

RAPPELS de PROBABILITÉ

Préparation à l'Agrégation Bordeaux 1

Année 2013 - 2014

Jean-Jacques Ruch et Marie-Line Chabanol

Table des Matières

Chapitre I. Introduction à la théorie des probabilités	5
1. Espaces probabilisés	5
2. Probabilités discrètes	6
3. Probabilités conditionnelles, événements indépendants	7
4. Lemme de Borel-Cantelli	8
Chapitre II. Variables aléatoires	11
1. Définitions	11
1.A. Variable aléatoire discrète	11
1.B. Variable aléatoire à densité	12
2. Tribu engendrée par une variable aléatoire	12
3. Fonction de répartition d'une variable aléatoire	13
4. Vecteur aléatoire	14
5. Variables aléatoires indépendantes	14
5.A. Définitions	14
5.B. Cas particuliers des variables aléatoires discrètes et à densité	16
5.C. Propriétés	16
5.D. Sommes de variables aléatoires indépendantes	16
5.E. Loi du 0-1	17
Chapitre III. Espérance	19
1. Définition	19
2. Moments d'une variable aléatoire	20
3. Propriétés	21
4. Covariance et corrélation	22
5. Fonction génératrice	24
6. Fonctions caractéristiques	25
Chapitre IV. Convergence de variables aléatoires	29
1. Les différentes notions de convergences	29
2. Loi des grands nombres ; théorème de Glivenko Cantelli	32
3. Lois faibles des grands nombres et démonstrations	34
4. La convergence en loi	35
5. Propriétés de la convergence en loi et lien avec les autres notions de convergence	37
6. Théorème central limite	38
7. Convergence vers la loi de Poisson	40
8. Appendice : démonstration de la loi forte des grands nombres	41
Bibliographie	47

CHAPITRE I

Introduction à la théorie des probabilités

1. Espaces probabilisés

Définition 1. Soit E un ensemble. On dit que $\mathcal{A} \subset \mathcal{P}(E)$ est une tribu (ou une σ -algèbre) si :

- $E \in \mathcal{A}$;
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
- $\forall n \geq 0, A_n \in \mathcal{A} \Rightarrow \bigcup_{n \geq 0} A_n \in \mathcal{A}$.

Les éléments de \mathcal{A} sont appelés *parties mesurables* (ou \mathcal{A} -mesurables). On dit alors que (E, \mathcal{A}) est un *espace mesurable*.

Il est facile de vérifier que l'ensemble vide appartient toujours à une tribu, et qu'une tribu est stable par intersection dénombrable.

Exemples :

- $\mathcal{P}(E), \{\emptyset, E\}$;
- $\mathcal{C} \subset \mathcal{P}(E)$ alors $\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu}, \mathcal{C} \subset \mathcal{A}} \mathcal{A}$;
- *tribu des boréliens* $\mathcal{B}(E)$: tribu engendrée par les ouverts d'un espace topologique ; si $E = \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par les $]a, b[$ ou $] -\infty, a[$ ($a \in \mathbb{R}$ ou \mathbb{Q}).

En probabilité on utilise un vocabulaire particulier :

- Ω est appelé espace des réalisations ou univers ; c'est aussi l'événement certain ;
- $\omega \in \Omega$ est appelé une réalisation ;
- \emptyset est l'événement impossible ;
- $A \in \mathcal{A}$ est un événement ;
- $A \cap B$ est l'événement A et B ;
- $A \cup B$ est l'événement A ou B .

Définition 2. On appelle mesure positive sur un espace mesurable (E, \mathcal{A}) une application $\mu : \mathcal{A} \rightarrow [0, +\infty]$ telle que :

- $\mu(\emptyset) = 0$
- pour toute famille $(A_n)_{n \geq 0}$ de parties mesurables disjointes : $\mu \left(\bigcup_{n \geq 0} A_n \right) = \sum_{n \geq 0} \mu(A_n)$.

On dit que μ est finie si $\mu(E) < +\infty$, et que μ est σ -finie s'il existe une suite croissante de parties mesurées $(A_n)_{n \geq 0}$ telle que $\bigcup_{n \geq 0} A_n = E$ et $\mu(A_n) < +\infty$ pour tout $n \geq 0$.

Définition 3. Une mesure positive est une probabilité si $\mu(E) = 1$.

On dit que $x \in E$ est un *atome* si $\mu(\{x\}) > 0$. Une mesure est diffuse si elle n'a pas d'atome.

Enfin, on appelle *mesure de Lebesgue*, l'unique mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $\lambda([a, b]) = b - a$.

On dit que $(\Omega, \mathcal{A}, \mathbb{P})$ est un *espace probabilisé* si (Ω, \mathcal{A}) est un espace mesurable sur lequel on a défini une probabilité \mathbb{P} .

Exemples :

- On lance un dé deux fois :

$$\Omega = \{1, 2, \dots, 6\}^2, \mathcal{A} = \mathcal{P}(\Omega), \mathbb{P}(A) = \frac{\text{Card}(A)}{36}$$

La probabilité choisie rend tous les tirages possibles équiprobables.

- On lance un dé jusqu'à obtenir 6. L'expérience définie ici peut amener à une infinité de lancers, on a donc :

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}^*}.$$

Les éléments de Ω sont donc les suites $\omega = (\omega_1, \omega_2, \dots)$ qui représentent les tirages successifs. La tribu \mathcal{A} sur Ω est la plus petite tribu qui rend mesurable tous les ensembles :

$$\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}$$

où $n \geq 1$ et $i_1, \dots, i_n \in \{1, \dots, 6\}$, et \mathbb{P} est l'unique probabilité sur Ω telle que

$$\mathbb{P}(\{\omega : \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\}) = \left(\frac{1}{6}\right)^n.$$

Une probabilité \mathbb{P} sur (Ω, \mathcal{T}) on peut vérifier les propriétés suivantes :

$$A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

$$\mathbb{P}\left(\bigcup_{1 \leq j \leq n} A_j\right) = \sum_{1 \leq j \leq n} (-1)^{k-1} \left(\sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \right), \text{ formule de Poincaré}$$

$$A_n \subset A_{n+1} \Rightarrow \mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n)$$

$$B_{n+1} \subset B_n \Rightarrow \mathbb{P}\left(\bigcap_{n \geq 0} B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n)$$

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

2. Probabilités discrètes

Il existe un cas particulier important lorsqu'on considère un univers Ω dénombrable. On prend $\mathcal{T} = \mathcal{P}(\Omega)$. Alors pour construire une probabilité sur (Ω, \mathcal{T}) il suffit de se donner les probabilités $\mathbb{P}(\{\omega\})$ pour toute réalisation $\omega \in \Omega$, telles que $\sum \mathbb{P}(\{\omega\}) = 1$. Il est alors immédiat de vérifier que pour tout événement A , $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.

Cette méthode s'applique en particulier lorsque Ω est fini et que les ω sont équiprobables, c'est-à-dire que

$$\mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}(\Omega)}$$

On obtient alors la règle d'équiprobabilité

$$\mathbb{P}(A) = \frac{\text{nombre d'issues favorables à } A}{\text{nombre d'issues possibles}} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$

Pour calculer en détails ces probabilités on utilise des résultats de combinatoire, dont on rappelle les plus classiques.

- Nombre de permutations d'un ensemble à n éléments : $n!$.
- Nombre de p -uplets dans un ensemble à n éléments : n^p .
- Nombre de p -uplets d'éléments distincts dans un ensemble à n éléments : $A_n^p = n(n-1)\dots(n-p+1)$.
- Nombre de parties d'un ensemble à n éléments : 2^n
- Nombre de parties à p éléments dans un ensemble à n éléments : $C_n^p = \binom{n}{p}$.

3. Probabilités conditionnelles, événements indépendants

Le fait de connaître un événement, de savoir qu'il s'est produit, peut modifier la probabilité d'un autre événement. il est donc important de pouvoir définir une nouvelle probabilité.

Définition 4. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et A un événement de probabilité non nulle. On peut définir une nouvelle probabilité sur (Ω, \mathcal{T}) , appelée probabilité conditionnelle sachant A , en posant pour tout $B \in \mathcal{T}$,

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

On peut facilement voir que l'on a

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

que l'on généralise en

$$\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

De plus, si $\mathbb{P}(A^c) > 0$ alors on a

$$\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)$$

et la formule de Bayes

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c)}.$$

Il est possible de généraliser ces dernières relations lorsqu'on considère une partition $(A_n)_{n \in \mathbb{N}}$ de Ω ($A_k \cap A_l = \emptyset$, $\cup A_n = \Omega$ et $\mathbb{P}(A_n) > 0$) :

$$\mathbb{P}(B) = \sum_{n \geq 0} \mathbb{P}(A_n)\mathbb{P}(B|A_n) \quad (\text{formule des probabilités totales})$$

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n)\mathbb{P}(B|A_n)}{\sum_{n \geq 0} \mathbb{P}(A_n)\mathbb{P}(B|A_n)} \quad (\text{formule de Bayes})$$

On peut remarquer que lorsque $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ alors $\mathbb{P}(B|A) = \mathbb{P}(B)$, autrement dit le fait de savoir que A est réalisé ne donne pas d'information sur la réalisation ou non de l'événement B ; d'où la définition suivante.

Définition 5. On considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Si $A, B \in \mathcal{A}$ sont deux événements, on dit que A et B sont indépendants si

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

Plus généralement on a :

Définition 6. On dit que n événements A_1, \dots, A_n sont indépendants si, pour tout sous-ensemble non vide $\{j_1, \dots, j_p\}$ de $\{1, \dots, n\}$, on a

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \dots \mathbb{P}(A_{j_p}).$$

Attention ceci n'est pas équivalent $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$ ou à, pour chaque paire $\{i, j\} \subset \{1, \dots, n\}$ les événements A_i et A_j sont indépendants.

Exemples :

Considérons l'espace correspondant à deux lancers de pile ou face et prenons $A = \{\text{pile au premier lancer}\}$, $B = \{\text{pile au second lancer}\}$ et $C = \{\text{même résultats aux deux lancers}\}$. Les trois événements sont 2 à 2 indépendants, mais ne sont pas indépendants.

Proposition 7. Les n événements A_1, \dots, A_n sont indépendants si et seulement si

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n)$$

pour $B_i \in \sigma(A_i) = \{\emptyset, \Omega, A_i, A_i^c\}$.

4. Lemme de Borel-Cantelli

Si $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements on note

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

De façon équivalente, $\omega \in \limsup_{n \rightarrow \infty} A_n \Leftrightarrow \omega$ appartient à une infinité d'événements A_n

On peut rappeler qu'on a aussi la définition $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k$. De façon équivalente,

$\omega \in \liminf_{n \rightarrow \infty} A_n \Leftrightarrow \omega$ appartient à tous les A_n sauf éventuellement un nombre fini.

Enfin, puisqu'une suite à valeurs dans $\{0, 1\}$ a pour limite supérieure 1 si et seulement si elle prend une infinité de fois la valeur 1, on a une interprétation simple en terme de fonctions indicatrices :

$$1_{\limsup A_n} = \limsup(1_{A_n}) \quad \text{et} \quad 1_{\liminf A_n} = \liminf(1_{A_n})$$

Le résultat suivant est le *lemme de Borel-Cantelli*, mais vu son importance cela pourrait être un théorème.

Lemme 8. Lemme de Borel Cantelli Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements

(i) Si $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$ alors

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$$

ou de manière équivalente

$\{n \in \mathbb{N} : \omega \in A_n\}$ est fini p.s.

(ii) Si $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty$ et si les événements A_n sont indépendants, alors

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$$

ou de manière équivalente

$\{n \in \mathbb{N} : \omega \in A_n\}$ est infini p.s.

L'hypothèse d'indépendance est nécessaire dans le (ii), comme le montre l'exemple trivial où $A_n = A$ pour tout $n \in \mathbb{N}$, avec $0 < \mathbb{P}(A) < 1$.

DÉMONSTRATION. (i) Supposons $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$. On a pour tout entier m

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=m}^{\infty} A_k$$

et donc $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) \leq \sum_{k \geq m} \mathbb{P}(A_k)$ qui tend vers 0 avec m si la série converge.

(ii) Pour le deuxième point, remarquons d'abord que pour tout n et tout $N \geq n$

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \leq k \leq N} A_k\right) &= 1 - \mathbb{P}\left(\bigcap_{n \leq k \leq N} A_k^c\right) \\ &= 1 - \prod_{n \leq k \leq N} (1 - \mathbb{P}(A_k)) \end{aligned}$$

Comme $1 - x \leq e^{-x}$ pour tout $x \geq 0$, on en déduit

$$\mathbb{P}\left(\bigcup_{n \leq k \leq N} A_k\right) \geq 1 - \exp\left(-\sum_{n \leq k \leq N} \mathbb{P}(A_k)\right)$$

Lorsque N tend vers l'infini, la somme dans l'exponentielle tend, pour tout n , vers l'infini par hypothèse, et donc $\mathbb{P}(\bigcup_{k \geq n} A_k) = 1$. Il ne reste alors plus qu'à remarquer que $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(\bigcup_{k \geq n} A_k)$. \square

CHAPITRE II

Variables aléatoires

1. Définitions

Définition 1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et (E, \mathcal{E}) un espace mesurable. Une application mesurable $X : \Omega \rightarrow E$ est appelée variable aléatoire à valeurs dans E . Lorsque $E \subset \mathbb{R}$, on parle de variable aléatoire réelle.

Exemple :

Si on lance deux dés on peut définir la variable aléatoire qui est égale à la somme des résultats des dés : $X((i, j)) = i + j$. Alors X définit une variable aléatoire sur $\Omega = \{1, 2, \dots, 6\}^2$ à valeurs dans $E = \{1, 2, \dots, 12\}$.

Si on lance un dé une infinité de fois on peut définir une variable aléatoire Y comme étant le premier instant où l'on obtient 6. Alors on a : $Y(\omega) = \inf\{j : \omega_j = 6\}$, avec la convention $\inf \emptyset = \infty$; Y est à valeurs dans $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Dans ce cas, pour vérifier la mesurabilité on observe que, pour tout $k \geq 1$

$$Y^{-1}(\{k\}) = \{\omega \in \Omega : \omega_1 \neq 6, \omega_2 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\}.$$

Définition 2. La loi d'une variable aléatoire $X : \Omega \rightarrow E$ est la mesure-image de \mathbb{P} par X . C'est donc la probabilité sur (E, \mathcal{E}) , notée \mathbb{P}_X , définie par : $\forall B \in \mathcal{E}$

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}).$$

1.A. Variable aléatoire discrète.

