Université Paris-Dauphine **2022-2023**
Cours de “Intégrale de Lebesgue et probabilités" Novembre 2022

Chapitre 7 - Fondement des probabilités

Table des matières

Dans ce chapitre on désigne par $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé, c'est-à-dire, un espace mesurable $(\Omega, \mathscr{A})$ muni d'une probabilité $\mathbf{P}$.

Sont écrites en rouge les parties hors programme, en violet les parties traitées en TD (résultats à connaitre pour sa culture) et en orange les parties modifiées (précisées) par rapport au cours en amphi.

1 Variable aléatoire, loi d'une variable aléatoire, espérance

Une variable aléatoire X est une application mesurable définie sur un espace probabilisé $(\Omega, \mathscr{A}, P)$ et à valeurs dans un espace mesurable $(E, \mathcal{E})$. Sans précision supplémentaire, on utilisera habituellement le terme variable aléatoire pour désigner une variable aléatoire réelle, c'est-à-dire, à valeurs dans $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. On utilisera également l'acronyme “var”. On dira que X est un vecteur aléatoire lorsque $(E, \mathcal{E}) = (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, $d \geq 2$.

Dans ce cours, l'espace probabilisé $(\Omega, \mathscr{A}, P)$ sera fixé et on considérera diverses variables aléatoires définies sur cet espace (dont on étudiera des propriétés). Au moment de “construire” les objets $(\Omega, \mathscr{A}, P)$ et X, il est parfois utile de procéder dans le sens inverse. On fixe $(\Omega, \mathscr{A})$, $(E, \mathcal{E}) = (\Omega, \mathscr{A})$, $X =$ l'identité, soit donc $X(\omega) = \omega$ pour tout $\omega \in \Omega$, et on construit P.

Le vocabulaire en théorie des probabilités est sensiblement différent de celui de la théorie de la mesure et de l'intégration vu jusqu'à maintenant. On le gardera pour des objets “probabilistes” qui sont par essence connus “sans certitude”. Un petit lexique de correspondances est donné à la fin du chapitre. On dira qu'une propriété est vraie presque sûrement (p.s.) si elle est vraie pour tout $\omega \in A \in \mathcal{A}$ avec $P(\Omega \backslash A) = 0$.

Définition 1.1. *On appelle loi de X la mesure image $X_\sharp P$ de P par X sur $(E, \mathcal{E})$. On notera $\mathcal{L}(X) = P^X = X_\sharp P\,(= X(P))$ et $X \sim \mathcal{L}(X)$. On a donc par définition*

$$\forall\, B \in \mathcal{E}, \quad P^X(B) := P(X \in B) = P(X^{-1}(B)).$$

En particulier, P^X est une mesure de probabilité.

Exemples 1.2 (de mesures/lois de probabilité)**.**
(i) Dans $\Omega := \{a, b\}$, on appelle loi de Bernouilli de paramètre $p \in [0,1]$ la mesure de probabilité

$$P := p\delta_b + (1-p)\delta_a.$$

C'est la loi de probabilité d'un lancé d'une pièce (éventuellement truquée) dont la réalisation est **a** *(disons pile) avec probabilité $1-p$ et* **b** *(disons face) avec probabilité p.*
(ii) Dans $\Omega := \{1, \ldots, n\}$, on appelle loi uniforme la mesure de probabilité

$$P := \sum_{i=1}^{n} \frac{1}{n}\delta_i.$$

(iii) Dans $\Omega := \{0, \ldots, n\}$, on appelle loi binomiale de paramètres $n \geq 1$ et $p \in [0,1]$, on note $\mathcal{B}(n,p)$, la mesure de probabilité

$$P := \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k}\delta_k.$$

La loi $\mathcal{B}(1,p)$ est une loi de Bernouilli sur $\{0,1\}$.
(iv) Dans $\Omega := \mathbb{N}$, on appelle loi de Poisson de paramètre $\lambda > 0$, on note $\mathcal{P}(\lambda)$, la mesure de probabilité

$$P := e^{-\lambda}\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\delta_k.$$

