonnable d'en déduire que les femmes fument plus que les hommes en France?

4.29. Dans une population de 56680 familles lapins ayant 8 petits chacune (donc 429440 lapereaux au total) il y a 221023 mâles et 208417 femelles. On se demande si le nombre de lapereaux mâles est significativement plus élevé que celui de lapereaux femelles? Autrement dit on voudrait tester l'hypothèse : les chances sont égale d'avoir un mâle ou une femelle avec un risque d'erreur de 5%.

4.30. Suivant l'hypothèse de Mendel sur la couleur des yeux (bleu récessif, marron dominant) une personne prise au hasard a une chance sur quatre d'avoir les yeux bleus. Si cette hypothèse est valide, combien de personnes doit-on observer pour être quasi-certain avec une probabilité de 99,8% que la proportion de personnes aux yeux marron sera comprise entre 0,7 et 0,8?

Chapitre 5

Sujets d'examens 2001-2003

Sujet d'examen partiel

Licence de Mathématiques - Probabilités
 mardi 27 mars 2001 - durée 3 heures - calculatrice interdite
 documents (cours manuscrit ou photocopie) autorisés

Exercice I.

Sur l'espace de probabilité $(]0,1], \mathcal{B}(]0,1]), \lambda$, où λ note la mesure de Lebesgue sur $]0,1]$, on considère la variable aléatoire $S(\omega) = [1/\omega]$, où $[\cdot]$ note la fonction partie entière.

- 1) Trouver la loi de S (on pourra chercher les valeurs possibles de S et les probabilités avec lesquelles sont prises ces valeurs);
- 2) Étudier l'existence des moments d'ordre un et deux de S .
- 3) Que vaut $P(S \geq 100)$?

Exercice II.

Le vecteur aléatoire (X, Y) a la densité de probabilité donnée par

$$f_{X,Y}(x,y) = \begin{cases} e^{-x-y}, & x > 0, y > 0 \\ 0 & \text{sinon.} \end{cases}$$

Trouver :

- 1) les densités marginales de X et de Y et leurs lois;
- 2) les fonctions de répartition $F_{(X,Y)}$, F_X et F_Y du vecteur et de ses coordonnées;
- 3) les fonctions caractéristiques $\varphi_{(X,Y)}$, φ_X et φ_Y du vecteur et de ses coordonnées;
- 4) l'espérance et la matrice de covariance du vecteur (X, Y) ;
- 5) la densité du vecteur $(Z, W) = (X, X + Y)$ et les lois des variables aléatoires $X + Y$ et $U = X/(X + Y)$ (on pourra calculer successivement les densités de $(Z, W) = (X, X + Y)$ et $(U, V) = (Z/W, W) = (X/(X + Y), X + Y)$, ensuite déduire les marginales en question).

Exercice III.

Une urne contient des boules rouges et noires. La proportion de boules rouges est $p \in]0,1[$. On effectue des tirages avec remise dans l'urne en notant à chaque fois la couleur obtenue.

On dit avoir obtenu un "succès" si une boule rouge est sortie. On cesse les tirages après avoir noté l'obtention de n boules rouges, $n \geq 1$ (n succès). Soit T la variable aléatoire nombre total de boules tirées.

- 1) Dire quel est l'ensemble de valeurs possibles de T .
- 2) Calculer $P(T = n)$ et $P(T = n + 1)$.
- 3) Donner une explication précise et succincte de la formule suivante :

$$P(T = k) = C_{k-1}^{n-1} p^n (1-p)^{k-n}.$$

pour k appartenant à l'ensemble trouvé au point 1).

- 4) Quelle loi obtient-on si $n = 1$?
- 5) Calculer $E(T)$, $\text{Var}(T)$. Calculer $E[(n-1)/(T-1)]$ lorsque $n \geq 2$.

Exercice IV.

Soient (Ω, \mathcal{A}, P) un espace de probabilité, \mathcal{T} une sous-tribu de \mathcal{A} et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle intégrable. On rappelle qu'une variable aléatoire réelle Z est \mathcal{T} -mesurable si :

$$Z^{-1}(B) \in \mathcal{T}, \text{ pour tout borélien } B \in \mathcal{B}(\mathbb{R}).$$

On admet qu'il existe une variable aléatoire réelle Z \mathcal{T} -mesurable, intégrable et satisfaisant :

$$E(Z \mathbb{1}_A) = E(X \mathbb{1}_A), \text{ pour tout } A \in \mathcal{T}. \quad (5.1)$$

Montrer que :

- 1) $E(Z) = E(X)$;
- 2) si $X \geq 0$ p.s. alors $Z \geq 0$ p.s. (on justifiera qu'on peut prendre dans (5.1) $A = \{Z < 0\}$ et on montrera que $P(A) = 0$);
- 3) la variable Z est unique p.s., c'est-à-dire que si \hat{Z} est une autre variable aléatoire réelle \mathcal{T} -mesurable, intégrable telle que :

$$E(\hat{Z} \mathbb{1}_A) = E(X \mathbb{1}_A), \text{ pour tout } A \in \mathcal{T}, \quad (5.2)$$

alors $Z = \hat{Z}$ p.s. (on justifiera qu'on peut prendre dans une égalité déduite à partir de (5.1) et (5.2), d'abord $A = \{Z > \hat{Z}\}$ et ensuite $A = \{Z < \hat{Z}\}$).