Définition 3. Une variable aléatoire $X : \Omega \rightarrow E$ est une variable aléatoire discrète si E (en fait $X(\Omega)$) est un ensemble fini ou dénombrable. Dans ce cas, la tribu associée est, en général, l'ensemble des parties $\mathcal{P}(E)$.

Pour une variable discrète X , la loi est donnée par

$$\mathbb{P}_X = \sum_{x \in E} p_x \delta_x$$

où $p_x = \mathbb{P}(X = x)$ et δ_x désigne la mesure de Dirac en x . Ceci provient du fait suivant :

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(\bigcup_{x \in B} \{X = x\}\right) = \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in B} p_x \delta_x(B)$$

En pratique déterminer la loi d'une variable aléatoire discrète, revient à calculer toutes les probabilités $\mathbb{P}(X = x)$ pour $x \in E$.

On supposera dans la suite sans perte de généralité que pour les variables discrètes $E \subset \mathbb{Z}$.

1.B. Variable aléatoire à densité.

Définition 4. Une variable aléatoire X à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est à densité si \mathbb{P}_X est absolument continue par rapport à la mesure de Lebesgue λ .

Le théorème de Radon-Nikodym nous assure alors qu'il existe une fonction borélienne $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ telle que

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(B) = \int_B f(x) dx.$$

On a en particulier $\int_{\mathbb{R}^d} f(x) dx = P(X \in \mathbb{R}^d) = 1$. La fonction f (unique à un ensemble de mesure de Lebesgue nulle près), est appelée la *densité de la loi* de X .

Si $d = 1$, on a en particulier, pour tout $\alpha \leq \beta$,

$$\mathbb{P}(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x) dx.$$

2. Tribu engendrée par une variable aléatoire

Soit X une variable aléatoire à valeurs dans un espace mesurable quelconque (E, \mathcal{E}) . La tribu engendrée par X , notée $\sigma(X)$, est par définition la plus petite tribu sur Ω qui rende X mesurable :

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

On peut généraliser cette définition à une famille quelconque $(X_i)_{i \in I}$ de variables aléatoires, X_i étant à valeurs dans (E_i, \mathcal{E}_i) . Dans ce cas,

$$\sigma(X_i, i \in I) = \{X_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i \in I\}.$$

Proposition 5. Soit X une variable aléatoire à valeurs dans un espace (E, \mathcal{E}) , et soit Y une variable aléatoire réelle. Il y a équivalence entre :

(i) Y est $\sigma(X)$ -mesurable ;

(ii) il existe une fonction mesurable f de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $Y = f(X)$.

DÉMONSTRATION. L'implication (ii) \Rightarrow (i) est facile puisque la composée de fonctions mesurables est mesurable.

Pour la réciproque on traite d'abord le cas où Y est étagée : $Y = \sum_{i=1}^n \lambda_i 1_{A_i}$ où $\lambda_i \in \mathbb{R}$ et $A_i \in \sigma(X)$. Pour tout i , on peut trouver $B_i \in \mathcal{E}$ tel que $A_i = X^{-1}(B_i)$, et on a

$$Y = \sum_{i=1}^n \lambda_i 1_{A_i} = \sum_{i=1}^n \lambda_i 1_{B_i} \circ X = f \circ X$$

où $f = \sum_{i=1}^n \lambda_i 1_{B_i}$ est \mathcal{E} -mesurable. Dans le cas général, on sait que Y est limite simple d'une suite de variables aléatoires Y_n étagées et $\sigma(X)$ -mesurables. D'après la première étape, on peut écrire, pour tout n , $Y_n = f_n(X)$, où $f_n : E \rightarrow \mathbb{R}$ est mesurable. On pose alors : pour tout $x \in E$

$$f(x) = \begin{cases} \lim_{n \rightarrow \infty} f_n(x) & \text{si cette limite existe} \\ 0 & \text{sinon} \end{cases}$$

qui est mesurable. On obtient ainsi $Y = f(X)$. □

3. Fonction de répartition d'une variable aléatoire

Définition 6. Si X est une variable aléatoire réelle, la fonction de répartition de X est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par : $\forall t \in \mathbb{R}$,

$$F_X(t) = \mathbb{P}(X \leq t).$$

On peut vérifier les propriétés suivantes :

- la fonction F_X est croissante ;
- la fonction F_X est continue à droite ;
- la fonction F_X a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Inversement, si on se donne une fonction F ayant ces propriétés, d'après le cours d'intégration, il existe une (unique) mesure de probabilité μ telle que $\mu(]-\infty, t]) = F(t)$ pour tout $t \in \mathbb{R}$. Cela montre qu'on peut interpréter F comme la fonction de répartition d'une variable aléatoire réelle. Il découle, que F_X caractérise la loi \mathbb{P}_X de X . On a :

$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= F_X(b) - F_X(a^-) \text{ si } a \leq b, \\ \mathbb{P}(a < X < b) &= F_X(b^-) - F_X(a^-) \text{ si } a < b. \end{aligned}$$

et les sauts de F_X correspondent aux atomes de \mathbb{P}_X . Ainsi une variable aléatoire à densité aura une fonction de répartition continue, alors que la fonction de répartition d'une variable aléatoire discrète sera en escalier.

Remarque : il existe des variables aléatoires qui ne sont ni discrète, ni à densité. Il suffit de considérer une fonction de répartition qui ne soit ni continue ni en escalier... Un exemple explicite assez simple consiste à considérer un mélange des deux lois : par exemple, si $\Omega = [0, 1]$, la variable aléatoire définie par $X(\omega) = 1_{[0, \frac{1}{2}]}$ (ω) + $\omega 1_{] \frac{1}{2}, 1]}$ (ω) vaut 1 avec probabilité $\frac{1}{2}$, mais elle n'est pas à densité.

Mais on peut aussi construire un exemple de variable aléatoire avec fonction de répartition continue mais qui ne soit pas à densité (elle aura alors une loi qui sera une mesure de probabilité sur \mathbb{R} sans atome, mais étrangère à la mesure de Lebesgue). Par exemple, si les ϵ_n sont des variables de Bernoulli de paramètre $\frac{1}{2}$ indépendantes, il est facile de vérifier que la série $\sum_{n=1}^{\infty} \frac{2\epsilon_n}{3^n}$ converge, et la variable aléatoire limite appartient avec probabilité 1 à l'ensemble de Cantor, elle n'est donc pas à densité.

La croissance de la fonction de répartition F entraîne l'existence de la fonction suivante.

Définition 7. Soit F une fonction de répartition. On appelle fonction quantile la fonction

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

Si F est bijective d'un intervalle $I \subset \mathbb{R}$ dans $]0, 1[$, on a simplement $F^- = F^{-1}$.

Cette fonction est très utile pour la simulation informatique de variables aléatoires, grâce à la proposition suivante :

Proposition 8. Si U est une variable de loi uniforme sur $[0, 1]$ et si F est une fonction de répartition, alors la variable $X = F^-(U)$ a pour fonction de répartition F .

Ainsi, si l'on dispose d'un générateur de variable de loi uniforme sur $[0, 1]$, pour simuler une variable de fonction de répartition F il suffit d'écrire un algorithme qui calcule F^- .

DÉMONSTRATION. On va vérifier $\forall t \in \mathbb{R}, \mathbb{P}(F^-(U) \leq t) = F(t)$. Il suffit pour cela de vérifier $\forall (u, t) \in]0, 1[\times \mathbb{R}, F^-(u) \leq t \Leftrightarrow u \leq F(t)$.

Si on suppose $u \leq F(t)$, alors $F^-(u) \leq t$. Réciproquement, si on suppose $u > F(t)$, puisque F est

continue à droite en t , il existe $t_1 > t$ tel que $F(t_1) < u$. Puisque F est croissante, $\forall x \leq t_1, F(x) < u$ et par conséquent $F^-(u) \geq t_1 > t$. \square

4. Vecteur aléatoire

Soient X et Y deux variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) .

Définition 9. La loi du couple (X, Y) , ou loi conjointe de X et Y , est la probabilité sur $(E_1 \times E_2, (\mathcal{E}_1 \otimes \mathcal{E}_2))$ notée $\mathbb{P}_{(X,Y)}$ définie par

$$\forall B \in \mathcal{E}_1 \otimes \mathcal{E}_2, \mathbb{P}_{(X,Y)}(B) = \mathbb{P}((X, Y) \in B)$$

Dans le cas de variables aléatoires discrètes, elle est caractérisée par l'application p définie sur $E_1 \times E_2$ par

$$(x, y) \mapsto \mathbb{P}(X = x, Y = y).$$

On généralise trivialement cette définition à un n -uplet.

Connaissant la loi d'un couple, on détermine la probabilité d'un événement par une somme double lorsque les variables sont discrètes, et par une intégrale double lorsque le couple de variables aléatoires est à densité.

Connaissant la loi du couple (X, Y) il est facile de retrouver la loi de X et de Y qui sont alors appelées *lois marginales* :

– Si X et Y sont discrètes,

$$\mathbb{P}(X = x) = \sum_{y \in E_2} \mathbb{P}(X = x, Y = y) \quad \mathbb{P}(Y = y) = \sum_{x \in E_1} \mathbb{P}(X = x, Y = y).$$

– Si (X, Y) à valeurs dans \mathbb{R}^{m+n} est à densité $f_{X,Y}$, la densité de X est

$$\forall x \in \mathbb{R}^m, f_X(x) = \int_{\mathbb{R}^n} f(x, y) dy$$

Les lois marginales se déduisent donc de la loi conjointe mais la réciproque est fautive comme le montre l'exemple ci-dessous.

Exemple :

Considérons l'espace correspondant au lancer de deux dés. Si X représente le résultat du premier dé et Y celui du second. Alors la loi conjointe de (X, Y) est équirépartie sur $\{1, \dots, 6\}^2$, mais $\mathbb{P}(X = 1, X = 2) = 0$. C'est-à-dire que les couples (X, Y) et (X, X) n'ont pas la même loi alors que les lois marginales sont toutes égales.

On peut contourner ce problème quand les variables sont indépendantes.

A partir de la loi conjointe, on peut de manière générale obtenir la loi de n'importe quelle fonction (mesurable) de (X, Y) $Z = \phi(X, Y)$ à partir de la fonction de répartition de Z : il faut pour cela déterminer pour tout réel t l'ensemble $A_t = \{(x, y) \in \mathbb{R}^2, \phi(x, y) \leq t\}$ et calculer $P(Z \leq t) = P(X, Y) \in A_t$. Dans le cas où les variables sont indépendantes et où $Z = X + Y$, on se ramène simplement un produit de convolution, comme on va le voir.

5. Variables aléatoires indépendantes

5.A. Définitions. La notion la plus générale est celle de *tribus indépendantes*.

Définition 10. On dit que les tribus, $\mathcal{B}_1, \dots, \mathcal{B}_n$ incluses dans \mathcal{A} , sont indépendantes si

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

Définition 11. On dit que les variables aléatoires, X_1, \dots, X_n à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$, sont indépendantes si les tribus $\sigma(X_1), \dots, \sigma(X_n)$ le sont. Cela équivaut à

$$\forall F_1 \in \mathcal{E}_1, \dots, \forall F_n \in \mathcal{E}_n, \quad \mathbb{P}(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\}) = \mathbb{P}(X_1 \in F_1) \dots \mathbb{P}(X_n \in F_n).$$

En effet, on sait que $\sigma(X_i) = \{X_i^{-1}(F), F \in \mathcal{E}_i\}$.

De manière intuitive les variables aléatoires, X_1, \dots, X_n sont indépendantes si la connaissance de certaines d'entre elles ne donne pas d'information sur les autres.

Si $\mathcal{B}_1, \dots, \mathcal{B}_n$ sont n sous-tribus indépendantes, et si pour tout i , X_i est une variable aléatoire \mathcal{B}_i -mesurable, alors les variables aléatoires X_1, \dots, X_n sont indépendantes.

Les n événements A_1, \dots, A_n sont indépendants si et seulement si les tribus $\sigma(A_1), \dots, \sigma(A_n)$ le sont.

Théorème 12.

Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$, on note X le vecteur aléatoire (X_1, \dots, X_n) qui est à valeurs dans $E_1 \times \dots \times E_n$ muni de la tribu produit $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$.

Les n variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si la loi du vecteur aléatoire $X = (X_1, \dots, X_n)$ est le produit des lois de X_1, \dots, X_n :

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

DÉMONSTRATION. Soit $F_i \in \mathcal{E}_i$ pour tout $i \in \{1, \dots, n\}$. On a

$$\mathbb{P}_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = \mathbb{P}(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\})$$

et

$$\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}(F_1 \times \dots \times F_n) = \prod_{i=1}^n \mathbb{P}_{X_i}(F_i) = \prod_{i=1}^n \mathbb{P}(X_i \in F_i).$$

L'indépendance de X_1, \dots, X_n est équivalente à, les deux mesures de probabilités $\mathbb{P}_{(X_1, \dots, X_n)}$ et $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ prennent les mêmes valeurs sur les pavés $F_1 \times \dots \times F_n$. Mais, d'après le lemme de classe monotone, une mesure de probabilité sur un espace produit est caractérisée par ses valeurs sur les pavés, d'où les deux mesures sont égales. \square

On peut généraliser ces définitions pour des familles infinies :

Définition 13. Soit $(\mathcal{B}_i)_{i \in I}$ une famille quelconque de sous-tribus de \mathcal{A} . On dit que cette famille est indépendante si pour tout sous-ensemble fini $\{i_1, \dots, i_p\}$ de I , les tribus $\mathcal{B}_{i_1}, \dots, \mathcal{B}_{i_p}$ sont indépendantes. De même, si $(X_i)_{i \in I}$ est une famille quelconque de variables aléatoires, cette famille est dite indépendante si la famille des tribus $(\sigma(X_i))_{i \in I}$ l'est.

5.B. Cas particuliers des variables aléatoires discrètes et à densité.

Proposition 14. *Les deux variables aléatoires discrètes X et Y sont indépendantes ssi pour toutes les valeurs x et y les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants, autrement dit*

$$\forall (x, y) \in E_1 \times E_2, \quad \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Comme déjà mentionné, la loi du couple est alors le produit (au sens des mesures) des deux lois marginales.