(v) Dans $\Omega := [a,b]$, on appelle loi uniforme, on note $\mathcal{U}(a,b)$, la mesure de probabilité

$$P := \frac{1}{b-a}\,dx$$

(vi) Dans $\Omega := \mathbb{R}_+$, on appelle loi exponentielle de paramètre $\lambda > 0$, on note $\mathcal{E}xp(\lambda)$, la mesure de probabilité

$$P := \lambda e^{-\lambda x}\,dx.$$

Définition 1.3. *Dans $\Omega := \mathbb{R}$, on appelle loi gaussienne unidimensionnelle $\mathcal{N}(\mu, \sigma^2)$ de paramètres $\sigma > 0$ et $\mu \in \mathbb{R}$, la mesure de probabilité*

$$P = g_{\lambda,\sigma}(x)\,dx, \quad g_{\lambda,\sigma} := \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Par extension, on appelle loi gaussienne unidimensionnelle $\mathcal{N}(\mu, 0)$ de paramètres $\sigma = 0$ et $\mu \in \mathbb{R}$, la mesure de Dirac δ_μ. On dit que la loi gaussienne est
- (centrée) réduite *si $\mu = 0$ et $\sigma = 1$;*
- centrée *si $\mu = 0$;*
- dégénérée *si $\sigma = 0$ (elle n'a alors pas de densité par rapport à la mesure de Lebesgue).*

Définition 1.4. - *Soit X une variable aléatoire à valeurs positives. On appelle espérance de X, on note $\mathbf{E}(X)$, la quantité*

$$\mathbf{E}(X) = \int_\Omega X dP \in [0, \infty].$$

- De même si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$ telle que $\mathbf{E}(|X|) < \infty$, on dit que X est intégrable, on note $X \in \mathcal{L}^1$, et on appelle espérance de X, on note $\mathbf{E}(X)$, la quantité définie ci-dessus. On dit que $X \in \mathcal{L}^1$ est centrée si $\mathbf{E}(X) = 0$.

Lemme 1.5 (de transport)**.** *Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$, $\phi : (E; \mathcal{E}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ une fonction mesurable avec $\phi \geq 0$ ou $\phi \circ X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, alors*

$$\mathbf{E}(\phi(X)) = \int_\Omega \phi \circ X dP = \int_E \phi(x)\, dP^X(x),$$

et ces quantités sont bien définies.

L'inégalité de Tchebychev s'écrit également

Proposition 1.6 (Inégalité de Markov)**.** *Si $X \geq 0$ et $a > 0$, alors*

$$\mathbf{P}(X \geq a) \leq \frac{1}{a}\mathbf{E}(X).$$

Définition 1.7. *Pour une X une variable aléatoire réelle (ou vectorielle), on notera $X \in \mathcal{L}^p$ si $\mathbf{E}(|X|^p) < \infty$ pour $p \in [1, \infty)$, $|X| \leq C$ p.s. si $p = \infty$. On dira que X est intégrable si $X \in \mathcal{L}^1$ et de carré intégrable si $X \in \mathcal{L}^2$.*

Il convient de remarquer que $X \in \mathcal{L}^p$, $p \in [1, \infty]$, implique $X \in \mathcal{L}^q$ pour tout $q \in [1, p[$, soit donc $\mathcal{L}^p \subset \mathcal{L}^q$. C'est une conséquence de l'inégalité de Hölder. Losque $1 \leq q < p < \infty$, on a en effet avec $r := p/q \in]1, \infty[$,

$$\mathbf{E}(|X|^q) \leq (\mathbf{E}(|X|^p))^{1/r}(\mathbf{E}(\mathbf{1}))^{1/r'} = (\mathbf{E}(|X|^p)^{1/r} < \infty.$$

Lorsque $1 \leq q < p = \infty$, on a plus simplement $\mathbf{E}(|X|^q) \leq \|X\|^q_{\mathcal{L}^\infty}\mathbf{E}(1) = \|X\|^q_{\mathcal{L}^\infty} < \infty$.

Définition 1.8. *Soit X une variable aléatoire réelle de carré intégrable, son espérance $\mathbf{E}(X)$ est donc définie. On appelle variance de X, on note $\mathrm{Var}(X)$, la quantité*

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

Il convient d'observer que $\mathrm{Var}(X - \mathbf{E}(X)) = \mathrm{Var}(X)$ et que si X est centrée alors $\mathrm{Var}(X) = \mathbf{E}(X^2)$. Plus généralement, si X une variable aléatoire $\mathcal{L}^2$ à valeurs dans $\mathbb{R}^d$, on définit la matrice de covariance comme étant la matrice de coefficients

$$\mathrm{cov}(X)_{ij} = \mathrm{cov}(X_j, X_k) = \mathbf{E}[(X_j - \mathbf{E}X_j)(X_k - \mathbf{E}X_k)] = \mathbf{E}(X_jX_k) - \mathbf{E}(X_j)\mathbf{E}(X_k).$$

Exercice 1.9. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, $a \in \mathbb{R}^d$ et $M \in M_{d,k}(\mathbb{R})$. Montrer que $\mathrm{cov}(X + a) = \mathrm{cov}(X)$ et $\mathrm{cov}(MX) = M\mathrm{cov}(X)\,{}^tM$.*

Exemples 1.10. *- Si $X \sim \mathcal{B}(n, p)$, alors $\mathbf{E}(X) = np$ et $Var(X) = np(1-p)$.*
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $\mathbf{E}(X) = \mu$ et $Var(X) = \sigma^2$.
- Si $X \sim \mathcal{P}(\lambda)$, alors $\mathbf{E}(X) = \lambda$ et $Var(X) = \lambda$.
- Si $X \sim \mathcal{E}xp(\lambda)$, alors $\mathbf{E}(X) = 1/\lambda$ et $Var(X) = 1/\lambda^2$.