Exercice V.

Soit (Ω, \mathcal{A}, P) un espace de probabilité et soient $A, B \in \mathcal{A}$ deux événements quelconques. Montrer que :

$$|P(A \cap B) - P(A)P(B)| \leq \frac{1}{4}$$

(on pourra déduire une inégalité utile en étudiant le maximum de la fonction $x \mapsto x(1-x)$ définie sur $[0,1]$ à valeurs dans \mathbb{R}).

Sujet d'examen

Licence de Mathématiques - Probabilités
jeudi 31 mai 2001 - durée 3 heures - calculatrice interdite
documents (cours manuscrit ou photocopié) autorisés

Exercice I.

Un point aléatoire (X, Y) est uniformément distribué dans le carré \mathcal{R} de sommets $(1, 0)$, $(0, 1)$, $(-1, 0)$ et $(0, -1)$; ainsi la densité est de la forme $f_{(X, Y)}(x, y) = c \mathbb{1}_{\mathcal{R}}(x, y)$. Trouver la constante c et étudier l'indépendance et la corrélation des coordonnées X, Y .

Exercice II.

Soient $c > 0$ et $f, g : [0, 1] \rightarrow \mathbb{R}$ deux fonctions continues telles que pour tout $x \in]0, 1[$, $0 < f(x) \leq cg(x)$.

- 1) Soit $(U_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes uniformément distribuées sur $[0, 1]$. Justifier la convergence presque sûre et préciser les valeurs des limites des suites suivantes:

$$X_n = \frac{1}{n} \sum_{j=1}^n f(U_j) \quad \text{et} \quad Y_n = \frac{1}{n} \sum_{j=1}^n g(U_j).$$

- 2) Calculer les limites presque sûre des suites $1/Y_n$ et X_n/Y_n .
3) Justifier la convergence et calculer $\lim_{n \rightarrow \infty} E(X_n/Y_n)$.
4) En déduire

$$\lim_{n \rightarrow \infty} \underbrace{\int_0^1 \dots \int_0^1}_{n \text{ fois}} \frac{f(x_1) + \dots + f(x_n)}{g(x_1) + \dots + g(x_n)} dx_1 \dots dx_n = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

- 5) Calculer :

$$\lim_{n \rightarrow \infty} \underbrace{\int_0^1 \dots \int_0^1}_{n \text{ fois}} \frac{x_1^2 + \dots + x_n^2}{x_1 + \dots + x_n} dx_1 \dots dx_n$$

Exercice III.

Soient X_1, \dots, X_n variables aléatoires strictement positives indépendantes et de même loi. Calculer, pour $k \in \{1, \dots, n\}$,

$$E\left(\frac{X_1 + \dots + X_k}{X_1 + \dots + X_n}\right).$$

On pourra justifier d'abord que les variables $\frac{X_j}{X_1 + \dots + X_n}$, $j = 1, \dots, n$, ont la même loi.

Exercice IV.

Soient D_1, D_2, \dots des variables aléatoires indépendantes de même loi $\mathcal{E}(\lambda)$, $\lambda > 0$. On regarde les variables D_j comme les durées de temps entre deux arrivées de clients devant un guichet. On s'intéresse au nombre N_t de clients arrivés jusqu'au moment t .

On définit $N_0 = 0$ et pour $t > 0$:

$$N_t = \max\{n : D_0 + D_1 + \dots + D_n \leq t\}, \quad \text{où } D_0 \equiv 0.$$

- 1) Donner la formule de la densité de la variable $D_1 + \dots + D_n$. S'agit-il d'une loi remarquable? On pourra utiliser la fonction caractéristique ou le produit de convolution.
- 2) Justifier l'égalité $P(N_t = 0) = P(D_1 > t)$, pour $t > 0$, et calculer cette probabilité.
- 3) Justifier que, pour $n \geq 1$ entier,

$$\{N_t = n\} = \{D_1 + \dots + D_n \leq t \text{ et } D_1 + \dots + D_{n+1} > t\}.$$

En déduire que, pour $n \geq 1$ entier,

$$P(N_t = n) = P(D_1 + \dots + D_n \leq t) - P(D_1 + \dots + D_{n+1} \leq t).$$

- 4) Montrer que, pour $t > 0$ et pour $n \geq 0$ entier,

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Quelle est la loi de N_t ? Calculer le nombre moyen de clients arrivés jusqu'au temps t .
Que vaut et que signifie $E(N_t/t)$? Montrer (par calcul et passage à la limite) que

$$\lim_{t \rightarrow \infty} E\left(\frac{t}{1 + N_t}\right) = E(D_1).$$

- 5) Calculer la limite en loi de la suite $(\frac{N_n}{\sqrt{n}} - \lambda\sqrt{n})_{n \geq 1}$, lorsque $n \rightarrow \infty$.

Sujet d'examen de septembre

Licence de Mathématiques - Probabilités

jeudi 6 septembre 2001 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou photocopié) autorisés

Exercice I.

Soient X et Y deux variables aléatoires indépendantes de même loi exponentielle d'espérance 1.