Dans le cas des variables aléatoire à densité, la définition en terme de mesure produit donne également un critère assez simple :

Proposition 15. *Soient X_1, \dots, X_n des variables aléatoires à densité à valeurs réelles. Elles sont indépendantes si et seulement si la densité du vecteur (X_1, \dots, X_n) est le produit des densités des X_i : $f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$ pour presque tout (x_1, \dots, x_n) .*

5.C. Propriétés.

Proposition 16. Regroupement par paquets

Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ des tribus indépendantes, et soient $n_0 = 0 < n_1 < \dots < n_p = n$. Alors les tribus

$$\begin{aligned} \mathcal{D}_1 &= \mathcal{B}_1 \wedge \dots \wedge \mathcal{B}_{n_1} = \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1}) \\ \mathcal{D}_2 &= \mathcal{B}_{n_1+1} \wedge \dots \wedge \mathcal{B}_{n_2} \\ &\quad \dots \\ \mathcal{D}_p &= \mathcal{B}_{n_{p-1}+1} \wedge \dots \wedge \mathcal{B}_{n_p} \end{aligned}$$

sont indépendantes.

En particulier, si X_1, \dots, X_n sont indépendantes, les variables aléatoires

$$Y_1 = (X_1, \dots, X_{n_1}), \dots, Y_p = (X_{n_{p-1}+1}, \dots, X_{n_p})$$

sont indépendantes.

Proposition 17. *Si X_1, \dots, X_n sont n variables aléatoires indépendantes et si g_1, \dots, g_n sont n fonctions mesurables, alors les variables aléatoires $Y_1 = g_1(X_1), \dots, Y_n = g_n(X_n)$ sont indépendantes.*

C'est une conséquence immédiate des inclusions $\forall 1 \leq i \leq n, \sigma(g(X_i)) \subset \sigma(X_i)$.

5.D. Sommes de variables aléatoires indépendantes. On commence par rappeler la définition du produit de convolution $*$: si μ et ν sont deux mesures de probabilités sur \mathbb{R}^d on note $\mu * \nu$ la mesure-image de $\mu \otimes \nu$ par l'application $(x, y) \rightarrow x + y$; pour toute fonction mesurable positive φ sur \mathbb{R}^d

$$\int_{\mathbb{R}^d} \varphi(z) \mu * \nu(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x + y) \mu(dx) \nu(dy).$$

Proposition 18. Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{R}^d . Alors la loi de $X + Y$ est $\mathbb{P}_X * \mathbb{P}_Y$. En particulier :

(i) Si X et Y sont à valeurs dans \mathbb{N} , on a

$$\forall n \in \mathbb{N}, P(X + Y = n) = \sum_{k=0}^n P(X = k)P(Y = n - k)$$

(ii) Si X a une densité f_X et si Y a une densité f_Y alors, $X + Y$ a une densité $f_X * f_Y$ définie par

$$f_X * f_Y(z) = \int_{\mathbb{R}^d} f_X(x)f_Y(z - x)dx = \int_{\mathbb{R}^d} f_X(z - y)f_Y(y)dy.$$

5.E. Loi du 0-1.

Définition 19. Soit $(\mathcal{T}_n)_{n \in \mathbb{N}}$ une suite de tribus indépendantes sur $(\Omega, \mathcal{A}, \mathbb{P})$. On note \mathcal{B}_n la tribu engendrée par $\mathcal{T}_n, \mathcal{T}_{n+1}, \dots$. La tribu $\mathcal{B}_\infty = \bigcap_{n \in \mathbb{N}} \mathcal{B}_n$ est appelée tribu terminale.

Intuitivement, les événements de la tribu terminale sont caractérisés par des propriétés “asymptotiques” : par exemple, $\limsup_{n \rightarrow \infty} \mathcal{B}_n$ est un élément de la tribu terminale. La loi du 0-1 de Kolmogorov peut être vue comme une généralisation du lemme de Borel-Cantelli :

Théorème 20.

Loi du 0-1 de Kolmogorov Une tribu terminale vérifie la loi du 0 – 1 (ou loi du tout ou rien), c'est-à-dire que pour événement A de la tribu terminale $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

DÉMONSTRATION. Soit $B \in \mathcal{B}_\infty$. On considère la classe monotone des événements indépendants de B :

$$\mathcal{M} = \{A \in \mathcal{A} : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}.$$

On va montrer que $\mathcal{B}_\infty \subset \mathcal{M}$. Alors, on aura $B \in \mathcal{M}$ et donc $\mathbb{P}(B) = \mathbb{P}(B \cap B) = \mathbb{P}(B)^2$, c'est-à-dire $\mathbb{P}(B) = 0$ ou 1. On pose

$$\mathcal{C}_n = \sigma(\mathcal{T}_0, \dots, \mathcal{T}_n).$$

Comme les tribus au départ sont indépendantes, pour tout $n \in \mathbb{N}$ les tribus \mathcal{C}_n et \mathcal{B}_{n+1} le sont aussi ; par conséquent pour tout $n \in \mathbb{N}$ les tribus \mathcal{C}_n et \mathcal{B}_∞ sont indépendantes. On en déduit que :

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

Ceci entraîne que $\mathcal{C}_\infty = \bigcup_{n=1}^{\infty} \mathcal{C}_n \subset \mathcal{M}$ et par le théorème des classes monotones $\sigma(\mathcal{C}_\infty) = \mathcal{M}(\mathcal{C}_\infty) \subset \mathcal{M}$. D'autre part on a :

$$\overline{\mathcal{T}}_k \subset \mathcal{C}_k \subset \mathcal{C}_\infty \subset \sigma(\mathcal{C}_\infty)$$

et par conséquent

$$\mathcal{B}_n = \sigma(\overline{\mathcal{T}}_k, k \geq n) \subset \sigma(\mathcal{C}_\infty)$$

d'où le résultat. □

CHAPITRE III

Espérance

1. Définition

Définition 1. Soit X une variable aléatoire réelle intégrable. On appelle espérance mathématique de X la quantité suivante :

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$$

– Si X est à valeurs dans \mathbb{Z} , X est intégrable si et seulement si la série $\sum_{n \in \mathbb{Z}} |n| \mathbb{P}(X = n)$ converge. On a alors

$$\mathbb{E}[X] = \sum_{n \in \mathbb{Z}} n \mathbb{P}(X = n)$$

– Si X a pour densité f , X est intégrable si et seulement si $x \mapsto xf$ l'est et on a alors

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x) dx$$

On étend cette définition au cas où $X = (X_1, \dots, X_d)$ est à valeurs dans \mathbb{R}^d en prenant alors $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$, pourvu que chacune des espérances $\mathbb{E}[X_i]$ soit bien définie.

On a, si $X = 1_B$ alors $\mathbb{E}[X] = \mathbb{P}(B)$. En général, $\mathbb{E}[X]$ s'interprète comme la moyenne de la variable aléatoire X .

Dans le cas particulier où Ω est fini et \mathbb{P} est équidistribuée, $\mathbb{E}[X]$ est bien la moyenne arithmétique au sens usuel des valeurs prises par X .

Proposition 2. Soit X une variable aléatoire à valeurs dans (E, \mathcal{E}) . Pour toute fonction mesurable $f : E \rightarrow [0, \infty]$, on a

$$\mathbb{E}[f(X)] = \int_E f(x) d\mathbb{P}_X(x).$$

DÉMONSTRATION. On remarque que le résultat est vrai par définition pour $f = 1_B$ puis par linéarité pour toute fonction étagée positive. Dans le cas général, on utilise le théorème de convergence monotone et le fait que toute fonction mesurable positive est limite croissante d'une suite de fonctions étagées positives. \square

Si f est de signe quelconque, la formule de la proposition reste vraie à condition que les intégrales soient bien définies, ce qui revient à $\mathbb{E}[|f(X)|] < +\infty$.

La donnée de \mathbb{P}_X permet donc de calculer la valeur moyenne de variables aléatoires de la forme $f(X)$. Inversement, on peut utiliser cette proposition pour calculer la loi d'une variable aléatoire X : si on arrive

à écrire

$$\mathbb{E}[f(X)] = \int f d\mu$$

pour toute fonction f “suffisamment” générale (par exemple toute fonction borélienne bornée), alors on peut identifier μ à la loi de X . Cela donne encore un autre moyen de calculer la loi d’une fonction d’une variable aléatoire X . Dans le cas où f est une fonction puissance $x \mapsto x^k$, $\mathbb{E}[X^k]$ (si elle existe) est appelée *moment d’ordre k de X* .

2. Moments d’une variable aléatoire

Définition 3. Soit X une variable aléatoire réelle et soit $k \geq 1$ un entier. Le moment d’ordre k de X est par définition la quantité $\mathbb{E}[X^k]$ qui n’est définie que si $|X|^k$ est intégrable ($\mathbb{E}[|X|^k] < +\infty$).

Si X est à valeurs dans \mathbb{Z} , X possède un moment d’ordre $k \in \mathbb{N}$, si la série $\sum_{n \in \mathbb{Z}} |n|^k \mathbb{P}(X = n)$ converge. Le moment est alors noté

$$\mathbb{E}[X^k] = \sum_{n \in \mathbb{Z}} n^k \mathbb{P}(X = n)$$

En particulier le moment d’ordre 1 est simplement l’espérance de X . On dit qu’une variable aléatoire réelle est centrée si elle est intégrable et si $\mathbb{E}[X] = 0$.

Le cas du moment d’ordre 2 est tout particulièrement important car l’espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$ muni du produit scalaire $\langle X, Y \rangle = E[XY]$ est un espace de Hilbert.

Définition 4. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La variance de X est

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

et l’écart-type de X est

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

De manière informelle, $\text{Var}(X)$ représente la dispersion de X autour de sa moyenne $E[X]$.

De manière plus géométrique, en terme de norme dans l’espace de Hilbert $L^2(\Omega, \mathcal{A}, \mathbb{P})$, la variance mesure la distance de X à son espérance

$$\sqrt{\text{Var}(X)} = \|X - \mathbb{E}(X)\|_2.$$

On considère souvent la variable centrée réduite associée à une variable aléatoire X , définie par $X^* = \frac{X - \mathbb{E}[X]}{\sigma(X)}$, qui vérifie $\mathbb{E}[X^*] = 0$ et $\sigma(X^*) = 1$.

Proposition 5. Si le moment d’ordre k existe tous les moments d’ordre k' , $k' \leq k$, existent également.

DÉMONSTRATION. Ce n’est rien d’autre que l’inclusion $L^{k'} \subset L^k$ si $k' \leq k$ dans le cas d’une mesure finie. Rappel de la démonstration On considère deux entiers tels que $1 \leq k' \leq k$. On a $0 \leq |x|^{k'} \leq \sup\{|x|^k, 1\} \leq 1 + |x|^k$ et donc

$$\int_E |x|^{k'} d\mathbb{P}(x) \leq \int_E (1 + |x|^k) d\mathbb{P}(X = x) = 1 + \int_E |x|^k d\mathbb{P}(x)$$

d’où le résultat. □

3. Propriétés

Les propriétés ci-dessous se démontrent sans difficulté.

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $X \geq 0 \Rightarrow \mathbb{E}[X] \geq 0$
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\text{Var}(X) = 0 \Rightarrow X$ est constante presque sûrement
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $\mathbb{E}[(X - a)^2] = \text{Var}(X) + (\mathbb{E}[X] - a)^2$
- $\text{Var}(X) = \inf_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2]$.

Ces deux dernières propriétés expriment le fait que l'espérance $\mathbb{E}[X]$ est la projection orthogonale dans L^2 de la variable aléatoire X sur l'espace H des variables aléatoires constantes, et que $\text{Var}(X)$ est le carré de la distance de X à H .

Les inégalités suivantes sont très utiles.

Proposition 6. Inégalité de Markov

Si X est une variable aléatoire positive et si $a > 0$, alors on a

$$\mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[X].$$

Plus généralement, si $p > 0$ et si $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{P}(X \geq a) \leq \frac{1}{a^p} \mathbb{E}[X^p].$$

Il suffit pour la démontrer de remarquer que sur l'ensemble $\{X \geq a\}$, on a $1 \leq \frac{X}{a}$.

Une conséquence directe de ce résultat est la proposition suivante.

Proposition 7. Inégalité de Bienaymé - Tchebytchev

Si $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et si $a > 0$, alors on a

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} \text{Var}(X).$$

De plus, l'espérance mathématique étant un cas particulier d'intégrale par rapport à une mesure positive, on peut donc lui appliquer tous les théorèmes généraux d'intégrations.

Théorème 8.

Théorèmes de convergences.

Convergence monotone : $X_n \geq 0$, $X_n \uparrow X \Rightarrow \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.

Lemme de Fatou : $X_n \geq 0 \Rightarrow \mathbb{E}[\liminf X_n] \leq \liminf \mathbb{E}[X_n]$.

Convergence dominée : $|X_n| \leq Z$, $\mathbb{E}[Z] < \infty$, $X_n \rightarrow X$ p.s. $\Rightarrow \mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

En probabilité on utilise l'expression presque sûrement plutôt que le presque partout de la théorie de la mesure.

Les espaces $L^p(\Omega, \mathcal{A}, \mathbb{P})$ sont définies pour tout $p \in [1, \infty]$ comme dans le cours d'intégration.

L'inégalité de Hölder s'écrit

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}$$

pourvu que $\frac{1}{p} + \frac{1}{q} = 1$. En prenant $Y = 1$ on trouve $\|X\|_1 \leq \|X\|_p$, ce qui se généralise aussitôt à $\|X\|_r \leq \|X\|_p$ si $r \leq p$; en particulier $L^p(\Omega, \mathcal{A}, \mathbb{P}) \subset L^r(\Omega, \mathcal{A}, \mathbb{P})$ si $r \leq p$.

L'inégalité de Cauchy-Schwarz s'écrit

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^2]^{1/2} \mathbb{E}[|Y|^2]^{1/2}$$

et le cas particulier où $Y = 1$, $\mathbb{E}[|X|]^2 \leq \mathbb{E}[|X|^2]$, est souvent très utile.

4. Covariance et corrélation

Définition 9. On appelle covariance des variables aléatoires réelles X et Y la quantité, lorsqu'elle existe,

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

L'inégalité de Cauchy-Schwartz garantit que la covariance est bien définie dès que X et Y possèdent un moment d'ordre 2. On remarque de plus facilement que $\text{Cov}(X, X) = \text{Var}(X)$ et $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Il est également facile de vérifier que la covariance est bilinéaire, et que si a est un réel, $\text{Cov}(X, a) = 0$. En particulier, si a, b, c, d sont 4 réels, $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.

Théorème 10.