Proposition 1.11. *Si $X \in \mathcal{L}^2$ alors*

$$\mathbf{E}[(X - a)^2] = Var(X) + (\mathbf{E}(X) - a)^2, \quad \forall\, a \in \mathbb{R},$$

et en particulier

$$Var(X) = \min_{a \in \mathbb{R}} \mathbf{E}[(X - a)^2].$$

Preuve de la Proposition 1.11. Il suffit de développer

$$\mathbf{E}[(X-a)^2] = \mathbf{E}[X^2] - 2a\mathbf{E}[X] + a^2 = \mathbf{E}[X^2] - (\mathbf{E}(X))^2 + (\mathbf{E}(X)-a)^2,$$

et de prendre éventuellement $a := \mathbf{E}(X)$. □

Proposition 1.12 (inégalité de Bienaymé-Tchebychev)**.** *Si $X \in \mathcal{L}^2$ et $a > 0$, alors*

$$\mathbf{P}(|X - \mathbf{E}X| \geq a) \leq \frac{1}{a^2}\,\mathrm{Var}(X).$$

Preuve de la Proposition 1.12. Il suffit d'appliquer l'inégalité de Markov à la var $(X - \mathbf{E}X)^2$. □

Il existe différentes façons de caractériser la loi d'une variable aléatoire.

Définition 1.13. *Pour un vecteur aléatoire X, on appelle fonction caractéristique de X, on note φ^X, la transformée de Fourier de la loi P^X qui est donc définie par*

$$\varphi^X(\xi) := \int_{\mathbb{R}^d} e^{i\xi\cdot x}\,dP^X(x) = \mathbf{E}(e^{i\xi\cdot X}), \quad \forall\,\xi \in \mathbb{R}^d.$$

La fonction caractéristique φ^X caractérise la loi de X, puisque la transformation de Fourier caractérisant une mesure de probabilité : si φ^X est la transformation de Fourier d'une mesure de probabilité μ, alors $X \sim \mu$.

Exemples 1.14. *(i) $\varphi(t) := (1 - it/\lambda)^{-1}$ est fonction caractéristique de $X \sim \mathcal{E}xp(\lambda)$.*
(ii) $\varphi(t) := \exp(\lambda(e^{it} - 1))$ est fonction caractéristique de $X \sim \mathcal{P}(\lambda)$.
(iii) $\varphi(t) := \exp(\mu i t - \sigma^2 t^2/2)$ est fonction caractéristique de $X \sim \mathcal{N}(\mu,\sigma)$.

Définition 1.15. *Pour une va réelle X, on appelle fonction de répartition de X, on note F^X, la fonction de répartition de la loi P^X. On a donc*

$$\forall\, t \in \mathbb{R}, \quad F^X(t) = P^X(]-\infty,t]) = P(X \leq t) = \mathbf{E}(\mathbf{1}_{X\leq t}).$$

Plus généralement, pour un vecteur aléatoire X, on appelle fonction de répartition de X, on note F^X, la fonction de répartition de la loi P^X. On a donc

$$\begin{aligned}\forall\,(t_1,\dots,t_d) \in \mathbb{R}^d, \quad F^X(t_1,\dots,t_d) &= P^X(]-\infty,t_1]\times\cdots\times]-\infty,t_d]) \\ &= P(X_1 \leq t_1,\dots,X_d \leq t_d). \\ &= \mathbf{E}(\mathbf{1}_{X_1\leq t_1,\dots,X_d\leq t_d}).\end{aligned}$$

La fonction de répartition F^X caractérise la loi de X puisque la fonction de répartition caratctérise une mesure de probabilité : si F^X est la fonction de répartition d'une mesure de probabilité μ, alors $X \sim \mu$.

Exemples 1.16. *(i) $F(t) := 1 - e^{-\lambda t}$ est fonction de répartition de la loi exponentielle $\mathcal{E}xp(\lambda)$.*
(ii) $F := \mathbf{1}_{[x,\infty[}$ est la fonction de répartition de la masse de Dirac δ_x, $x \in \mathbb{R}$, donc d'une var $X = a$ ps.
(iii) $F(t) := \sum_{0\leq k\leq n} C_n^k p^k (1-p)^{n-k}\mathbf{1}_{[k,\infty[}(t)$ est fonction de répartition de la loi binomiale $\mathcal{B}(n,p)$.
(iii) $F(t) := t$ est la fonction de répartition de la loi uniforme $\mathcal{U}_{[0,1]}$.