- 1) Trouver la densité de la variable aléatoire $Z = X - Y$ (on pourra commencer par étudier le couple aléatoire (X, Z)). S'agit-il d'une loi remarquable?
- 2) Effectuer le calcul de la fonction caractéristique φ_X . Que vaut φ_Y ? En déduire la fonction caractéristique de Z (ne pas calculer d'intégrale mais utiliser les fonctions φ_X et φ_Y).
- 3) Trouver la fonction de répartition F_Z .
- 4) Calculer l'espérance et la variance de Z .
- 5) Soit $W \sim \mathcal{U}_{[0, e^2 - 1]}$ une variable aléatoire indépendante du couple (X, Y) . Dire pourquoi W est indépendante de Z . En déduire que $\text{Var}(Z) = (e^2 - 1)\text{E}(e^{-W|Z|})$ (on pourra calculer l'espérance du membre de droite).

Exercice II.

Soient X et Y deux variables aléatoires discrètes telles que

$$P(X = n, Y = m) = \frac{a}{(n + m - 1)(n + m)(n + m + 1)}, \quad n, m = 1, 2, 3, \dots$$

- 1) Trouver les lois marginales de X et Y et calculer ensuite la valeur de la constante a .
- 2) Calculer $P(\frac{1}{2} < X < 3)$ et $P(Y \geq r)$, où $r \in \mathbb{N}^*$.
- 3) Pour quelles valeurs de p , $X \in L^p$? X est-elle intégrable?
- 4) X et Y sont-elles indépendantes? non corrélées?
- 5) Calculer $P(X + Y = r)$ pour r entier supérieur à une valeur que l'on indiquera.

Exercice III.

Soit $(X_n)_{n \geq 1}$ et X des variables aléatoires telles que la série $\sum_{n \geq 1} \|X_n - X\|_p^p$ soit convergente pour un $p \geq 1$. Prouver que $X_n \xrightarrow{\text{p.s.}} X$, lorsque $n \rightarrow \infty$.

Exercice IV.

On considère $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi donnée par :

$$P(X_n = \ell) = \frac{e^{-1}}{(\ell + 1)!}, \quad \ell = -1, 0, 1, 2, 3, \dots$$

- 1) Trouver les lois de $1 + X_1$ et de $n + S_n$, où $S_n = X_1 + \dots + X_n$.
- 2) Peut-on utiliser le théorème central limite? Énoncer et appliquer ce résultat.
- 3) En déduire que les suites $P(S_n \leq 0)$ et $P(S_n \geq 0)$ convergent vers $\frac{1}{2}$ et que :

$$\lim_{n \rightarrow \infty} \left(e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \right) = \frac{1}{2}.$$

Sujet d'examen partiel

Licence de Mathématiques - Probabilités

mardi 26 mars 2002 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou polycopié) autorisés

Exercice I.

- 1) Sur l'espace de probabilité $(]0,1], \mathcal{B}(]0,1]), \lambda)$, où λ est la mesure de Lebesgue sur $]0,1]$, on considère la variable aléatoire $X(\omega) = -2 \ln \omega$.
 - a) Trouver la fonction de répartition de X et montrer qu'il s'agit d'une loi exponentielle de variance 4.
 - b) Montrer que pour tous $u, v > 0$, $P(X > u + v) = P(X > u)P(X > v)$.
- 2) Soit Y une variable aléatoire sur un espace de probabilité (Ω, \mathcal{A}, P) de même loi que X . On pose $Z = 1 + [Y]$, où $[\cdot]$ note la fonction partie entière.
 - a) Calculer $P(Z = 1)$ et $P(Z = 2)$. Trouver la loi de Z .
 - b) Sur une cible de surface 1 ayant la forme d'un disque, la partie centrale (noire) est de rayon $\sqrt{p/\pi}$, $0 < p < 1$. On tire T fois sur cette cible jusqu'à ce qu'on touche la première fois la partie centrale. Pour quel choix particulier de p les variables aléatoires T et Z ont la même loi?
 - c) Énoncer et prouver une égalité pour Z similaire à celle du point 1b) pour $u, v \in \mathbb{N}^*$.
 - d) Calculer $G_Z(s) = E(s^Z)$. En déduire $E(Z)$ et $\text{Var}(Z)$.
- 3) On rappelle que Y est de même loi que X et on note $W = Y - [Y]$.
 - a) Montrer que $\{W \leq t\} = \cup_{n \geq 0} \{n \leq Y \leq n+t\}$. En déduire la fonction de répartition de W . Possède-t-elle une densité?
 - b) Calculer $\varphi_W(t) = E(e^{itW})$ et étudier l'existence des moments de la variable W .

Exercice II.Soient (Ω, \mathcal{A}, P) un espace de probabilité et $\{A_n : n \geq 1\}$ une suite d'événements.

- 1) Montrer que si $B_n := \cup_{m \geq n} A_m$, la suite d'événements $\{B_n : n \geq 1\}$ est décroissante et converge, lorsque $n \rightarrow \infty$, vers l'événement $\{\sum_{m=1}^{\infty} \mathbb{1}_{A_m} = \infty\}$.
- 2) Supposons que la série $\sum_{n \geq 1} P(A_n)$ est convergente. Calculer $\lim_{n \rightarrow \infty} P(B_n)$.
- 3) En utilisant les points précédents, vérifier l'implication

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies \sum_{n=1}^{\infty} \mathbb{1}_{A_n} < \infty \text{ p.s.}$$

Exercice III.Soient X et Y deux variables aléatoires réelles admettant des moments d'ordre deux, centrées, réduites, de covariance ρ .