Si X et Y sont deux variables aléatoires indépendantes possédant une espérance et telles que XY possède une espérance, alors

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Plus généralement, si X et Y sont indépendantes et si f et g sont deux fonctions mesurables telles que $f(X)$ et $f(Y)$ possèdent une espérance, on a alors

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

DÉMONSTRATION. La démonstration est très simple dans le cas discret : on a alors

$$\mathbb{E}[XY] = \sum_{x,y} xy\mathbb{P}(X=x, Y=y) = \sum_{x,y} xy\mathbb{P}(X=x)\mathbb{P}(Y=y) = \sum_x x\mathbb{P}(X=x) \sum_y y\mathbb{P}(Y=y) = \mathbb{E}[X]\mathbb{E}[Y].$$

Dans le cas général, c'est une conséquence du théorème de Fubini ; en effet si les variables aléatoires sont indépendantes la loi de (X, Y) est le produit des deux lois ; l'intégrabilité de XY permet alors d'utiliser le théorème de Fubini et d'écrire :

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy d\mathbb{P}(x)d\mathbb{P}(y) = \int_{\mathbb{R}} x d\mathbb{P}(x) \int_{\mathbb{R}} y d\mathbb{P}(y) = \mathbb{E}[X]\mathbb{E}[Y]$$

Il suffit pour la généralisation de remarquer que $f(X)$ et $g(Y)$ sont alors indépendantes. \square

On en déduit facilement le résultat suivant.

Proposition 11. Si X et Y sont deux variables aléatoires indépendantes ayant une covariance, alors cette covariance est nulle.

Attention la réciproque est fautive. Par exemple on peut prendre X et Y telles que :

$$\begin{aligned} P(X = 0, Y = -1) &= 0 & P(X = 0, Y = 0) &= 1/2 & P(X = 0, Y = 1) &= 0 \\ P(X = 1, Y = -1) &= 1/4 & P(X = 1, Y = 0) &= 0 & P(X = 1, Y = 1) &= 1/4 \end{aligned}$$

ou encore si on s'intéresse à des variables aléatoires à densité :

soit X_1 une variable aléatoire réelle de loi $\mathcal{N}(0, 1)$. Soit ϵ une deuxième variable aléatoire à valeurs dans $\{-1, 1\}$, indépendante de X_1 et telle que $\mathbb{P}(\epsilon = -1) = \mathbb{P}(\epsilon = 1) = 1/2$. Si on prend $X_2 = \epsilon X_1$, on voit immédiatement que $\text{Cov}(X_1, X_2) = 0$, alors que X_1 et X_2 ne sont pas indépendantes. En effet, sinon $|X_1|$ et $|X_2| = |X_1|$ le seraient. Or, si une variable aléatoire réelle est indépendante d'elle-même elle doit être constante presque sûrement et donc sa loi est une mesure de Dirac, d'où une contradiction.

La covariance permet de calculer la variance d'une somme de variables aléatoires :

Théorème 12.

Soient X_1, \dots, X_n des variables avec un moment d'ordre 2, alors

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

En particulier si X_1, \dots, X_n sont deux à deux indépendantes, alors $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$

Définition 13. Le coefficient de corrélation des variables aléatoires X et Y est défini par (quand il existe) :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Si $\rho(X, Y) = 0$ on dit que les variables sont non-corrélées.

Evidemment, si X et Y sont indépendantes leur coefficient de corrélation est nul. La réciproque est fautive. Le coefficient de corrélation donne donc une indication plus ou moins précise de l'indépendance de deux variables aléatoires. Cependant la réciproque est exacte si (X, Y) est un vecteur gaussien (cas fréquent), d'où l'intérêt de ce coefficient.

Donnons quelques propriétés du coefficient de corrélation.

Proposition 14. On a :

- $\rho(X, Y) \in [-1, 1]$
- Si $Y = aX + b$ ($a \neq 0$) alors $|\rho(X, Y)| = 1$
- Si $|\rho(X, Y)| = 1$ alors on a $\frac{X - \mathbb{E}[X]}{\sigma(X)} - \rho(X, Y) \frac{Y - \mathbb{E}[Y]}{\sigma(Y)} = 0$ (et donc Y est une fonction affine de X).

DÉMONSTRATION. Ce n'est rien d'autre que l'inégalité de Cauchy-Schwartz pour le produit scalaire dans L^2 $\langle X, Y \rangle = E[XY]$; rappel de la démonstration : on considère la quantité positive suivante :

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E} \left[\left(X - \mathbb{E}[X] + \lambda(Y - \mathbb{E}[Y]) \right)^2 \right]; \text{ on a donc}$$

$$\forall \lambda \in \mathbb{R}, \quad \lambda^2 \text{Var}(Y) + 2\lambda \text{Cov}(X, Y) + \text{Var}(X) \geq 0.$$

Le discriminant de ce trinôme est négatif ou nul, c'est-à-dire $\text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y)$, d'où $\rho(X, Y)^2 \leq 1$.

Maintenant si $|\rho(X, Y)| = 1$ le discriminant est nul et par conséquent le trinôme admet une racine double $\lambda_0 = -\text{Cov}(X, Y)/\text{Var}(Y)$ et donc

$$\mathbb{E}[(X - \mathbb{E}[X] + \lambda_0(Y - \mathbb{E}[Y]))^2] = 0$$

Ceci entraîne que $(X - \mathbb{E}[X] + \lambda_0(Y - \mathbb{E}[Y])) = 0$ et par conséquent le résultat.

Pour le deuxième résultat il suffit de remplacer Y par son expression et de faire les calculs. \square

5. Fonction génératrice

On ne donnera ici que la définition pour des variables aléatoires discrètes à valeurs dans \mathbb{N} .

Définition 15. Soit X une variable aléatoire à valeurs entières positives ($X(\Omega) \subset \mathbb{N}$). La fonction génératrice de X est définie par :

$$\forall s \in \mathbb{C}, |s| \leq 1, \quad G_X(s) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) s^k = \mathbb{E}(s^X).$$

Le rayon de convergence de cette série entière est au moins 1 puisque :

$$G_X(1) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) = 1.$$

Une fonction génératrice est donc en particulier X^∞ sur $] -1; 1[$. La connaissance de la fonction génératrice de X entraîne la connaissance de la loi puisque $\mathbb{P}(X = k)$ est le coefficient de s^k dans la série.

Théorème 16.

Si X et Y sont des variables aléatoires indépendantes alors $G_{X+Y}(s) = G_X(s)G_Y(s)$.
Plus généralement, si X_1, \dots, X_n sont des variables indépendantes $G_{X_1+\dots+X_n}(s) = G_{X_1}(s) \dots G_{X_n}(s)$.

DÉMONSTRATION. Par indépendance des variables aléatoires X et Y on a $\mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X]\mathbb{E}[s^Y]$. \square

Attention la réciproque de ce résultat est fausse.

Proposition 17. Soit X une variable aléatoire à valeur dans \mathbb{N} . Si la série converge on a

$$G_X^{(m)}(1) = \sum_{k=m}^{\infty} k(k-1)\dots(k-m+1)\mathbb{P}(X = k) = \mathbb{E}[X(X-1)\dots(X-m+1)]$$

et en particulier

$$\begin{aligned} G_X'(1) &= \mathbb{E}[X] \\ G_X''(1) &= \mathbb{E}[X^2] - \mathbb{E}[X]. \end{aligned}$$

La fonction génératrice permet donc d'obtenir l'espérance et la variance, puisqu'on a $\text{Var}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$.

Théorème 18.

Somme d'un nombre aléatoire de variables aléatoires Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées à valeurs dans \mathbb{N} . Soit N une variable aléatoire à valeurs dans \mathbb{N}^* indépendante des X_i . On pose $S_N = \sum_{i=1}^N X_i$. Alors la fonction génératrice de S_N vérifie :

$$G_{S_N}(s) = G_N \circ G_{X_1}(s).$$

De plus si X_1 et N admettent une espérance, S_N également et

$$\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X_1]$$

DÉMONSTRATION. On a

$$\mathbb{P}(S_N = k) = \sum_{n=1}^{\infty} \mathbb{P}(N = n)\mathbb{P}(S_N = k \mid N = n) = \sum_{n=1}^{\infty} \mathbb{P}(N = n)\mathbb{P}\left(\sum_{i=1}^n X_i = k\right)$$

ce qui entraîne le résultat. De plus si G_N et G_{X_1} sont dérivables en 1, comme on a $G_{X_1}(1) = 1$, G_{S_N} est dérivable en 1 et $G'_{S_N}(1) = G'_{X_1}(1)G'_N(1)$ d'où le résultat. \square

6. Fonctions caractéristiques

Pour les variables aléatoires à densité, on préfère utiliser la notion de fonction caractéristique.

Définition 19. Si X est une variable aléatoire à valeurs dans \mathbb{R}^d , la fonction caractéristique de X est la fonction : $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par

$$\forall t \in \mathbb{R}^d, \quad \Phi_X(t) = \mathbb{E}[\exp(it \cdot X)] = \int_{\mathbb{R}^d} \exp(it \cdot x) \mathbb{P}_X(dx).$$

Ceci permet de voir la fonction caractéristique comme la transformée de Fourier de la loi de X . Elle existe toujours car $|\exp(it \cdot X)| = 1$ et P_X est une mesure bornée. De plus d'après le théorème de convergence dominée, Φ_X est continue et bornée sur \mathbb{R}^d .

Lorsque X est une variable aléatoire réelle à densité,

$$\Phi_X(t) = \int_{\mathbb{R}} \exp(itx) f(x) dx.$$

Proposition 20. Soit X une variable aléatoire réelle. On note a et b deux constantes réelles. On a alors

$$\begin{aligned} \Phi_{aX}(t) &= \Phi_X(at) \\ \Phi_{X+b}(t) &= \exp(itb)\Phi_X(t) \end{aligned}$$

On en déduit en particulier, si X est une variable aléatoire d'espérance m et d'écart-type σ , en posant $U = (X - m)/\sigma$

$$\Phi_U(t) = \Phi_{\frac{X-m}{\sigma}}(t) = \exp\left(-\frac{itm}{\sigma}\right)\Phi_X\left(\frac{t}{\sigma}\right) \iff \Phi_X(t) = \exp(itm)\Phi_U(\sigma t).$$

La fonction caractéristique se prête bien aux additions de variables aléatoires indépendantes. En effet la fonction caractéristique d'une somme de variables aléatoires indépendantes est égale au produit de leurs fonctions caractéristiques :

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

En effet, si on considère X_1 et X_2 deux variables aléatoires indépendantes,

$$\Phi_{X_1+X_2}(t) = \mathbb{E}[\exp(it(X_1 + X_2))] = \mathbb{E}[\exp(itX_1)\exp(itX_2)]$$

L'indépendance de X_1 et X_2 entraîne que $\exp(itX_1)$ et $\exp(itX_2)$ sont indépendantes et donc que l'espérance du produit est égal au produit des espérances. Notons qu'il ne s'agit pas d'une condition nécessaire et suffisante.

Proposition 21. *Soit X une variable aléatoire symétrique par rapport à l'origine. Alors la fonction caractéristique de X est réelle.*

En effet

$$\begin{aligned}\Phi_X(-t) &= \int_{\mathbb{R}^d} \exp(-it \cdot x) \mathbb{P}_X(dx) = \overline{\Phi_X(t)} \\ &= \int_{\mathbb{R}^d} \exp(it \cdot x) \mathbb{P}_X(-dx) = \Phi_X(t)\end{aligned}$$

car d'après la symétrie $\mathbb{P}_X(-dx) = \mathbb{P}_X(dx)$.

Proposition 22. *On a $\Phi_X(0) = 1$. De plus si les dérivées à l'origine existent jusqu'à l'ordre $k \in \mathbb{N}^*$*

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k].$$

En particulier on a

$$\Phi_X'(0) = i\mathbb{E}[X] \quad \Phi_X''(0) = -\mathbb{E}[X^2].$$

Si Φ_X est indéfiniment dérivable en 0, la formule de Mac-Laurin donne

$$\Phi_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} i^k \mathbb{E}[X^k].$$

D'après les propriétés des transformées de Fourier, deux variables aléatoires ayant même fonctions caractéristiques ont même loi de probabilité. La fonction caractéristique détermine donc de manière unique une distribution de probabilité d'où son nom.

La formule d'inversion de la transformée de Fourier permet d'obtenir la loi de X connaissant Φ_X .

Proposition 23. *Si $\int_{\mathbb{R}} |\Phi_X(t)| dt < \infty$ alors X admet une densité $f(x)$ continue et :*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \Phi_X(t) \exp(-itx) dt.$$

Sinon on a toujours le résultat suivant :

$$F_X(b) - F_X(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \Phi_X(t) \frac{\exp(-ita) - \exp(-itb)}{it} dt.$$

Ce paragraphe est illustré par les calculs réalisés pour des variables aléatoires gaussiennes.

Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$ (gaussienne centrée réduite). Alors on a

$$\begin{aligned}\Phi_X(t) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \exp(itx) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{1}{2}[x - it]^2 - \frac{t^2}{2}\right) dx \\ &= \exp(-t^2/2) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}[x - it]^2\right) dx = \exp(-t^2/2).\end{aligned}$$

En effet l'intégrale vaut 1 : on peut le deviner car c'est l'intégrale d'une variable de Gauss imaginaire de moyenne it et de variance 1 ; plus rigoureusement une utilisation du théorème des résidus fournit une preuve.

Une autre méthode pour trouver ce résultat est d'utiliser la formule de Mac-Laurin. On sait que $\mathbb{E}[X^k] = 0$ si k est impair et $\mathbb{E}[X^{2k}] = \frac{(2k)!}{2^k k!}$, d'où

$$\Phi_X(t) = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} (-1)^k \frac{(2k)!}{2^k k!} = \sum_{k=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^k}{k!} = \exp(-t^2/2).$$

Soit X une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$, alors $U = (X - m)/\sigma$ est une variable aléatoire gaussienne centrée réduite. D'après la proposition 18

$$\Phi_X(t) = \exp(itm) \Phi_U(\sigma t) = \exp(itm) \exp(-\sigma^2 t^2/2).$$

On en déduit que la somme de deux variables aléatoires gaussiennes $X_1 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m_2, \sigma_2^2)$ a pour fonction caractéristique

$$\Phi_{X_1+X_2}(t) = \Phi_{X_1}(t) \Phi_{X_2}(t) = \exp(it(m_1 + m_2)) \exp(-(\sigma_1^2 + \sigma_2^2)t^2/2),$$

donc $X_1 + X_2$ suit une loi gaussienne $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

CHAPITRE IV

Convergence de variables aléatoires

1. Les différentes notions de convergences

Soit $(X_n)_{n \geq 1}$ et X des variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans \mathbb{R}^d .