Définition 1.17. *Pour un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$, on appelle fonction génératrice de moments, on note M^X, la transformée de Laplace de la loi P^X si celle-ci est définie. Plus précisément, en supposant qu'il existe $r > 0$ tel que $e^{r|X|} \in \mathcal{L}^1$ et en notant $r^* \in (0,\infty]$ le suppremum de ces nombres, on définit*

$$M^X(z) := \int_{\mathbb{R}^d} e^{z\cdot x}\,dP^X(x) = \mathbf{E}(e^{z\cdot X}), \quad \forall\, z \in \mathbb{C}^d,\ |z| < r^*.$$

Lemme 1.18. *Lorsque celle-ci est définie, la fonction génératrice de moments M^X caractérise la loi de X puisque la fonction génératrice de moments caratctérise une mesure de probabilité : si M^X est la fonction génératrice de moments d'une mesure de probabilité μ, alors $X \sim \mu$. De plus, en dimension $d = 1$ (pour simpifier), on a*

$$M^X(z) = \sum_{n \geq 0} \frac{z^n}{n!} \mathbf{E}(X^n), \quad \forall\, z \in \mathbb{C},\, |z| < r^*.$$

Afin de voir que la fonction génératrice de moments M^X caractérise la loi de X, on peut remarque que $\frac{d^n}{dz^n} M^X(0) = \mathbf{E}(X^n)$, et que ceux-ci caractérisent la loi de X d'après le corollaire VI.2.9. du théorème de Weierstrass.

Définition 1.19. *Pour une va X à valeurs dans $\mathbb{N}$, la fonction génératrice de X est la série entière*

$$G^X(t) := \mathbf{E}(t^X) = \sum_{k=0}^{\infty} t^k \mathbf{P}(X = k) = \sum_{k=0}^{\infty} t^k \mathbf{P}^X(k), \quad \forall\, t \in \mathbb{R},\, |t| \leq 1.$$

Exemples 1.20. *Pour la loi de Poisson $\mathscr{P}(\lambda)$, on calcule la fonction génératrice des moments $M_X(t) = \mathbf{E}(e^{tX}) = \exp(\lambda(e^t - 1))$ et fonction génératrice $G_X(t) = \mathbf{E}(t^X) = e^{\lambda(t-1)}$.*

Pour la loi exponentielle $\mathcal{E}xp(\lambda)$, on calcule la fonction de répartition $F_X = 1 - e^{-\lambda t}$ et la fonction génératrice des moments $M_X(t) = (1 - t/\lambda)^{-1}$.

2 Indépendance

Le concept probablement le plus central de la théorie des probabilités est celui d'indépendance, dont nous rappelons la définition la plus simple.

Définition 2.1. *Deux évenements A et B sont indépendants (on note parfois $A \perp\!\!\!\perp B$) si*

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

Une famille d'évenements $(A_i)_{i \in I}$ est mutuellement indépendante si pour tout $J \subset I$ fini

$$\mathbf{P}\Big(\bigcap_{j \in J} A_j \Big) = \prod_{j \in J} \mathbf{P}(A_j). \tag{1}$$

Remarque 2.2. *(i) L'indépendance de deux événements A et B traduit l'idée que la connaissance de la réalisation de l'événement B ne donne pas d'information sur la réalisation ou non de l'événement A. Lorsque $P(B) > 0$, on peut écrire de manière équivalente que la probabilité conditionnelle de A sachant B est égale à la probabilité de A, soit donc*

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

(ii) Pour l'indépendance d'une famille d'événements de cardinal $\sharp I \geq 3$, il ne suffit pas d'avoir l'indépendance deux à deux des événements, c'est-à-dire

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j).$$

pour toute paire (i, j) de I. Donnons un exemple et considérons deux lancers de pile ou face équilibrés de sorte que $\Omega := \{0, 1\}^2$, $\mathbf{P}(\{a, b\}) = 1/4$. Prenons les événements

$$\begin{aligned} A_1 &:= \{\text{pile au premier lancer}\}, \quad \mathbf{P}(A_1) = 1/2, \\ A_2 &:= \{\text{pile au second lancer}\}, \quad \mathbf{P}(A_2) = 1/2, \\ A_3 &:= \{\text{même résulat aux deux lancers}\}, \quad \mathbf{P}(A_3) = 1/2. \end{aligned}$$