- 1) Montrer que $|\rho| \leq 1$. A quelle condition sur X et Y a-t-on l'égalité dans cette inégalité?
- 2) Justifier que $\max\{X^2, Y^2\}$ est une variable aléatoire. Pourquoi est-elle intégrable?
- 3) Montrer que $E[\max\{X^2, Y^2\}] \leq 1 + \sqrt{1 - \rho^2}$ (on remarquera que $\max\{u, v\} = \frac{1}{2}(u + v + |u - v|)$).

Exercice IV.

- 1) Soit le vecteur aléatoire (X, Y) de fonction de répartition $F(s, t) = P(X \leq s, Y \leq t)$. Soient les réels $a < b$, $c < d$ et on considère les événements $A = \{a < X \leq b, Y \leq d\}$ et $B = \{X \leq b, c < Y \leq d\}$.

- a) Exprimer les probabilités $P(A)$, $P(B)$ et $P(A \cup B)$ à l'aide de la fonction F (on pourra faire un dessin).
 b) En déduire $P(A \cap B)$ et montrer que

$$0 \leq F(b, d) - F(a, d) - F(b, c) + F(a, c) = P(a < X \leq b, c < Y \leq d).$$

- 2) On considère $F : \mathbb{R}^2 \rightarrow [0, 1]$ définie par

$$F(s, t) = \left(\frac{1}{2} e^{s/2} \mathbb{1}_{\{s < 0\}} + \left(1 - \frac{1}{2} e^{-s/2}\right) \mathbb{1}_{\{s \geq 0\}} \right) \left(\frac{1}{2} + \frac{1}{\pi} \arctan \frac{t}{2} \right), s, t \in \mathbb{R}.$$

- a) Montrer qu'il existe un couple aléatoire (X, Y) ayant pour fonction de répartition F (on pourra dériver F , sur le domaine où elle est dérivable et voir si on peut obtenir une densité de probabilité).
 b) Trouver, lorsqu'elles existent, f_X , f_Y . Que peut-on remarquer?
 c) En utilisant des résultats sur deux lois remarquables, calculer φ_X et φ_Y . En déduire (quand ces quantités existent) $E(X)$, $E(Y)$ et $\text{Var}(X)$, $\text{Var}(Y)$.
 d) Quelles sont les lois de $X/2$, $Y/2$ et $2/Y$?

Sujet d'examen

Licence de Mathématiques - Probabilités

lundi 10 juin 2002 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou photocopié) autorisés

Exercice I.Soient X_1, X_2, X_3, X_4 quatre variables aléatoires réelles indépendantes de même loi $\mathcal{N}(0,1)$.

1. Déterminer la fonction caractéristique de $X_1 X_4$.
2. Trouver la fonction caractéristique et loi de la variable aléatoire $D = \det \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$.
3. En déduire la loi de la variable $|D|$.

Exercice II.Soient X et Y deux variables aléatoires réelles indépendantes de même loi, centrées et réduites. On notera φ la fonction caractéristique commune.

1. Supposons que la loi commune est $\mathcal{N}(0,1)$. Montrer que $\frac{X+Y}{\sqrt{2}} \sim X \sim Y$.
2. On ne suppose plus connue la loi commune mais seulement $\frac{X+Y}{\sqrt{2}} \sim X \sim Y$.
 - a) Montrer que $\varphi^2(t/\sqrt{2}) = \varphi(t)$. Que vaut $\varphi^4(t/2)$?
 - b) En déduire que si X_1, X_2, X_3, X_4 sont quatre variables aléatoires indépendantes de même loi que X et Y , alors $\frac{X_1 + X_2 + X_3 + X_4}{2} \sim X_1 \sim X_2 \sim X_3 \sim X_4$.
 - c) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi que X et Y . On note $S_n = X_1 + \dots + X_n$. Quelle est la loi de $S_{2^n} / \sqrt{2^n}$?
 - d) En utilisant un résultat important de convergence en loi, énoncer et prouver la phrase réciproque de 1.

Exercice III.Supposons que les variables $(X_n)_{n \geq 1}$ sont indépendantes de loi $\mathcal{U}_{]-1,1[}$.

1. Montrer que $\frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow \frac{1}{3}$ en probabilité quand $n \rightarrow \infty$.
2. On note \mathbf{X} le vecteur aléatoire (X_1, \dots, X_n) et $\|\mathbf{X}\|_{(n)}$ sa norme euclidienne $(\sum_{i=1}^n X_i^2)^{1/2}$. Que vaut la limite en probabilité de $\|\mathbf{X}\|_{(n)} / \sqrt{n}$?
3. Pour $\varepsilon > 0$ on note $C_{n,\varepsilon}$ la couronne $\{\mathbf{x} \in \mathbb{R}^n : \sqrt{\frac{1}{3}} - \varepsilon < \frac{\|\mathbf{x}\|_{(n)}}{\sqrt{n}} < \sqrt{\frac{1}{3}} + \varepsilon\}$ et soit D_n le cube $\{\mathbf{x} \in \mathbb{R}^n : \max_{1 \leq i \leq n} |x_i| < 1\}$. Montrer que

$$\lim_{n \rightarrow \infty} \frac{1}{2^n} \lambda_n(C_{n,\varepsilon} \cap D_n) = 1,$$

où λ_n est la mesure de Lebesgue sur \mathbb{R}^n .**Exercice IV.**Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{N} , indépendantes, de même loi et de fonction génératrice commune : $G(s) = E(s^{X_i})$, $|s| \leq 1$, $i \geq 1$. On pose $T_n = \sum_{i=1}^n X_i$. On considère N une autre variable aléatoire à valeurs dans \mathbb{N} , indépendante de la suite $(X_i)_{i \geq 1}$,

ayant la fonction génératrice $G_N(s) = E(s^N)$. On note T_N la variable aléatoire définie par $T_N(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega)$.