Définition 1.

– On dit que la suite (X_n) converge presque sûrement vers X , et on note $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$, si

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\}) = 1.$$

– On dit que la suite (X_n) converge dans L^p vers X , et on note $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$, si

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

– On dit que la suite (X_n) converge en probabilité vers X , et on note $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$, si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Nous verrons dans la suite de ce chapitre une dernière notion, celle de convergence en loi.

On peut vérifier que l'on a

$$\{\omega \in \Omega : X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)\} = \bigcap_{p \geq 1} \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < 1/p\}$$

Proposition 2. Soit $\mathcal{L}_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, \mathbb{P})$ l'espace de toutes les variables aléatoires à valeurs dans \mathbb{R}^d , et soit $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, \mathbb{P})$ son quotient par la relation d'équivalence $X \sim Y$ si et seulement si $X = Y$ p.s. Alors la formule

$$d(X, Y) = \mathbb{E}[|X - Y| \wedge 1]$$

définit une distance sur $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, \mathbb{P})$ qui est compatible avec la convergence en probabilité au sens suivant :

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X \iff d(X_n, X) \xrightarrow[n \rightarrow \infty]{} 0.$$

De plus, l'espace $L_{\mathbb{R}^d}^0(\Omega, \mathcal{A}, \mathbb{P})$ est complet pour cette distance.

DÉMONSTRATION. On vérifie facilement que d est une distance. De plus si (X_n) converge en probabilité vers X , on a pour tout $\varepsilon > 0$,

$$\mathbb{E}[|X_n - X| \wedge 1] \leq \mathbb{E}[|X_n - X| 1_{|X_n - X| \leq \varepsilon}] + \mathbb{E}[(|X_n - X| \wedge 1) 1_{|X_n - X| > \varepsilon}] \leq \varepsilon + \mathbb{P}(|X_n - X| > \varepsilon).$$

D'après la définition de la convergence en probabilité, cela entraîne que $\limsup_{n \rightarrow \infty} d(X_n, X) \leq \varepsilon$, et puisque ε est arbitraire on a $d(X_n, X) \rightarrow 0$.

Réciproquement, si $d(X_n, X) \rightarrow 0$, alors, pour tout $\varepsilon \in]0, 1[$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-1} \mathbb{E}[|X_n - X| \wedge 1] = \varepsilon^{-1} d(X_n, X) \xrightarrow[n \rightarrow \infty]{} 0.$$

Il reste à montrer que L^0 est complet pour la distance d . Soit (X_n) une suite de Cauchy pour d . On peut trouver une sous-suite $Y_k = X_{n_k}$ telle que, pour $k \geq 1$, $d(Y_k, Y_{k+1}) \leq 2^{-k}$. Alors,

$$\mathbb{E} \left[\sum_{k=1}^{\infty} (|Y_{k+1} - Y_k| \wedge 1) \right] = \sum_{k=1}^{\infty} d(Y_{k+1}, Y_k) < \infty$$

ce qui entraîne $\sum_{k=1}^{\infty} |Y_{k+1} - Y_k| \wedge 1 < \infty$ p.s., et donc aussi $\sum_{k=1}^{\infty} |Y_{k+1} - Y_k| < \infty$. On définit ensuite une variable aléatoire dans L^0 en posant $X = Y_1 + \sum_{k=1}^{\infty} (Y_{k+1} - Y_k)$. Par construction la suite (Y_k) converge presque sûrement vers X et cela entraîne le résultat. \square

Une conséquence de la démonstration précédente est :

Proposition 3. *Si (X_n) converge en probabilité vers X , alors, il existe une sous-suite (X_{n_k}) qui converge presque sûrement vers X .*

Une conséquence immédiate du théorème de convergence dominée et de l'inégalité de Markov est l'implication ci-dessous :

Proposition 4. *Si (X_n) converge presque sûrement ou dans L^p vers X , alors, elle converge aussi en probabilité vers X .*

Enfin, on a

Proposition 5. *Si (X_n) converge en probabilité vers X et s'il existe $r \in]1, \infty[$ tel que (X_n) soit bornée dans L^r , alors, pour tout $p \in]1, r[$, la suite (X_n) converge dans L^p .*

DÉMONSTRATION. Soit $p \in]1, r[$. X_n étant bornée dans L^r , il existe un réel M tel que $\forall n, \mathbb{E}[|X_n|^r] \leq M$. Le lemme de Fatou appliqué à une sous-suite qui converge p.s, assure alors qu'on a également $\mathbb{E}[|X|^r] \leq M$. Il suffit alors puisque $\frac{r}{p} > 1$ d'utiliser l'inégalité de Hölder : pour tout $\varepsilon > 0$,

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n - X|^p 1_{|X_n - X| > \varepsilon}] + \mathbb{E}[|X_n - X|^p 1_{|X_n - X| \leq \varepsilon}] \leq (\mathbb{E}[|X_n - X|^r])^{p/r} (\mathbb{P}(|X_n - X| > \varepsilon))^{1-p/r} + \varepsilon^p$$

On en déduit $\mathbb{E}[|X_n - X|^p] \leq (2M)^{p/r} \mathbb{P}(|X_n - X| > \varepsilon)^{1-p/r} + \varepsilon^p$

En faisant tendre n vers l'infini, on obtient pour tout $\varepsilon > 0$, $\limsup \mathbb{E}[|X_n - X|^p] \leq \varepsilon^p$ et donc $\lim \mathbb{E}[|X_n - X|^p] = 0$. \square

On peut donner une conséquence directe du *lemme de Borel-Cantelli* pour les variables aléatoires.

Lemme 6. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.

(i) Si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$ alors $X_n \rightarrow X$ p.s.

(ii) Si les $(X_n)_{n \in \mathbb{N}}$ sont indépendantes alors $X_n \rightarrow 0$ p.s. si et seulement si pour tout $\varepsilon > 0$, $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) < \infty$.

DÉMONSTRATION. – Soit $k \in \mathbb{N}^*$. On note $A_{n,k}$ l'événement $|X_n - X| \geq \frac{1}{k}$. Si on suppose

$\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \frac{1}{k}) < \infty$, le lemme de Borel-Cantelli assure que $P(\limsup_n A_{n,k}) = 0$, c'est-à-dire

que $P(E_k) = 1$ où on a noté $E_k = \{\omega \in \Omega, \omega \in \text{un nombre fini de } A_{n,k}\}$. On a donc également $P(\cap_k E_k) = 1$: donc pour presque tout ω , quel que soit l'entier k , il existe un rang N tel que $\forall n \geq N |X_n - X| < \frac{1}{k}$: ainsi, X_n converge p.s. vers X .

– Réciproquement, si $\sum_{n \in \mathbb{N}} \mathbb{P}(|X_n| \geq \varepsilon) = \infty$ et si les X_n sont indépendantes, on déduit de Borel-Cantelli que pour presque tout ω , l'ensemble des entiers tels que $|X_n| \geq \varepsilon$ est infini : donc X_n ne peut pas converger vers 0. □

Le résultat ci-dessous est très utile en statistique :

Proposition 7. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles de carré intégrable définies sur $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose qu'il existe une constante a telle que $\lim_n \mathbb{E}[X_n] = a$ et que $\lim_n \text{Var}(X_n) = 0$. Alors X_n converge vers a dans L^2 et en probabilité.

DÉMONSTRATION. C'est une conséquence immédiate de l'égalité

$$\mathbb{E}[(X_n - a)^2] = \text{Var}(X_n) + (\mathbb{E}[X_n - a])^2. \quad \square$$

D'autre part ces convergences sont compatibles avec un certain nombre d'opérations :

Proposition 8.

– Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d qui converge p.s., respectivement en probabilités, vers une variable aléatoire X et f une fonction continue de \mathbb{R}^d dans \mathbb{R}^m . Alors la suite $f(X_n)$ converge p.s., respectivement en probabilités, vers $f(X)$.

– Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de variables aléatoires réelles qui convergent respectivement vers X et Y , p.s. ou en probabilités ou dans L^p . Alors $(X_n + Y_n)$ converge vers $X + Y$ p.s. ou en probabilités ou dans L^p .

DÉMONSTRATION. – Pour la convergence presque sûre, c'est clair. On suppose donc que X_n converge en probabilités vers X . Soit $\epsilon > 0$, et soit $\delta > 0$. On va montrer que pour n assez grand, $P(|f(X_n) - f(X)| > \epsilon) < 2\delta$. Puisque $\lim_m P(|X| > m) = 0$, il existe a tel que $P(|X| > a) \leq \delta$. D'autre part f étant continue elle est équicontinue sur la boule compacte $B(0, 2a)$. Ainsi il existe $\eta > 0$ tel que

$$\forall (x, y) \in B(0, 2a)^2, |f(x) - f(y)| > \epsilon \Rightarrow |x - y| > \eta$$

On en déduit

$$\forall (x, y) \in (\mathbb{R}^d)^2, |f(x) - f(y)| > \epsilon \Rightarrow |x| > a \text{ ou } |x - y| > \min(\eta, a)$$

Finalement

$$\mathbb{P}(|f(X) - f(X_n)| > \epsilon) \leq P(|X| > a) + P(|X_n - X| > \min(\eta, a)) \leq \delta + P(|X_n - X| > \min(\eta, a))$$

Ce dernier terme peut être rendu inférieur à δ en prenant n suffisamment grand.

- Le seul énoncé nouveau concerne la convergence en probabilités. On peut voir cette propriété comme une conséquence de la métrisabilité de cette convergence (en vertu du premier théorème du chapitre). C'est aussi une conséquence du point précédent : si (X_n) et (Y_n) convergent en probabilités, c'est également le cas du couple (X_n, Y_n) qui converge vers (X, Y) . Il suffit alors de lui appliquer la fonction continue $(x, y) \mapsto x + y$.

□

2. Loi des grands nombres ; théorème de Glivenko Cantelli

Donnons tout de suite l'énoncé le plus fort, celui de la loi forte des grands nombres. On trouvera en annexe sa démonstration ; on présentera dans la section suivante des démonstrations de versions plus faibles, plus abordables, et souvent plus facilement généralisables à d'autres situations.

Théorème 9.

Loi forte des grands nombres

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées, intégrables alors

$$\frac{\sum_{k=1}^n X_k}{n} \text{ converge p.s. vers } \mathbb{E}[X_1]$$

En statistique la moyenne arithmétique $\frac{\sum_{k=1}^n X_k}{n}$ sera en général notée \bar{X}_n .

Une autre version un peu plus forte de ce résultat est la suivante :

- pour toute suite de variables aléatoires indépendantes identiquement distribuées, on a

$$\frac{\sum_{k=1}^n X_k}{n} \text{ converge p.s. } \iff \mathbb{E}[|X_1|] < +\infty.$$

De plus si l'une des deux conditions est vérifiée alors la limite de la moyenne empirique est $\mathbb{E}[X_1]$.

Un cas particulier très souvent utilisé est explicité dans le corollaire suivant :

Corollaire 10. Si $(A_n)_{n \geq 1}$ est une suite d'événements indépendants de même probabilité, alors on a

$$\frac{1}{n} \sum_{k=1}^n 1_{A_k} \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{P}(A_1)$$

Exemple Soit X une variable aléatoire de loi uniforme sur $[0, 1]$ et Y_n la suite de ses décimales. On peut montrer facilement que les Y_n sont indépendantes et uniformément distribuées sur $\{0, \dots, 9\}$. Par conséquent on a p.s. $\frac{1}{n} \sum_{k=1}^n 1_{Y_k=5} = \mathbb{P}(Y_1 = 5) = \frac{1}{10}$. Dit autrement, un nombre réel choisi au hasard uniformément dans $[0, 1]$ aura presque sûrement une proportion asymptotique de 5 dans son développement décimal de $\frac{1}{10}$ (5 ne joue pas de rôle particulier, il en est de même pour tous les entiers entre 0 et 9). On montrerait ainsi plus généralement que presque tous les réels sont **normaux**.

Une conséquence fondamentale du corollaire précédent, qui est à la base du test de Kolmogorov-Smirnov est le théorème de Glivenko-Cantelli, qui concerne la convergence de la fonction de répartition empirique vers la fonction de répartition théorique.

Définition 11. Fonction de répartition empirique Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de même loi, de fonction de répartition F . La fonction de répartition empirique du vecteur (X_1, \dots, X_n) est la fonction aléatoire F_n de $\mathbb{R} \times \Omega$ définie par

$$\forall (x, \omega) \in \mathbb{R} \times \Omega, F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^n 1_{X_i(\omega) \leq x}$$

La fonction de répartition empirique est donc une fonction aléatoire en escalier, qui mesure pour chaque x la proportion des X_i qui sont inférieures à x .

Le théorème de Glivenko-Cantelli assure qu'elle converge p.s. uniformément vers la "vraie" fonction de répartition : ainsi lorsqu'on observe un échantillon de loi inconnue, la fonction de répartition empirique permet bien d'avoir une idée assez précise de la fonction de répartition. De plus on sait que la différence entre les deux est d'ordre $\frac{1}{\sqrt{n}}$. Cela sera la base en statistique du test de Kolmogorov Smirnov.

Théorème 12.

Théorème de Glivenko-Cantelli Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes de même loi, de fonction de répartition F . Alors pour presque tout ω , la suite de fonctions de répartition empiriques $F_n(\cdot, \omega)$ converge uniformément vers la fonction de répartition F :

$$\text{Pour presque tout } \omega, \lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| = 0$$

DÉMONSTRATION. On doit démontrer qu'il existe un ensemble Ω' de probabilité 1 tel que pour tout $\omega \in \Omega'$, pour tout entier k , il existe n_0 tel que pour tout $n \geq n_0$, pour tout $x \in \mathbb{R}$, $|F_n(x, \omega) - F(x)| \leq \frac{1}{k}$. Pour alléger les notations on notera $F(x^-) = \mathbb{P}(X < x)$ la limite à gauche en x de F ; de même $F_n(x^-) = \frac{1}{n} \sum_{i=1}^n 1_{X_j < x}$.

On peut déjà commencer par remarquer que si x est un réel fixé quelconque, les événements $1_{X_j \leq x}$ sont indépendants et de même probabilité, donc en vertu du corollaire précédent on a pour tout x réel, il existe un ensemble de probabilité 1 $\Omega_{1,x}$ tel que pour tout $\omega \in \Omega_{1,x}$, $F_n(x, \omega)$ converge vers $F(x)$.