On calcule

$$\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j) = 1/4, \quad \forall\, i,j \in \{1,2,3\},\ i \neq j,$$

et pourtant

$$\mathbf{P}(A_1 \cap A_1 \cap A_3) = 1/4 \neq 1/8 = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3).$$

(iii) L'indépendance définie par (1) *est équivalente à la condition*

$$\mathbf{P}\Big(\bigcap_{j\in J} B_j\Big) = \prod_{j\in J} \mathbf{P}(B_j), \tag{2}$$

pour tout $J \subset I$ *fini et tout* $B_j \in \sigma(A_j) := \{\emptyset, A_i, A_i^c, \Omega\}$*. Le sens* (2) *implique* (1) *est clair. Démontrons que* (1) *implique* (2) *lorsque* $J := \{1,\dots,n\}$ *pour simplifier les notations. Pour* $B_1 = A_1^c$ *et* $B_j = A_j$ *pour tout* $2 \leq j \leq n$*, on calcule*

$$\begin{aligned}
\mathbf{P}(A_1^c \cap A_2 \cap \dots \cap A_n) &= \mathbf{P}(A_2 \cap \dots A_n) - \mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\
&= \mathbf{P}(A_2)\dots\mathbf{P}(A_n) - \mathbf{P}(A_1)\dots\mathbf{P}(A_n) \\
&= (1 - \mathbf{P}(A_1))\mathbf{P}(A_2)\dots\mathbf{P}(A_n) = \mathbf{P}(A_1^c)\mathbf{P}(A_2)\dots\mathbf{P}(A_n).
\end{aligned}$$

Le cas $B_1 = \emptyset$ *et* $B_1 = \Omega$ *étant immédiat, nous avons démontré* (2) *pour tout* $B_j \in \sigma(A_j)$ *pour* $j \leq i$ *et* $B_j = A_j$ *pour* $j \geq i+1$ *dans le cas* $i = 1$*. En itérant le procédé, on montre successivement que* (2) *est vrai pour* $i = 2,\dots,n-1$*, ce qui est le résultat souhaité.*

Définition 2.3. *Une famille* $(\mathscr{C}_i)_{i\in I}$ *d'ensembles d'événements (de sous-tribus, d'algèbres, de* π*-systèmes, ...) est mutuellement indépendante si toute famille d'évènements* $(A_i)_{i\in I}$*,* $A_i \in \mathscr{C}_i$*, est mutuellement indépendante, soit donc*

$$\mathbf{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbf{P}(A_j), \quad \forall\, A_j \in \mathscr{C}_j,\ \forall\, J \subset I \text{ fini.}$$

Remarque 2.4. *La notion fondamentale en théorie des probabilités est la notion d'indépendance de sous-tribus. Les sous-tribus étant de gros ensembles d'événements, il est parfois commode d'avoir des conditions suffisantes pour l'indépendance de sous-tribus. C'est l'objet du résultat qui suit.*

Lemme 2.5. *- Deux algèbres* $\mathscr{C}_1$ *et* $\mathscr{C}_2$ *sont indépendantes si, et seulement si, les deux tribus* $\sigma(\mathscr{C}_1)$ *et* $\sigma(\mathscr{C}_2)$ *sont indépendantes.*
- De la même manière, n *algèbres* $\mathscr{C}_1,\dots,\mathscr{C}_n$ *sont indépendantes si, et seulement si, les* n *tribus* $\sigma(\mathscr{C}_1),\dots,\sigma(\mathscr{C}_n)$ *sont indépendantes, une famille d'algèbres* $(\mathscr{C}_i)_{i\in I}$ *est mutuellement indépendante si, et seulement si, la famille de tribus* $(\sigma(\mathscr{C}_i))_{i\in I}$ *est mutuellement indépendante.*
- On peut encore généraliser : une famille $(\mathscr{C}_i)_{i\in I}$ *de* π*-systèmes (chaque* $\mathscr{C}_i$ *est stable par intersections) est mutuellement indépendante si, et seulement si, la famille de tribus* $(\sigma(\mathscr{C}_i))_{i\in I}$ *est mutuellement indépendante. En particulier, on retrouve que les événements* A_i*,* $i \in I$*, sont indépendants si, et seulement si, les tribus* $\sigma(A_i)$*,* $i \in I$*, sont indépendantes.*

Preuve du Lemme 2.5. C'est un argument de classe monotone. On ne traite que le cas $n = 2$. On note $\mathscr{B}_i := \sigma(\mathscr{C}_i)$. Pour $C_2 \in \mathscr{C}_2$ fixé, on définit

$$\mathcal{M}_1 := \{B_1 \in \mathscr{B}_1,\ P(B_1 \cap C_2) = P(B_1)\,P(C_2)\}.$$