1. Montrer que la fonction génératrice de T_N est donnée par

$$G_{T_N}(s) = (G_N \circ G)(s) = G_N(G(s)), \quad |s| \leq 1.$$

On pourra d'abord vérifier l'égalité $s^{T_N} = \sum_{n \geq 1} s^{T_n} \mathbb{1}_{\{N=n\}}$.

2. En déduire que $E(T_N) = E(N)E(X_1)$ et $\text{Var}(T_N) = \text{Var}(N)[E(X_1)]^2 + E(N)\text{Var}(X_1)$.
3. Le nombre d'accidents de moto en un an sur une portion de route est une variable aléatoire de Poisson de paramètre $\theta > 0$. A chaque accident, le motocycliste est blessé avec une probabilité $p \in]0,1[$ (et sort indemne avec une probabilité $1 - p$). Tous ces phénomènes aléatoires sont indépendants les uns des autres. Trouver la loi du nombre de motocyclistes blessés en un an et aussi le nombre moyen.

Exercice V.

On note $Z = \frac{X}{|Y|}$ et $Z_n = \frac{X}{\frac{1}{n} + |Y|}$, $n \geq 1$, où $X \sim Y \sim \mathcal{N}(0,1)$ indépendantes.

1. Montrer que Z est définie presque sûrement et que $\lim_{n \rightarrow \infty} Z_n = Z$ presque sûrement.
2. Vérifier que $Z \notin L^1$ et que $Z_n \in L^1$ pour tout $n \geq 1$. Calculer $E(Z_n)$.
3. Montrer que Z_n ne converge pas vers Z dans L^1 quand $n \rightarrow \infty$.
La suite $(Z_n)_{n \geq 1}$ est-elle uniformément intégrable?
4. Prouver que Z est une variable aléatoire de loi de Cauchy.

Sujet d'examen de septembre

Licence de Mathématiques - Probabilités

mercredi 11 septembre 2002 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou photocopié) autorisés

Exercice I.

Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{N}^* et de même loi donnée par la suite $p_n = 2^{-n}$, $n = 1, 2, \dots$

- 1) Identifier la loi commune et calculer sa fonction caractéristique. Calculer pour tout t réel, $P(X > t)$. Quelle est la loi de $\min\{X, Y\}$?
- 2) Calculer $P(X \geq kY)$ pour un entier positif donné k . Que vaut $P(X \geq Y)$?
- 3) Calculer $P(X = rY)$ pour un rationnel positif r . Que valent $P(X = Y)$, $P(X > Y)$ et $P(X < Y)$?
- 4) Calculer $P(X \text{ divise } Y)$.

Exercice II.

Soit ρ une constante vérifiant $-1 < \rho < 1$. (X, Y) est un couple aléatoire de densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right).$$

- 1) Montrer que les variables X et Y ont la même loi et que le coefficient de corrélation entre X et Y est ρ .
- 2) Calculer le produit des densités marginales f_X et f_Y . En déduire une condition nécessaire et suffisante qui porte sur ρ pour que les variables X et Y soient indépendantes.
- 3) Montrer que les variables X et $Z = (Y - \rho X)/\sqrt{1-\rho^2}$ sont indépendantes de même loi $\mathcal{N}(0,1)$. En déduire $P(X > 0, Y > 0) = \frac{1}{4} + \frac{1}{2\pi} \arctan(\rho/\sqrt{1-\rho^2}) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$.
- 4) Supposons que les variables X et Y sont non-corrélées. Calculer de deux façons la probabilité $P(X > 0, Y > 0)$.

Exercice III.

La probabilité qu'une pièce montre pile lors d'un lancer est $p \in]0, 1[$. On lance cette pièce $n \in \mathbb{N}^*$ fois et on note respectivement P et F les nombres de piles et de faces obtenus lors de ces lancers.

- 1) On fixe le nombre de lancers. Montrer que $P \sim \mathcal{B}(n, p)$. Que vaut $P + F$? Calculer $G_{P,F}(s,t) := E[s^P t^F]$. En déduire les fonctions génératrices $G_P(s)$ et $G_F(t)$. Calculer ensuite les fonctions caractéristiques des variables P et F et étudier leur indépendance.
- 2) On fait varier n et on notera les variables P et F par P_n et F_n . Justifier pourquoi P_n/n tend en probabilité vers p , lorsque $n \rightarrow \infty$. En déduire que, lorsque $n \rightarrow \infty$ $P(2p - 1 - \varepsilon \leq \frac{1}{n}(P_n - F_n) \leq 2p - 1 + \varepsilon) \rightarrow 1$ (utiliser la valeur de la somme $P_n + F_n$).
- 3) Montrer que pour tout $n \in \mathbb{N}^*$, $E\left(\frac{1}{1+P_n}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}$ (on pourra utiliser l'égalité $\frac{1}{1+P_n} = \int_0^1 t^{P_n} dt$).
- 4) On suppose enfin que $p = \frac{\lambda}{n}$ ($0 < \lambda < n$). Montrer alors que P_n converge en loi, lorsque $n \rightarrow \infty$, vers une variable aléatoire de loi de Poisson de paramètre λ . Calculer par deux méthodes la limite $\lim_{n \rightarrow \infty} E\left(\frac{1}{1+P_n}\right)$.