De même on montre que pour tout x réel, il existe un ensemble de probabilité 1 $\Omega_{2,x}$ tel que pour tout $\omega \in \Omega_{2,x}$, $F_n(x^-, \omega)$ converge vers $F(x^-)$. Mais ces ensembles Ω_x dépendent de x or le théorème demande une convergence presque sûre uniforme. On va donc devoir considérer une famille dénombrables de tels x pour pouvoir prendre l'intersection. On va considérer pour tout entier $k > 0$, pour tout $0 \leq j \leq k$,

$$x_{j,k} = \inf \{x \in \mathbb{R}, F(x) \geq \frac{j}{k}\} = F^{-1}\left(\frac{j}{k}\right)$$

où F^{-1} désigne la fonction quantile introduite plus haut, en convenant que $\inf(\mathbb{R}) = -\infty$ et $\inf(\emptyset) = +\infty$. La famille $(x_{j,k})_{k \in \mathbb{N}^*, 0 \leq j \leq k}$ est dénombrable, il existe donc un ensemble Ω' de probabilité 1 tel que la convergence simple ait lieu pour tout ω dans Ω' et pour tout $x_{j,k}$.

Soit $\omega \in \Omega'$ et soit $k \in \mathbb{N}^*$. Il existe donc pour tout $0 \leq j \leq k$ un entier n_j tel que pour $n \geq n_j$,

$$|F_n(x_{j,k}, \omega) - F(x_{j,k})| \leq \frac{1}{k}$$

On pose maintenant $n_0 = \max_{0 \leq j \leq k} n_j$. Soit $n \geq n_0$, et soit $x \in \mathbb{R}$. Si x est l'un des $x_{j,k}$ on n'a rien à démontrer. Sinon, il existe l tel que $x \in]x_{l,k}, x_{l+1,k}[$. La croissance des fonctions considérées permet alors d'écrire

$$F(x_l) \leq F(x) \leq F(x_{l+1}^-)$$

$$F(x_l) - \frac{1}{k} \leq F_n(x_l, \omega) \leq F_n(x, \omega) \leq F_n(x_{l+1}^-, \omega) \leq F(x_{l+1}^-) + \frac{1}{k}$$

Il suffit maintenant de remarquer que, par continuité à droite de F , on a $F(x_l) \geq \frac{l}{k}$; de plus puisque $\forall y < x_{l+1}, F(y) < \frac{j+1}{k}$ on a $F(x_{l+1}^-) \leq \frac{l+1}{k}$: ainsi, $0 \leq F(x_{l+1}^-) - F(x_l) \leq \frac{1}{k}$ et on a finalement

$$|F_n(x, \omega) - F(x)| \leq \frac{2}{k}$$

□

3. Lois faibles des grands nombres et démonstrations

Les énoncés ci-dessous, certes plus faibles, sont aussi plus facilement généralisables, et leur démonstration est tout-à-fait abordable.

Théorème 13.

Loi faible des grands nombres

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi. Si on a $\mathbb{E}[X_1^2] < \infty$, alors

$$\frac{1}{n} (X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{L^2} \mathbb{E}[X_1].$$

DÉMONSTRATION. Par linéarité on a $\mathbb{E} \left[\frac{1}{n} (X_1 + \cdots + X_n) \right] = \mathbb{E}[X_1]$, et par conséquent :

$$\mathbb{E} \left[\left(\frac{1}{n} (X_1 + \cdots + X_n) - \mathbb{E}[X_1] \right)^2 \right] = \frac{1}{n^2} \text{Var}(X_1 + \cdots + X_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n} \text{Var}(X_1)$$

qui tend vers 0 quand $n \rightarrow \infty$. □

La preuve montre que le résultat reste vrai sous des hypothèses moins fortes. Au lieu, par exemple, de supposer que les X_n ont même loi, il suffit de demander que $\mathbb{E}[X_k] = \mathbb{E}[X_1]$ et que la suite $\mathbb{E}[X_n^2]$ soit bornée. Au lieu de l'indépendance, il suffit qu'on ait $\text{Cov}(X_m, X_n) = 0$ dès que $n \neq m$, ce qui est beaucoup plus faible.

Nous donnons un deuxième énoncé de loi forte des grands nombres, avec une hypothèse sur le moment d'ordre 4.

Proposition 14. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi. Si on a $\mathbb{E}[X_1^4] < \infty$, alors

$$\frac{1}{n} (X_1 + \cdots + X_n) \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X_1] \text{ p.s.}$$

DÉMONSTRATION. Quitte à remplacer X_k par $X_k - \mathbb{E}[X_k]$ on peut supposer que les variables sont centrées. Alors on a

$$\mathbb{E} \left[\left(\frac{1}{n} (X_1 + \dots + X_n) \right)^4 \right] = \frac{1}{n^4} \sum_{1 \leq i_1, \dots, i_4 \leq n} \mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}].$$

En utilisant l'indépendance et la propriété $\mathbb{E}[X_k] = 0$, on voit que les seuls termes non nuls de la somme sont ceux pour lesquels deux composantes au chaque valeur prise par une composante du quadruplet (i_1, i_2, i_3, i_4) apparaît au moins deux fois dans ce quadruplet. De plus, comme les X_k ont même loi, on trouve

$$\mathbb{E} \left[\left(\frac{1}{n} (X_1 + \dots + X_n) \right)^4 \right] = \frac{1}{n^4} (n\mathbb{E}[X_1^4] + 3n(n-1)\mathbb{E}[X_1^2 X_2^2]) \leq \frac{C}{n^2}$$

pour une constante positive $C < \infty$. Il en découle que

$$\sum_{n=1}^{\infty} \mathbb{E} \left[\left(\frac{1}{n} (X_1 + \dots + X_n) \right)^4 \right] < \infty$$

et en intervertissant somme et espérance, on obtient

$$\mathbb{E} \left[\sum_{n=1}^{\infty} \left(\frac{1}{n} (X_1 + \dots + X_n) \right)^4 \right] < \infty$$

d'où

$$\sum_{n=1}^{\infty} \left(\frac{1}{n} (X_1 + \dots + X_n) \right)^4 < \infty \text{ p.s.}$$

ce qui entraîne le résultat de la proposition. \square

4. La convergence en loi

Rappelons que $\mathcal{C}_b(\mathbb{R}^d)$ désigne l'espace des fonctions continues bornées de \mathbb{R}^d dans \mathbb{R} , qu'on munit de la norme sup

$$\|\varphi\| = \sup_{x \in \mathbb{R}^d} |\varphi(x)|.$$

Définition 15. Une suite (μ_n) de mesures de probabilité sur \mathbb{R}^d converge étroitement vers une mesure de probabilité μ sur \mathbb{R}^d on note $\mu_n \xrightarrow[n \rightarrow \infty]{(e)} \mu$ si

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^d), \quad \int \varphi d\mu_n \xrightarrow[n \rightarrow \infty]{} \int \varphi d\mu.$$

Définition 16. Une suite de variables aléatoires (X_n) à valeurs dans \mathbb{R}^d converge en loi vers une variable aléatoire X , on note $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ si la suite \mathbb{P}_{X_n} converge étroitement vers \mathbb{P}_X . Cela équivaut encore à

$$\forall \varphi \in \mathcal{C}_b(\mathbb{R}^d), \quad \mathbb{E}[\varphi(X_n)] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[\varphi(X)].$$

Il y a un abus de langage à dire que la suite de variables aléatoires (X_n) converge en loi vers X , car la variable aléatoire limite n'est pas définie de manière unique : seule sa loi l'est. On note d'ailleurs parfois qu'une suite (X_n) converge vers une loi μ , il faudra comprendre que la suite \mathbb{P}_{X_n} converge étroitement

vers μ . C'est la plus faible parmi les différentes notions de convergence. Attention, il n'y a pas obligation que les variables aléatoires soient toutes définies sur le même espace!

Remarque : Le fait qu'on soit dans des espaces de probabilités entraîne qu'on peut remplacer dans la définition ϕ continue bornée par ϕ continue à support compact.

Pour montrer qu'il y a convergence en loi on utilise souvent le résultat suivant :

Théorème 17.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d . La suite converge en loi vers X , si et seulement si la suite des fonctions caractéristiques converge vers la fonction caractéristique de X :

$$\forall t \in \mathbb{R}^d, \Phi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \Phi_X(t).$$

DÉMONSTRATION. $\forall t \in \mathbb{R}^d$, la fonction $x \mapsto e^{i\langle t, x \rangle}$ est continue bornée, donc une implication est claire.

Réciproquement, on suppose que $\forall t \in \mathbb{R}^d, \Phi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \Phi_X(t)$. Soit ψ une fonction continue à support compact sur \mathbb{R}^d . Elle admet une transformée de Fourier $\hat{\psi}$ et on a alors en utilisant Fubini :

$$\mathbb{E}[\psi(X_n)] = \frac{1}{2\pi} \mathbb{E} \left[\int_{\mathbb{R}^d} e^{-i\langle t, X_n \rangle} \hat{\psi}(t) dt \right] = \frac{1}{2\pi} \int_{\mathbb{R}^d} \hat{\psi}(t) \Phi_{X_n}(t) dt$$

On peut alors appliquer le théorème de convergence dominée puis utiliser Fubini en sens inverse pour conclure. \square

Le théorème ci-après apporte quelques précisions.

Théorème 18.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d telles que $\Phi_{X_n} \xrightarrow[n \rightarrow \infty]{} \Psi$ avec Ψ continue en 0, alors Ψ est la fonction caractéristique d'une variable aléatoire X . De plus la suite $(X_n)_{n \geq 1}$ converge en loi vers X .

Voici enfin une traduction de la convergence en loi en terme de fonctions de répartition :

Théorème 19.

Soit $(X_n)_{n \geq 1}$ et X des variables aléatoires réelles dont on note F_{X_n} et F_X les fonctions de répartition. La suite $(X_n)_{n \geq 1}$ converge en loi vers X si et seulement si pour tout point de continuité de F_X , $F_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} F_X(t)$.

Insistons sur le fait qu'il suffit de vérifier la convergence aux points de continuité de la fonction de répartition limite!

DÉMONSTRATION. On suppose que (X_n) converge en loi vers X . Soit x_0 un point de continuité de F_X , et soit $\epsilon > 0$. On considère la fonction ϕ_ϵ la fonction affine par morceaux définie par $\phi_\epsilon(x) = 1$ si

$x \leq x_0$, $\phi_\epsilon(x) = 0$ si $x > x_0 + \epsilon$ et ϕ_ϵ affine sur $[x_0, x_0 + \epsilon]$. On a alors $\mathbb{P}(X_n \leq x_0) \leq \mathbb{E}[\phi_\epsilon(X_n)]$. En passant à la limite, $\limsup F_n(x_0) \leq F(x_0 + \epsilon)$.

On montre de manière analogue $\liminf F_n(x_0) \geq F(x_0 - \epsilon)$. En faisant tendre ϵ vers 0 on obtient bien $\lim F_n(x_0) = F(x_0)$.

Réciproquement, soit ϕ une fonction continue bornée. ϕ est alors limite uniforme de fonctions en escalier. Le nombre de points de discontinuité de F étant dénombrable, on peut s'arranger pour les éviter : si $\epsilon > 0$, il existe une fonction $g = \sum_{j=1}^k a_j 1_{]x_j, x_{j+1}]}$ où les x_j sont des points de continuité de F telle que $\|g - \phi\| \leq \epsilon$. On a alors

$$|\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \leq 2\epsilon + |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]|$$

Or $\mathbb{E}[g(X_n)] = \sum_{i=1}^n a_j (F_n(x_{j+1}) - F_n(x_j))$ converge vers $\mathbb{E}[g(X)]$. On a donc pour tout $\epsilon > 0$

$$0 \leq \limsup |\mathbb{E}[\phi(X_n)] - \mathbb{E}[\phi(X)]| \leq 2\epsilon$$

et on en déduit bien que (X_n) converge en loi vers X . \square

Dans le cas de variables aléatoires discrètes, on a une interprétation assez simple :

Proposition 20. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{N} et X une variable aléatoire à valeurs dans \mathbb{N} . Alors (X_n) converge en loi vers X si et seulement si $\forall k \in \mathbb{N}, \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.

DÉMONSTRATION. On suppose que (X_n) converge en loi vers X . Soit $k \in \mathbb{N}$. Puisque X est à valeurs entières, $\frac{1}{2}$ et $k + \frac{1}{2}$ sont des points de continuité de F_X , donc $\mathbb{P}(X_n) = k = F_{X_n}(k + \frac{1}{2}) - F_{X_n}(k - \frac{1}{2})$ converge vers $\mathbb{P}(X)$.

Réciproquement, si x est un point de continuité de F_X alors $F_{X_n}(x) = \sum_{k=0}^{E(x)} \mathbb{P}(X_n = k)$ est une somme finie qui converge par hypothèse vers $\mathbb{P}(X)$. \square

5. Propriétés de la convergence en loi et lien avec les autres notions de convergence

Attention la convergence en loi n'est pas compatible avec l'addition : si (X_n) converge en loi vers X et (Y_n) converge en loi vers Y on ne peut rien dire sur $(X_n + Y_n)$ (d'ailleurs X et Y ne sont éventuellement même pas définies sur un même espace de probabilités).

D'autre part, si le vecteur (X_n^1, \dots, X_n^d) à valeurs dans \mathbb{R}^d converge en loi vers (Y^1, \dots, Y^d) alors chaque suite de composantes X_n^i converge en loi vers Y^i mais la réciproque est en général fautive !

Néanmoins on peut effectuer certaines opérations :

Proposition 21. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d qui converge en loi vers une variable X , et soit f une fonction continue sur \mathbb{R}^d . Alors $(f(X_n))$ converge en loi vers $f(X)$.

DÉMONSTRATION. En effet si ϕ est continue bornée et si f est continue, $\phi \circ f$ est continue bornée. \square

La convergence en loi est la plus faible des convergences vues jusqu'ici : en effet

Proposition 22. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d qui converge en probabilité vers une variable X . Alors $(X_n)_{n \geq 1}$ converge en loi vers X .