On a $\mathscr{C}_1 \subset \mathcal{M}_1 \subset \mathscr{B}_1$ par hypothèse et $\mathcal{M}_1$ est clairement une classe monotone. Par le lemme des classes monotones, on obtient $\mathcal{M}_1 = \mathscr{B}_1$. On fixe maintenant $B_1 \in \mathscr{B}_1$ et on définit

$$\mathcal{M}_2 := \{B_2 \in \mathscr{B}_2,\ P(B_1 \cap B_2) = P(B_1)\,P(B_2)\}.$$

De la même manière, on montre $\mathcal{M}_2 = \mathscr{B}_2$, ce qui suffit pour conclure. $\square$

Exercice 2.6. *Démontrer le dernier point du Lemme 2.5 en utilisant la variante du lemme de classe monotone sur les π-λ système.*

Corollaire 2.7. *Soit $(\mathscr{A}_i)_{i\in I}$ une famille de sous-tribus, $(J_\ell)_{\ell\in L}$ une partition de l'ensemble d'indices I et $\mathscr{B}_\ell := \sigma(\mathscr{A}_i,\, i \in J_\ell)$. Si la famille $(\mathscr{A}_i)_{i\in I}$ est mutuellement indépendante alors la famille $(\mathscr{B}_\ell)_{\ell\in L}$ est mutuellement indépendante.*

Preuve du Corollaire 2.7. Il suffit de traiter le cas fini : $L := \{1, \dots, n\}$, $J_\ell := \{1, \dots, m(\ell)\}$. Soit alors $\mathscr{C}_\ell := \{\cap_j A_j,\ A_j \in \mathscr{A}_j,\ j \in J_\ell\}$. C'est un π-système qui contient $\cup_j \mathscr{A}_j$. Par construction, la famille des $(\mathscr{C}_\ell)$ est mutuellement indépendante. D'après le Lemme 2.5 et puisque $\mathscr{B}_\ell = \sigma(\mathscr{C}_\ell)$, la famille $(\mathscr{B}_\ell)_{\ell\in L}$ est mutuellement indépendante. □

Définition 2.8. *Une famille de variables aléatoires $(X_i)_{i\in I}$ est mutuellement indépendante si la famille des tribus engendrées $(\sigma(X_i))_{i\in I}$ est mutuellement indépendante, soit donc*

$$\mathbf{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbf{P}(A_j), \quad \forall\, A_j \in \sigma(X_j), \quad \forall\, J \subset I \text{ fini.}$$

Pour deux variables aléatoires indépendantes X et Y, on notera parfois $X \perp\!\!\!\perp Y$.

Théorème 2.9. *Une famille finie de variables aléatoires $X = (X_j)_{j\in J}$, $J := \{1, \dots, n\}$, à valeurs dans $(E, \mathcal{B})$ est mutuellement indépendante si, et seulement si, l'une des conditions équivalentes suivantes est réalisée*

(1) pour tous les ensembles mesurables $B_j \in \mathcal{B}$,

$$\mathbf{P}(X_j \in B_j, \forall\, j \in J) = \prod_{j\in J} \mathbf{P}(X_j \in B_j);$$

(2) la loi conjointe des X_j, $j \in J$, est le produit des lois marginales, soit donc

$$\mathbf{P}^X = \mathbf{P}^{(X_1,\dots,X_n)} = \mathbf{P}^{X_1} \otimes \cdots \otimes \mathbf{P}^{X_n};$$

(3) pour toute famille $(\phi_j)_{j\in J}$ de fonctions boréliennes telles que $\phi_j \ge 0$ pour tout $j \in J$ ou $\phi_j(X_j)$ est intégrable pour tout $j \in J$, on a

$$\mathbf{E}\Big(\prod_{j\in J} \phi_j(X_j)\Big) = \prod_{j\in J} \mathbf{E}\big(\phi_j(X_j)\big).$$

Si les X_i sont réelles (ou à valeurs dans $\mathbb{R}^d$), ces conditions sont également équivalentes à

(4) la fonction caractéristique du vecteur aléatoire X est le produit (tensoriel) des fonctions caractéristiques des X_j :

$$\varphi_X(\xi) = \prod_{j=1}^{n} \varphi_{X_j}(\xi_j), \quad \forall\, \xi = (\xi_{j_1}, \dots, \xi_{j_n}).$$

En particulier, si $(X_1, \dots, X_n) \in \mathcal{L}^2$ est une famille de var indépendantes alors la matrice de covariance associée est diagonale (mais la réciproque est fausse en général!).

Preuve du Théorème 2.9. On procède en plusieurs étapes.