Sujet d'examen partiel

Licence de Mathématiques - Probabilités

samedi 12 avril 2003 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou polycopié) autorisés

Exercice I. Soit $f : \mathbb{R}_+^* \rightarrow \mathbb{R}_+$ une densité de probabilité.

- 1) On définit $p(x,y) := \begin{cases} \frac{f(x+y)}{x+y}, & \text{si } x > 0, y > 0, \\ 0, & \text{sinon.} \end{cases}$

Montrer qu'il existe un couple aléatoire (X,Y) ayant p pour densité de probabilité.

- 2) Soit W une variable aléatoire de densité f . Supposons que $\int_0^\infty x^2 f(x) dx < \infty$, c'est-à-dire que $W \in L^2$. On note m et respectivement σ^2 l'espérance et la variance de W . Calculer, en fonction de m et σ^2 , le vecteur espérance et la matrice de covariance du couple aléatoire (X,Y) .
- 3) Calculer $E(X+Y)$ et $\text{Var}(X+Y)$. Trouver la densité de la variable aléatoire $X+Y$. (On pourra exprimer d'abord la densité du couple $(X+Y,Y)$ à l'aide de la fonction p .)
- 4) Soit $f(x) := \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \mathbb{1}_{x>0}$. Vérifier que f est une densité de probabilité. Dans la suite on supposera que la variable aléatoire W est de densité f . Calculer $E(W)$ et $\text{Var}(W)$. Dans ce cas, les variables X et Y , sont-elles corrélées? indépendantes?
- 5) Une autre variable aléatoire U est de loi uniforme sur $[0,1]$. Quelle est la loi de la variable $V = \mathbb{1}_{\{U \leq \frac{1}{2}\}} - \mathbb{1}_{\{U > \frac{1}{2}\}}$?
- 6) Supposons maintenant que les variables W et U soient indépendantes. Justifier que W et V sont aussi des variables aléatoires indépendantes. Calculer $E(VW)$ et $\text{Var}(VW)$. Quelle est la loi de VW ? (On pourra écrire sa fonction de répartition et sa densité.)
- 7) Enfin, on note $Z_a = -(VW) \mathbb{1}_{\{|VW| \leq a\}} + (VW) \mathbb{1}_{\{|VW| > a\}}$, $a > 0$ fixé. Pour ϕ une fonction borélienne positive quelconque calculer $E[\phi(Z_a)]$. (On pourra utiliser le fait que $\phi(Z_a)$ est une fonction de VW .) En déduire la densité de Z_a . Pourquoi $P(Z_a + VW = 0) > 0$?

Exercice II. Soit $g : [0,1] \rightarrow [0,1]$ une fonction continue et on note $I = \int_0^1 g(x) dx$. Supposons que le couple aléatoire (U,V) est de loi uniforme dans le carré $[0,1] \times [0,1]$. Quelles sont les lois de U et de V ? Montrer que $V > 0$ presque sûrement et que $P(\frac{g(U)}{V} \geq 1) = I$. On pose

$$X = \mathbb{1}_{\{V \leq g(U)\}}, Y = g(U) \text{ et } Z = \frac{1}{2}[g(U) + g(1-U)].$$

Montrer que $E(X) = E(Y) = E(Z) = I$ et que $\text{Var}(X) \geq \text{Var}(Y) \geq \text{Var}(Z)$.

Exercice III. Soit n un entier strictement positif fixé.

- 1) On possède une pièce qui montre pile avec probabilité $p \in]0,1[$. On lance cette pièce n fois et on note à chaque fois le résultat. Soit N le nombre de piles obtenus lors de ces n lancers. Expliquer pourquoi $N \sim \mathcal{B}(n,p)$.
- 2) Un probabiliste possède n pièces identiques dont chacune montre pile avec probabilité $q \in]0,1[$. Il les lance toutes en même temps; il reprend les pièces qui ont donné pile et il les lance une deuxième fois en même temps. Soit N' le nombre de pièces qui ont donné pile deux fois. Que doit valoir q pour que les variables N et N' aient la même loi?

- 3) Calculer la fonction caractéristique, l'espérance et la variance de $\frac{N}{n}$.
- 4) Soit f une fonction réelle définie et continue sur $[0,1]$. Montrer que $E[f(\frac{N}{n})]$ est un polynôme en p de degré n , noté $B_n(p)$.
- 5) Si $\|f\|_\infty$ désigne la norme uniforme de f sur $[0,1]$, montrer que, pour tout $p \in [0,1]$,

$$|f(p) - B_n(p)| \leq E \left[|f(\frac{N}{n}) - f(p)| \mathbf{1}_{\{|\frac{N}{n} - p| \geq \varepsilon\}} \right] + E \left[|f(\frac{N}{n}) - f(p)| \mathbf{1}_{\{|\frac{N}{n} - p| < \varepsilon\}} \right]$$

$$\leq 2\|f\|_\infty P(|\frac{N}{n} - p| \geq \varepsilon) + \varrho(\varepsilon; f) P(|\frac{N}{n} - p| < \varepsilon).$$

On a noté ici par $\varrho(\varepsilon; f) := \sup\{|f(t) - f(s)| : 0 \leq s < t \leq 1 \text{ avec } t - s < \varepsilon\}$ le module de continuité de f .