DÉMONSTRATION. En effet soit f une fonction continue bornée sur \mathbb{R}^d . On a alors pour tout $\epsilon > 0$,

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \epsilon + 2\|f\|\mathbb{P}(|X_n - X| > \epsilon)$$

On en déduit en passant à la limite que pour tout $\epsilon > 0$, $0 \leq \limsup(|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]|) \leq \epsilon$. \square

La réciproque est en général fausse, mais elle est vraie lorsque la limite est constante :

Proposition 23. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur le même espace de probabilités, à valeurs dans \mathbb{R}^d , qui converge en loi vers une variable aléatoire a constante. Alors $(X_n)_{n \geq 1}$ converge en probabilités vers a .

DÉMONSTRATION. On fait la démonstration dans le cas $d = 1$ (si on est en dimension supérieure, il suffit de travailler coordonnée par coordonnée). Soit $\epsilon > 0$.

$$\mathbb{P}(|X_n - a| > \epsilon) = \mathbb{P}(X_n > a + \epsilon) + \mathbb{P}(X_n < a - \epsilon) \leq 1 - F_n(a + \epsilon) + F_n(a - \epsilon)$$

Or $a + \epsilon$ et $a - \epsilon$ sont des points de continuité de la fonction de répartition de la variable constante égale à a , on en déduit $\lim_n 1 - F_n(a + \epsilon) + F_n(a - \epsilon) = 0$. \square

Le théorème de Slutsky ci-dessous et son corollaire sont très utiles, notamment en statistiques :

Théorème 24.

Théorème de Slutsky Soient $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ deux suites de variables aléatoires définies sur un même espace de probabilités, à valeurs dans \mathbb{R} , telles que (X_n) converge en loi vers une variable aléatoire X et (Y_n) converge en loi vers une variable aléatoire a constante. Alors le couple (X_n, Y_n) converge en loi vers le couple (X, a) .

On en déduit en particulier que $(X_n + Y_n)$ converge en loi vers $X + a$ et que $(X_n Y_n)$ converge en loi vers aX .

DÉMONSTRATION. On va passer par les fonctions caractéristiques : soit $(s, t) \in \mathbb{R}^2$.

$$|\mathbb{E}[e^{isX_n + itY_n}] - \mathbb{E}[e^{isX + ita}]| = |\mathbb{E}[e^{isX_n}(e^{itY_n} - e^{ita})] + \mathbb{E}[e^{ita}(e^{itX_n} - e^{itX})]| \leq \mathbb{E}[|e^{itY_n} - e^{ita}|] + |e^{ita}| \mathbb{E}[|e^{itX_n} - e^{itX}|]$$

La fonction $|y \mapsto e^{ity} - e^{ita}|$ est continue bornée : on en déduit que $\mathbb{E}[|e^{itY_n} - e^{ita}|]$ converge vers 0. Par hypothèse il en est de même de $\mathbb{E}[|e^{itX_n} - e^{itX}|]$. On en déduit finalement que (X_n, Y_n) converge en loi vers (X, a) .

Il suffit pour les derniers points d'appliquer les fonctions continues $(x, y) \mapsto x + y$ et $(x, y) \mapsto xy$ au couple (X_n, Y_n) pour conclure. \square

6. Théorème central limite

Soit $(X_n)_{n \geq 1}$ une suite de variable aléatoires réelles indépendantes et de même loi, dans L_1 . D'après la loi forte des grands nombres on a

$$\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow[n \rightarrow \infty]{\text{P.S.}} \mathbb{E}[X_1].$$

On peut alors se demander à quelle vitesse cette convergence a lieu, c'est-à-dire trouver un ordre de grandeur de la différence

$$\frac{1}{n}(X_1 + \dots + X_n) - \mathbb{E}[X_1]$$

quand n est grand.

Sous l'hypothèse supplémentaire que les variables X_i sont dans L^2 , on devine la réponse en calculant, comme dans la preuve de la loi faible des grands nombres,

$$\mathbb{E} [(X_1 + \dots + X_n - n\mathbb{E}[X_1])^2] = \text{Var}(X_1 + \dots + X_n) = n\text{Var}(X_1).$$

Ce calcul indique que la valeur moyenne de $(X_1 + \dots + X_n - n\mathbb{E}[X_1])^2$ croît linéairement avec n , donc suggère fortement que l'ordre de grandeur de $X_1 + \dots + X_n - n\mathbb{E}[X_1]$ est \sqrt{n} , ou encore que l'ordre de grandeur de $\frac{1}{n}(X_1 + \dots + X_n) - \mathbb{E}[X_1]$ est $1/\sqrt{n}$. Le théorème central limite rend ceci plus précis.

Théorème 25.

Théorème central limite

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi, dans L^2 . On pose $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \text{Var}(X_1)$, et la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors, la suite de variables aléatoires $\sqrt{n}(\bar{X}_n - \mu)$ converge en loi vers la loi gaussienne $\mathcal{N}(0, \sigma^2)$:

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

En utilisant la traduction de la convergence en loi en termes de fonctions de répartition on obtient ce qui permettra plus tard de donner un intervalle de confiance asymptotique pour l'espérance μ :

Corollaire 26. Avec les mêmes hypothèses et soit $a > 0$ alors

$$\mathbb{P} \left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}} \right] \right) \xrightarrow[n \rightarrow \infty]{} \frac{1}{\sqrt{2\pi}} \int_{-a}^a \exp(-x^2/2) dx.$$

Enfin, lorsque les (X_n) sont des variables de Bernoulli indépendantes de paramètre p , la loi de $\sum_{i=1}^n X_i$ est une loi binomiale; on obtient ainsi le théorème de Moivre-Laplace (au programme - mais sans démonstration - de la classe de Terminale S)

Théorème de Moivre-Laplace

Corollaire 27. Soit (S_n) une suite de variables aléatoires de loi binomiale de paramètres (n, p) . Alors la suite $Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}$ converge en loi vers une loi normale centrée réduite.

DÉMONSTRATION. Quitte à remplacer X_n par $X_n - \mathbb{E}[X_n]$ on peut supposer que les variables aléatoires sont centrées. Posons

$$Z_n = \frac{\sum_{k=1}^n X_k}{\sqrt{n}}.$$

Par indépendance des X_n , la fonction caractéristique de Z_n est alors :

$$\Phi_{Z_n}(t) = \mathbb{E} \left[\exp \left(it \frac{\sum_{k=1}^n X_k}{\sqrt{n}} \right) \right] = \prod_{k=1}^n \mathbb{E} \left[\exp \left(it \frac{X_k}{\sqrt{n}} \right) \right] = \mathbb{E} \left[\exp \left(it \frac{X_1}{\sqrt{n}} \right) \right]^n = \Phi_{X_1} \left(\frac{t}{\sqrt{n}} \right)^n$$

Comme X_1 est de carré intégrable, Φ_{X_1} est deux fois dérivable en 0 et on peut écrire :

$$\Phi_{X_1} \left(\frac{t}{\sqrt{n}} \right) = 1 - \frac{\sigma^2 t^2}{2n} + o \left(\frac{t}{n} \right).$$

En passant à la limite on obtient

$$\lim_{n \rightarrow +\infty} \Phi_{X_1} \left(\frac{t}{\sqrt{n}} \right)^n = \exp \left(-\frac{\sigma^2 t^2}{2} \right)$$

qui est la fonction caractéristique de la loi $\mathcal{N}(0, \sigma^2)$. □

On peut généraliser le théorème central limite à d dimensions ; la loi limite fait alors intervenir les vecteurs gaussiens.

Théorème 28.

Théorème central limite multidimensionnel Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d , indépendantes, de même loi, et dans L^2 . On note $\mu = \mathbb{E}[X_1]$, C la matrice de covariance de X_1 , et la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors, la suite de variables aléatoires $\sqrt{n}(\bar{X}_n - \mu)$ converge en loi vers un vecteur gaussien de loi $\mathcal{N}(0, C)$.

Enfin la généralisation suivante (parfois appelé Delta-Method) est parfois utilisé en statistique pour démontrer une convergence vers une loi normale ; plutôt que de l'apprendre par coeur, il vaut mieux savoir le redémontrer au cas par cas.

Proposition 29. Delta-méthode Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. On suppose qu'il existe deux réels θ et σ tels que la suite $\sqrt{n}(X_n - \theta)$ converge en loi vers une loi $\mathcal{N}(0, \sigma^2)$. Soit g une fonction définie sur \mathbb{R} , dérivable en θ telle que $g'(\theta) \neq 0$. Alors la suite $\sqrt{n}(g(X_n) - g(\theta))$ converge en loi vers une loi $\mathcal{N}(0, \sigma^2(g'(\theta))^2)$

DÉMONSTRATION. On commence par remarquer que $X_n - \theta$ converge en loi vers 0, donc X_n converge en probabilité vers θ .

D'autre part on considère la fonction h définie par $h(x) = \frac{g(x) - g(\theta)}{x - \theta}$ si $x \neq \theta$ et $h(\theta) = g'(\theta)$. Alors h est continue. Par conséquent $h(X_n)$ converge en probabilité vers $h(\theta)$. Il suffit alors d'écrire $\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n}(X_n - \theta)h(X_n)$ et d'utiliser le théorème de Slutsky pour en déduire que $\sqrt{n}(g(X_n) - g(\theta))$ converge en loi vers $g'(\theta)\mathcal{N}(0, \sigma^2)$. □

7. Convergence vers la loi de Poisson

Ce théorème explique pourquoi la loi de Poisson est aussi appelée loi des événements rares.

Théorème 30.

Soit $(S_n)_{n \geq 1}$ une suite de variables aléatoires de binomiales de paramètre λ et d'ordre n , i.e. $S_n \stackrel{\mathcal{L}}{\sim} \mathcal{B}(n, \frac{\lambda}{n})$. Alors, $(S_n)_{n \geq 1}$ converge en loi vers la loi de Poisson de paramètre λ .

DÉMONSTRATION. Les variables aléatoires S_n peuvent être vues comme des sommes de variables de Bernoulli indépendantes, plus précisément on peut poser $S_n = \sum_{i=1}^n X_{n,i}$ où les $X_{n,i}$ sont des variables aléatoires de Bernoulli, indépendantes, de paramètre λ/n . Comme on travaille avec des variables aléatoires discrètes on va étudier les fonctions génératrices. On a :

$$G_{X_{n,i}}(z) = 1 - \frac{\lambda}{n} + \frac{\lambda}{n}z = 1 + \frac{\lambda}{n}(z - 1)$$

d'où par indépendance on obtient

$$G_{S_n}(z) = \left[1 + \frac{\lambda}{n}(z - 1) \right]^n \xrightarrow{n \rightarrow +\infty} \exp(\lambda(z - 1))$$

qui est la fonction génératrice de la loi de Poisson de paramètre λ . □

8. Appendice : démonstration de la loi forte des grands nombres

Le but de ce paragraphe est de montrer que si (X_n) est une suite de variables aléatoires indépendantes et de même loi, dans L^1 , alors les moyennes $\frac{1}{n}(X_1 + \dots + X_n)$ convergent p.s. vers $\mathbb{E}[X_1]$, sous des hypothèses optimales.

Rappelons d'abord la loi du 0-1.

Proposition 31. Loi du tout ou rien

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, à valeurs dans des espaces mesurables quelconques. Pour tout $n \geq 1$ on pose $\mathcal{B}_n = \sigma(X_k; k \geq n)$. Alors la tribu terminale (ou asymptotique) \mathcal{B}_∞ définie par

$$\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n$$

est grossière au sens où $\mathbb{P}(B) = 0$ ou 1 pour tout $B \in \mathcal{B}_\infty$.

Théorème 32.**Loi forte des grands nombres**

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées, intégrables alors

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{n} = \mathbb{E}[X_1] \right) = 1$$

En statistique $\frac{S_n}{n} = \frac{\sum_{k=1}^n X_k}{n}$ est appelée la moyenne empirique (ou arithmétique) des X_k et est notée \bar{X}_n .

Une autre manière d'énoncer ce résultat est de dire :

- pour toute suite de variables aléatoires indépendantes identiquement distribuées, on a

$$\frac{\sum_{k=1}^n X_k}{n} \text{ converge p.s.} \iff \mathbb{E}[|X_1|] < +\infty.$$

De plus si l'une des deux conditions est vérifiée alors la limite de la moyenne empirique est $\mathbb{E}[X_1]$.

DÉMONSTRATION. Supposons dans un premier temps que les X_n sont centrées. On verra à la fin de la démonstration comment supprimer cette hypothèse.

ETAPE 1 : troncature

On pose

$$X'_n = X_n 1_{|X_n| \leq n} \quad \text{et} \quad S'_n = X'_1 + \dots + X'_n.$$

Lemme 33. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes identiquement distribuées, intégrables alors

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X_k}{n} = 0\right) = 1 \iff \mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n X'_k}{n} = 0\right) = 1.$$

En effet, on a en posant $A_n = \{\omega \in \Omega \mid X_n(\omega) \neq X'_n(\omega)\}$

$$\begin{aligned} \sum_{n \geq 1} \mathbb{P}(A_n) &= \sum_{n \geq 1} \mathbb{P}(X_n \neq X'_n) = \sum_{n \geq 1} \mathbb{P}(|X_n| > n) = \sum_{n \geq 1} \mathbb{P}(|X_1| > n) \\ &= \sum_{n \geq 1} \mathbb{E}[1_{|X_1| > n}] = \mathbb{E}\left[\sum_{n \geq 1} 1_{|X_1| > n}\right] \end{aligned}$$

cette dernière égalité étant une conséquence du théorème de convergence monotone. Remarquons que pour $x \geq 0$ on a

$$\phi(x) = \sum_{n \geq 1} 1_{x > n} = [x] - 1 \leq x$$

On obtient donc

$$\sum_{n \geq 1} \mathbb{P}(A_n) = \mathbb{E}\left[\sum_{n \geq 1} 1_{|X_1| > n}\right] = \mathbb{E}[\phi(|X_1|)] \leq \mathbb{E}[|X_1|] < +\infty$$

D'après le lemme de Borel-Cantelli, il suit que $\mathbb{P}(\limsup A_n) = 0$, ou encore que $\mathbb{P}((\limsup A_n)^c) = \mathbb{P}(\liminf (A_n^c)) = 1$.

Par conséquent si $\omega \in E = (\limsup A_n)^c$ alors la série $\sum_{k \geq 1} (X_k(\omega) - X'_k(\omega))$ est une série convergente, puisque, en dehors d'un nombre fini d'entre eux, tous les termes sont nuls. Ainsi la suite des sommes partielles

$$\sum_{k=1}^n (X_k(\omega) - X'_k(\omega)) = S_n(\omega) - S'_n(\omega),$$

est une suite convergente et donc bornée, ce qui entraîne que

$$\lim_{n \rightarrow \infty} \frac{S_n(\omega) - S'_n(\omega)}{n} = 0.$$

D'où, en posant $F = \{\omega \in \Omega, \lim_{n \rightarrow \infty} \frac{S_n(\omega) - S'_n(\omega)}{n} = 0\}$, on obtient $\mathbb{P}(F) \geq \mathbb{P}(E) = \mathbb{P}((\limsup A_n)^c) = 1$.