Etape 1. En rapelant que $\sigma(X_j) = \sigma(\{X_j \in B_j, \forall\, B_j \in \mathcal{B}\})$, soit donc $A_j \in \sigma(X_j)$ si, et seulement si, $A_j = \{X_j \in B_j\}$ pour un certain $B_j \in \mathcal{B}$, on a clairement que les (X_j) sont mutuellement indépendants si, et seulement si (1).

Etape 2. Par définition, on a

$$\mathbf{P}^X(B_1 \times \cdots \times B_n) = \mathbf{P}(X_j \in B_j, \forall\, j \in J)$$

et
$$\mathbf{P}^{X_1} \otimes \cdots \otimes \mathbf{P}^{X_n}(B_1 \times \cdots \times B_n) = \mathbf{P}^{X_1}(B_1) \dots \mathbf{P}^{X_n}(B_n) = \prod_{j \in J} \mathbf{P}(X_j \in B_j).$$

On en déduit que (1) est équivalent à l'identité (2) sur les pavés de E^J, et donc (1) est équivalent à (2) d'après le théorème d'unicité de la mesure (classe monotone).

Etape 3. L'identité (2) sur les pavés de E^J est équivalente à l'identité (3) pour des fonctions ϕ_j qui sont des fonctions caractéristiques d'ensemble, puisque
$$\mathbf{E}(\prod \mathbf{1}_{X_j \in B_j}) = \mathbf{E}(\mathbf{1}_{X_1 \in B_1, \dots, X_n \in B_n}) = \mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n)$$
d'une part, et
$$\prod \mathbf{E}(\mathbf{1}_{X_j \in B_j}) = \prod \mathbf{P}(X_j \in B_j)$$
d'une part. L'équivalence entre (2) et (3) s'en déduit immédiatement (en utilisant la construction de l'intégrale de Lebesgue qui permet de passer aux fonctions étagées, puis mesurables positives, puis enfin intégrables).

Etape 3 (alternative). Si (2) est vraie, grâce au théorème de Fubini, on peut écrire
$$\begin{aligned}
\mathbf{E}\Big(\prod_{j \in J} \phi_j(X_j)\Big) &= \int_{E^N} \prod_{j \in J} \phi_j(x_j)\, dP^{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\
&= \int_{E^N} \prod_{j \in J} \phi_j(x_j)\, dP^{X_1}(x_1) \dots dP^{X_n}(x_n) \\
&= \prod_{j \in J} \int_E \phi_j(x_j)\, dP^{X_j}(x_j) = \prod_{j \in J} \mathbf{E}(\phi_j(X_j)),
\end{aligned}$$
ce qui démontre (3). La preuve de l'implication inverse est similaire.

Etape 4. Lorsque $E = \mathbb{R}^d$, il est clair que (3) implique (4) en prenant $\phi_j(x_j) := e^{-i\xi_j \cdot x_j}$. L'implication réciproque provient de la densité de l'espace vectoriel engendré par les fonctions circulaires, résultat démontré au cours de la preuve du Théorème de Lévy. □

Remarque 2.10. *(a) En utilisant le Lemme 2.5, on voit que si $\mathscr{C}$ est une algèbre telle que $\sigma(\mathscr{C}) = \mathcal{B}$, alors (1) est équivalent à l'assertion : pour tous les ensembles $C_j \in \mathscr{C}$,*
$$\mathbf{P}(X_j \in C_j, \forall\, j \in J) = \prod_{j \in J} \mathbf{P}(X_j \in C_j);$$
(b) Lorsque l'on considère une famille (quelconque, par exemple une suite) de variables aléatoires $(X_i)_{i \in I}$, le résultat précédent est vrai en remplaçant partout $(X_1, \dots, X_n)$ par $X_J := (X_{j_1}, \dots, X_{j_n})$ pour toute famille finie $J := \{j_1, \dots, j_n\} \subset I$.

Lemme 2.11. *Soit X et Y deux variables aléatoires indépendantes. Alors*
$$P^{X+Y} = P^X * P^Y$$
et en particulier
$$\phi^{X+Y} = \phi^X \phi^Y.$$

Preuve du Lemme 2.11. On fixe $\phi : \mathbb{R} \to \mathbb{R}_+$ mesurable, et on calcule
$$\begin{aligned}
\int_{\mathbb{R}} \phi(z)\, P^{X+Y}(dz) &= \mathbf{E}(\phi(X+Y)) \\
&= \int_{\mathbb{R}^2} \phi(x+y)\, P^{(X,Y)}(dx, dy) \\
&= \int_{\mathbb{R}^2} \phi(x+y)\, P^X(dx) P^Y(dy). \\
&= \int_{\mathbb{R}} \phi(z)\, (P^X * P^Y)(dz).
\end{aligned}$$

d'après la définition de la mesure $P^X * P^Y$ présentée au chapitre précédent. $\square$