- 6) Appliquer l'inégalité de Tchebychev pour montrer que pour tout $p \in [0,1]$ et tout $\varepsilon > 0$,

$$|f(p) - B_n(p)| \leq \frac{\|f\|_\infty}{2n\varepsilon^2} + \varrho(\varepsilon; f).$$

(On pourra justifier et utiliser l'inégalité $p(1-p) \leq \frac{1}{4}$.)

- 7) En déduire que $\lim_{n \rightarrow \infty} \|f - B_n\|_\infty = 0$. (On justifiera soigneusement les passages à la limite.) Quel résultat d'approximation concernant des fonctions continues a-t-on obtenu?

Sujet d'examen

Licence de Mathématiques - Probabilités
 mercredi 4 juin 2003 - durée 3 heures - calculatrice interdite
 documents (cours manuscrit ou polycopié) autorisés

Exercice I. Soient $a, b > 0$ et on désigne par Z_a et Z_b deux variables aléatoires indépendantes de lois $\gamma(a, 1)$ et $\gamma(b, 1)$ respectivement.

- 1) Trouver la densité du vecteur $(U, V) := \left(Z_a + Z_b, \frac{Z_a}{Z_a + Z_b} \right)$.
- 2) Les variables U, V sont-elles indépendantes? Montrer que leurs lois sont deux lois remarquables.
- 3) On suppose dans cette question que $a = b = \frac{1}{2}$. Montrer que $1/V$ a la même loi que la variable aléatoire $1 + C^2$, où C suit une loi de Cauchy. On pourra d'abord montrer que pour toute fonction borélienne positive g , $E[g(1/V)] = E[g(1 + C^2)]$ et ensuite conclure.

Exercice II. Soit (X, Y) un couple aléatoire tel que sa fonction caractéristique $E[e^{isX + itY}]$ est donnée par $\varphi_{X, Y}(s, t) = \exp[\alpha(e^{is} - 1) + \beta(e^{it} - 1) + \gamma(e^{i(s+t)} - 1)]$, avec $\alpha, \beta > 0$ et $\gamma \geq 0$.

- 1) Dédire les fonctions caractéristiques des variables marginales X et Y , mais aussi celle de la variable aléatoire $X + Y$.
- 2) Montrer que si $\gamma = 0$ alors les variables X et Y sont indépendantes. La réciproque est-elle vraie?
- 3) On continue de supposer que $\gamma = 0$. Calculer $P(X = k \mid X + Y = n)$, pour $n \in \mathbb{N}^*$ et $k \in \{0, \dots, n\}$. Montrer que la loi conditionnelle de X sachant $\{X + Y = n\}$ est une loi binomiale dont on indiquera les paramètres.
- 4) On suppose que $\gamma = 1$. Trouver la fonction caractéristique de la variable aléatoire $Z_\alpha := (X - E(X))/\sqrt{\text{Var}(X)}$. Peut-on indiquer la limite en loi de Z_α lorsque $\alpha \rightarrow \infty$? On pourra utiliser le développement limité à l'ordre deux de e^{ix} au voisinage de $x = 0$.

Exercice III. Soit $\{X_n : n \geq 1\}$ une suite de variables aléatoires indépendantes identiquement distribuées de loi de densité $f_\lambda(x) = e^{-\lambda x} \mathbb{1}_{[\lambda, \infty[}(x)$, avec $\lambda \geq 0$.

- 1) Trouver l'espérance et la variance communes des variables X_n .
- 2) Calculer la limite ℓ_1 de la suite $\frac{1}{n} \sum_{i=1}^n X_i$ et la limite ℓ_2 de la suite $\frac{1}{n} \sum_{i=1}^n X_i^2$, lorsque $n \rightarrow \infty$, et préciser quels types de convergences ont lieu. On prendra dans la suite les convergences au sens le plus fort.
- 3) Que vaut la limite de $\left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2$? On justifiera le résultat.
- 4) Combien de termes y a-t-il dans la somme $\sum_{1 \leq i < j \leq n} X_i X_j$? Calculer $E(X_i X_j)$ pour $i < j$. Peut-on appliquer la loi forte des grands nombres pour calculer la limite de la suite $\frac{1}{C_n^2} \sum_{1 \leq i < j \leq n} X_i X_j$, lorsque $n \rightarrow \infty$?
- 5) Vérifier que $\frac{1}{C_n^2} \sum_{1 \leq i < j \leq n} X_i X_j = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 - \frac{1}{n(n-1)} \sum_{i=1}^n X_i^2$.
- 6) Utiliser le résultat précédent pour calculer la limite de la suite $\frac{1}{C_n^2} \sum_{1 \leq i < j \leq n} X_i X_j$, lorsque $n \rightarrow \infty$. On justifiera soigneusement les passages à la limite dans les opérations avec des suites de variables aléatoires.

Exercice IV. Soient X et Y deux variables aléatoires à valeurs dans $\{0,1\}$. On note $a = P(X = 0, Y = 0)$, $b = P(X = 0, Y = 1)$, $c = P(X = 1, Y = 0)$ et $d = P(X = 1, Y = 1)$, avec $a + b + c + d = 1$.