Maintenant, si on note $G_1 = \{\omega \in \Omega, \lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = 0\}$ et $G_2 = \{\omega \in \Omega, \lim_{n \rightarrow \infty} \frac{S'_n(\omega)}{n} = 0\}$ alors on a clairement

$$F \cap G_1 \subset G_2 \quad F \cap G_2 \subset G_1$$

on en déduit donc

$$\mathbb{P}(G_1) = 1 \iff \mathbb{P}(G_2) = 1,$$

ce qui démontre le lemme.

ETAPE 2 : recentrage

Nous allons maintenant centrer les X'_k . Pour cela définissons :

$$\begin{aligned}\tilde{X}_k &= X'_k - \mathbb{E}[X'_k] \\ \tilde{S}_k &= \tilde{X}_1 + \dots + \tilde{X}_k\end{aligned}$$

et remarquons que :

Lemme 34. *Avec les notations précédentes :*

$$\mathbb{P}\left(\omega \in \Omega \mid \lim_{n \rightarrow \infty} \frac{S'_n}{n} = 0\right) = 1 \iff \mathbb{P}\left(\omega \in \Omega \mid \lim_{n \rightarrow \infty} \frac{\tilde{S}_n}{n} = 0\right) = 1.$$

En effet, on a

$$\frac{S'_n}{n} - \frac{\tilde{S}_n}{n} = \frac{\mathbb{E}[X'_1] + \dots + \mathbb{E}[X'_n]}{n}$$

qui ne dépend pas de ω . D'autre part, comme les X_k ont même loi, on a

$$\lim_{n \rightarrow \infty} \mathbb{E}[X'_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n 1_{|X_n| \leq n}] = \lim_{n \rightarrow \infty} \mathbb{E}[X_1 1_{|X_1| \leq n}].$$

De plus on constate que $\lim_{n \rightarrow \infty} X_1 1_{|X_1| \leq n} = X_1$ et $|X_1 1_{|X_1| \leq n}| \leq |X_1|$ qui est intégrale ; ce qui nous permet d'appliquer le théorème de convergence dominée, et d'obtenir

$$\lim_{n \rightarrow \infty} \mathbb{E}[X'_n] = \lim_{n \rightarrow \infty} \mathbb{E}[X_1 1_{|X_1| \leq n}] = \mathbb{E}[\lim_{n \rightarrow \infty} X_1 1_{|X_1| \leq n}] = \mathbb{E}[X_1] = 0.$$

Enfin, par le lemme de Cesaro, on obtient, pour tout $\omega \in \Omega$

$$\lim_{n \rightarrow \infty} \frac{S'_n(\omega)}{n} - \frac{\tilde{S}_n(\omega)}{n} = 0.$$

ETAPE 3 : inégalité de Kolmogorov

Nous allons maintenant démontrer une inégalité due à Kolmogorov (on utilisera ici l'indépendance des variables aléatoires).

Lemme 35. *Soit $(Y_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et centrées. Posons :*

$$V_n = Y_1 + \dots + Y_n$$

Alors, pour tout $x > 0$,

$$\mathbb{P}\left(\sup_{n \geq 1} |V_n| > x\right) \leq \frac{\sum_{k \geq 1} \text{Var}(Y_k)}{x^2}$$

Si la somme des variances diverge le résultat est vrai, on peut donc supposer $\sum_{k \geq 1} \text{Var}(Y_k) < +\infty$. On pose

$$\tau = \begin{cases} \inf\{k \geq 1, |V_k| > x\} \\ +\infty \end{cases} \quad \text{si } \{k \geq 1, |V_k| > x\} = \emptyset$$

Remarquons alors que, pour $k \leq n$, $V_k 1_{\tau=k} \perp V_n - V_k$. En effet, on a

$$\begin{aligned}\{\tau = k\} &= \{|V_1| \leq x, \dots, |V_{k-1}| \leq x, |V_k| > x\} \\ &= \{|Y_1| \leq x, |Y_1 + Y_2| \leq x, \dots, |Y_1 + \dots + Y_{k-1}| \leq x, |Y_1 + \dots + Y_k| > x\}\end{aligned}$$

est $\sigma(Y_1, \dots, Y_k)$ -mesurable alors que $V_n - V_k = Y_{k+1} + \dots + Y_n$ est $\sigma(Y_{k+1}, \dots, Y_n)$ -mesurable. L'indépendance de variables aléatoires entraîne donc le résultat. Ceci nous permet donc d'obtenir que

$$\begin{aligned} \sum_{k=1}^n \text{Var}(Y_k) &= \text{Var}(V_n) = \mathbb{E}[V_n^2] \\ &\geq \sum_{k=1}^n \mathbb{E}[V_n^2 1_{\tau=k}] = \sum_{k=1}^n \mathbb{E}[(V_n - V_k + V_k)^2 1_{\tau=k}] \\ &\geq \sum_{k=1}^n \mathbb{E}[V_k^2 1_{\tau=k}] \\ &\geq \mathbb{E}[x^2 \sum_{k=1}^n 1_{\tau=k}] = x^2 \mathbb{P}(\tau \leq n) \end{aligned}$$

En faisant tendre n vers l'infini on obtient

$$\sum_{k \geq 1} \text{Var}(Y_k) \geq x^2 \mathbb{P}(\tau < +\infty) \geq x^2 \mathbb{P}\left(\sup_{n \geq 1} |V_n| > x\right)$$

ETAPE 4 : convergence de séries de variable aléatoires

L'inégalité de Kolmogorov est, avec le lemme de Borel-Cantelli, l'ingrédient essentiel de la preuve de lemme suivant.

Lemme 36. *Soit une suite $(U_n)_{n \geq 1}$ de variables aléatoires réelles indépendantes et centrées. Si*

$$\sum_{k \geq 1} \text{Var}(U_k) < +\infty$$

alors la suite $Z_n = \sum_{k=1}^n U_k$ est convergente, ou encore, la série $\sum_{k \geq 1} U_k$ est convergente.

On applique l'inégalité de Kolmogorov à la suite $Y_n = U_{M+n}$. Avec les notations prises dans le lemme, on a

$$\begin{aligned} V_n &= Z_{M+n} - Z_M \\ \sum_{k \geq 1} \text{Var}(Y_k) &= \sum_{k \geq M} \text{Var}(U_k) \end{aligned}$$

et l'inégalité de Kolmogorov donne

$$\mathbb{P}\left(\sup_{n \geq 1} |Z_{M+n} - Z_M| > x\right) \leq \frac{\sum_{k \geq M} \text{Var}(U_k)}{x^2}.$$

On peut remarquer que la suite $\tilde{Z}_M = \sup_{n, m \geq 1} |Z_{M+n} - Z_{M+m}| = \sup_{k, l > M} |Z_k - Z_l|$ est décroissante. De plus elle vérifie

$$\tilde{Z}_M \leq \sup_{n, m \geq 1} (|Z_{M+n} - Z_M| + |Z_M - Z_{M+m}|) = 2 \sup_{n \geq 1} |Z_{M+n} - Z_M|.$$

On en déduit que, pour tout $k \geq 1$ et $M \geq 1$,

$$\mathbb{P}(\tilde{Z}_M > 1/k) \leq \mathbb{P}\left(\sup_{n \geq 1} |Z_{M+n} - Z_M| > 1/2k\right) \leq 4k^2 \sum_{n \geq M} \text{Var}(U_n).$$

Comme $\sum_{n \geq M} \text{Var}(U_n)$ tend vers 0 on peut choisir, pour tout $k \geq 1$, $M_k > M_{k-1}$ tel que

$$\mathbb{P}(\tilde{Z}_{M_k} > 1/k) \leq 2^{-k}.$$

Ainsi

$$\sum_{k \geq 1} \mathbb{P}(\tilde{Z}_{M_k} > 1/k) < +\infty$$

et le lemme de Borel-Cantelli entraîne que, presque sûrement, à partir d'un certain rang, \tilde{Z}_{M_k} est majorée par $1/k$, et donc que \tilde{Z}_{M_k} converge presque sûrement vers 0. Par ailleurs on a vu que \tilde{Z}_M est une suite décroissante. On a une suite décroissante, dont une sous-suite est convergente, elle est donc elle-même convergente. Par conséquent, \tilde{Z}_M converge presque sûrement vers 0. Or

$$\begin{aligned} \{\lim_M \tilde{Z}_M = 0\} &\iff \{Z_n \text{ est une suite de Cauchy}\} \\ &\iff \{Z_n \text{ est une suite convergente}\} \\ &\iff \{\sum_n U_n \text{ est une série convergente}\} \end{aligned}$$

ETAPE 5 : lemme de Kronecker Le point suivant est l'utilisation d'un lemme d'analyse du à Kronecker.

Lemme 37. Soit a_1, a_2, \dots une suite de réels tels que $\sum_{n \geq 1} (a_n/n)$ converge, alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n a_k = 0.$$

On pose $s_0 = 0$ et $s_n = \sum_{k=1}^n (a_k/k)$ et $R_n = \sum_{k=1}^n a_k$. Alors, par hypothèse, s_n converge vers une limite s . Comme $a_k = k(s_k - s_{k-1})$, on en déduit

$$R_{n+1} = \sum_{k=1}^{n+1} a_k = \sum_{k=1}^{n+1} k s_k - \sum_{k=1}^n k s_{k-1} = - \sum_{k=1}^n s_k + (n+1) s_{n+1}.$$

Par conséquent,

$$\frac{1}{n+1} R_{n+1} = s_{n+1} - \frac{n}{n+1} \frac{1}{n} \sum_{k=1}^n s_k \rightarrow s - 1 \cdot s = 0.$$

ETAPE 6 : conclusion dans le cas centré On applique maintenant ce lemme.

Lemme 38. En reprenant les notations du lemme 10 on a

$$\sum_{k \geq 1} \text{Var} \left(\frac{\tilde{X}_k}{k} \right) < +\infty.$$

Les calculs se font mieux si on remplace k au dénominateur par $k+1$ (les deux séries étant égales à une constante prêt). Pour $k \geq 2$ on a

$$\text{Var} \left(\frac{\tilde{X}_k}{k+1} \right) = \frac{\mathbb{E}[X_k'^2]}{(k+1)^2} - \frac{\mathbb{E}[X_k']^2}{(k+1)^2}.$$

On a $\frac{\mathbb{E}[X'_k]^2}{(k+1)^2} = o\left(\frac{1}{k^2}\right)$. La convergence de la série de terme général est donc équivalente à la convergence de la série

$$\sum_{k \geq 1} \frac{\mathbb{E}[X_k'^2]}{(k+1)^2}.$$

Or, on a

$$\begin{aligned} \sum_{k \geq 1} \frac{\mathbb{E}[X_k'^2]}{(k+1)^2} &= \sum_{k \geq 1} (k+1)^{-2} \mathbb{E}[X_1^2 1_{0 < |X_1| \leq k}] \leq \sum_{k \geq 1} \int_k^{k+1} x^{-2} \mathbb{E}[X_1^2 1_{0 < |X_1| \leq x}] dx \\ &= \int_1^{+\infty} x^{-2} \mathbb{E}[X_1^2 1_{0 < |X_1| \leq x}] dx = \mathbb{E} \left[X_1^2 1_{|X_1| > 0} \int_1^{+\infty} x^{-2} 1_{|X_1| \leq x} dx \right] \\ &\leq \mathbb{E} \left[X_1^2 1_{|X_1| > 0} \int_{|X_1|}^{+\infty} x^{-2} dx \right] = \mathbb{E}[X_1^2 1_{|X_1| > 0} |X_1|^{-1}] \leq \mathbb{E}[|X_1|] < +\infty \end{aligned}$$

Le résultat de ce lemme combiné avec celui du lemme 12 nous donne que la série $\sum_{n \geq 1} \frac{\tilde{X}_k}{k}$ converge presque sûrement, et par le lemme de Kronecker, on en déduit que, presque sûrement,

$$\lim_{n \rightarrow +\infty} \frac{\tilde{S}_n}{n} = 0.$$

ETAPE 7 : décentrage et conclusion générale Si on ne suppose plus les X_n centrées, mais seulement i.i.d et intégrables, on pose

$$\hat{X}_k = X_k - \mathbb{E}[X_k], \quad \hat{S}_n = \sum_{k=1}^n \hat{X}_k.$$

Les \hat{X}_k étant centrées i.i.d et intégrables, les étapes précédentes montrent que

$$\mathbb{P} \left(\omega \in \Omega \mid \lim_{n \rightarrow +\infty} \frac{\hat{S}_n(\omega)}{n} = 0 \right) = 1.$$

Or, on remarque que

$$\frac{\hat{S}_n(\omega)}{n} = \frac{S_n(\omega) - n\mathbb{E}[X_1]}{n} = \frac{S_n(\omega)}{n} - \mathbb{E}[X_1];$$

d'où

$$\mathbb{P} \left(\omega \in \Omega \mid \lim_{n \rightarrow +\infty} \frac{\hat{S}_n(\omega)}{n} = 0 \right) = \mathbb{P} \left(\omega \in \Omega \mid \lim_{n \rightarrow +\infty} \frac{S_n(\omega)}{n} = \mathbb{E}[X_1] \right)$$

ce qui démontre le résultat. □

Bibliographie

- [1] Jean-Yves Oувrard *Probabilités 1 et 2*, Cassini
- [2] Marie Cottrell et al. *Exercices de probabilités*, Cassini
- [3] Dominique Foata, Jacques Franchi, Aimé Fuchs *Calcul des probabilités. Cours, exercices et problèmes corrigés*, Dunod
- [4] Dominique Foata et Aimé Fuchs *Processus stochastiques. Processus de Poisson, Chaînes de markov et martingales*, Dunod
- [5] Bernard Bercu et Djalil Chafaï *Modélisation stochastique et simulation. Cours et applications*, Dunod
- [6] Vincent Rivoirard et Gilles Stoltz *Statistique mathématique en action*, Vuibert
- [7] Benoit Cadre et Céline Vial *Statistique mathématique. Cours et exercices corrigés*, Ellipses
- [8] Geoffrey Grimmett and David Stirzaker *Probability and random processes*, Oxford