- 1) Trouver une condition nécessaire et suffisante pour que X et Y soient non-corrélées.
- 2) Trouver une condition nécessaire et suffisante pour que X et Y soient indépendantes. Montrer qu'une seule relation (sur les quatre) suffit pour exprimer la condition. On pourra d'abord vérifier que si une relation a lieu, alors $ad = bc$. Comparer avec le résultat du point précédent.
- 3) On suppose que $a = b = c = d = \frac{1}{4}$. Quelles sont les lois des variables X et Y ? Sont-elles indépendantes?
- 4) On continue de supposer que $a = b = c = d = \frac{1}{4}$. Calculer $P(X + Y = 0)$, $P(|X - Y| = 0)$ et $P(X + Y = 0, |X - Y| = 0)$. Que peut-on déduire? Les variables aléatoires $X + Y$ et $|X - Y|$ sont-elles non-corrélées?

Sujet d'examen de septembre

Licence de Mathématiques - Probabilités

mardi 2 septembre 2003 - durée 3 heures - calculatrice interdite

documents (cours manuscrit ou photocopié) autorisés

Exercice I. Soit X une variable aléatoire discrète ayant ses valeurs dans \mathbb{N}^* . Montrer que

$$E(X) = \sum_{n=0}^{\infty} P(X > n).$$

Application: Une urne contient b boules bleues et r boules rouges. On effectue des tirages au hasard en enlevant chaque fois la boule tirée, jusqu'à ce qu'on tire la première boule bleue. Montrer que le nombre moyen de tirages est $(b+r+1)/(b+1)$.

On pourra utiliser l'égalité combinatoire suivante: $\sum_{n=0}^r C_{n+b}^b = C_{r+b+1}^{b+1}$.

Exercice II. Trouver la densité de la variable aléatoire $Z = X+Y$, lorsque le couple aléatoire (X,Y) est de densité $f(x,y) = \frac{1}{3\sqrt{2\pi}}\sqrt{x+y}e^{-(x+y)/2}$, $x,y \geq 0$. Z a-t-elle une loi remarquable?

Exercice III. Soit $m \in \mathbb{N}^*$ et soient les variables aléatoires indépendantes X_1, \dots, X_m de même loi exponentielle de paramètre $\lambda > 0$.

- 1) Effectuer et justifier le calcul de la fonction caractéristique commune des variables X_i .
- 2) Trouver la fonction caractéristique et ensuite la loi de la variable $S = X_1 + \dots + X_m$.
- 3) On fixe $t > 0$ et soit $T \sim \mathcal{P}(\lambda t)$. Prouver l'égalité suivante: $P(T < m) = P(S > t)$.

Exercice IV. Soient X et Y deux variables aléatoires indépendantes à densité. À une fonction borélienne positive $\Psi : \mathbb{R}^2 \rightarrow [0, \infty[$, on associe

$$\psi_1(x) = E[\Psi(x,Y)] \text{ et } \psi_2(y) = E[\Psi(X,y)], x \in \mathbb{R}, y \in \mathbb{R}.$$

Vérifier la double égalité: $E[\Psi(X,Y)] = E[\psi_1(X)] = E[\psi_2(Y)]$.

Application: On suppose que les variables indépendantes sont telles que $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\mu)$, $\lambda, \mu > 0$. Calculer $P(X \leq Y)$. Que vaut $P(X = Y)$?

Exercice V. Sur l'espace de probabilité $([0,1], \mathcal{B}([0,1]), \lambda)$, où λ est la mesure de Lebesgue sur $[0,1]$, on considère la variable aléatoire $U(\omega) = \omega - 1$.

- 1) Trouver la fonction de répartition de U . Cette variable est-elle à densité? discrète?
- 2) Montrer que pour tout entier $p \geq 1$, $U \in L^p$ et calculer les moments μ_p d'ordre p de U .
- 3) Soit $\{U_n : n \geq 1\}$ une suite de variables aléatoires indépendantes de même loi que U et on pose

$$T_n = \frac{U_1^3 + \dots + U_n^3}{U_1^2 + \dots + U_n^2}, n \geq 1.$$

Montrer que la suite de variables aléatoires $\{T_n : n \geq 1\}$ converge presque sûrement vers une constante que l'on évaluera. On pourra étudier les deux suites $X_n = \frac{U_1^3 + \dots + U_n^3}{n}$ et $Y_n = \frac{U_1^2 + \dots + U_n^2}{n}$. Justifier soigneusement les passages à la limite.

- 4) Que vaut la limite en loi de la suite $Z_n = \sqrt{n}(X_n + \frac{1}{4})$. On pourra utiliser un résultat remarquable de convergence en loi.

Bibliographie

- [1] Barbe, Ph., Ledoux, M. (1998) *Probabilité*. Belin
- [2] Bertoin, J. (2001) *Probabilités : cours de licence*. Université Paris VI
- [3] Jacod, J. (2000) *Probabilités : cours de licence*. Université Paris VI
- [4] Neveu, J. (1995) *Probabilités*. École Polytechnique
- [5] Resnick, S.I. (2001) *A Probability Path*. Birkhäuser
- [6] Vallois, P. (2000) *Cours de Probabilités - licence*. Université de Nancy 1