On observe que si $X_1 \perp\!\!\!\perp X_2$ sont deux var $\mathcal{L}^2$ de même loi alors (puisque dans ce calcul on peut supposer la loi centrée)

$$\mathrm{Var}(X_1 + X_2) = \mathbf{E}(X_1 + X_2)^2 = 2\mathrm{Var}X_1.$$

De même, si X_i est une suite de var $\mathcal{L}^2$ indépendantes et de même variance, alors

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\mathrm{Var}(X_1), \quad \bar{X}_n := \frac{S_n}{n}, \quad S_n := X_1 + \cdots + X_n.$$

On peut supposer les var centrées, et on calcule alors

$$\begin{aligned}
\mathrm{Var}(\bar{X}_n) &= \mathbf{E}((\bar{X}_n)^2) = \frac{1}{n^2}\mathbf{E}(S_n^2) = \frac{1}{n^2}\mathbf{E}\big((\sum_{i=1}^n X_i)(\sum_{j=1}^n X_j)\big) \\
&= \frac{1}{n^2}\sum_{i\neq j}\mathbf{E}(X_iX_j) + \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}(X_i^2) \\
&= \frac{1}{n^2}\sum_{i\neq j}\mathbf{E}(X_i)\mathbf{E}(X_j) + \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}(X_1^2) = \frac{1}{n}\mathbf{E}(X_1^2) = \frac{1}{n}\mathrm{Var}(X_1).
\end{aligned}$$

Exercice 2.12. *Soient $X \sim \mathscr{P}(\lambda)$ et $Y \sim \mathscr{P}(\mu)$ indépendantes, montrer que $X + Y \sim \mathscr{P}(\lambda + \mu)$. Soient $X_1 \sim \mathcal{N}(\mu, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu, \sigma_2^2)$ indépendantes, montrer que $X_1 + X_2 \sim \mathcal{N}(\mu, \sigma_1^2 + \sigma_2^2)$. (Ind. Pour le premier cas, preuve à la main, en calculant $P(X + Y = n)$, ou en passant par la fonction caractéristique.)*

3 Convergence en loi et applications

On retiendra du catalogue de résultats de convergence démontrés au chapitre 2, les implications suivantes :

- Si $X_n \to X$ dans $\mathcal{L}^1$, alors $X_n \to X$ en probabilité, ce qui signifie

$$\forall\, \varepsilon > 0, \quad \mathbf{P}(|X_n - X| > \varepsilon) \to 0, \quad n \to \infty.$$

- Si $X_n \to X$ en probabilité, alors $X_{n_k} \to X$ p.s. pour une sous-suite.

Définition 3.1. *On dit qu'une suite de variables aléatoires (X_n) converge en loi vers une variable aléatoire X, on note $X_n \Rightarrow X$, si $\mathbf{P}^{X_n} \rightharpoonup \mathbf{P}^X$ faiblement, soit donc également*

$$\mathbf{E}(\varphi(X_n)) \to \mathbf{E}(\varphi(X)), \quad \forall\, \varphi \in C_b(E).$$

Nous avons vu que lorsque $E = \mathbb{R}^d$, la famille des fonctions tests utilisée dans cette définition peut être modifiée. Au lieu de $\mathcal{F} = C_b(\mathbb{R}^d)$, on peut prendre $\mathcal{F} = C_0(\mathbb{R}^d)$, $\mathcal{F} = C_c(\mathbb{R}^d)$ ou en fait même $\mathcal{F} = \{\varphi;\, \varphi(x) = \exp(-ix\cdot\xi),\, \xi \in \mathbb{R}^d\}$ (version faible du Théorème de Lévy).

Lemme 3.2. *Si $X_n \to X$ p.s., en probabilité ou $\mathcal{L}^1$, alors $X_n \Rightarrow X$.*

Preuve du Lemme 3.2. Si $X_n \to X$ p.s. alors pour tout $\varphi \in C_b$, on a $\varphi(X_n) \to \varphi(X)$ p.s. et $|\varphi(X_n)| \le \|\varphi\|_\infty \in \mathcal{L}^1$ de sorte que

$$\mathbf{E}(\varphi(X_n)) \to \mathbf{E}(\varphi(X)),$$

d'après le Théorème de convergence dominée. Cela signifie bien $X_n \Rightarrow X$.

Si $X_n \to X$ en probabilité alors pour une sous-suite (X_{n_k}) on a $X_{n_k} \to X$ ps et donc $X_{n_k} \Rightarrow X$ d'après l'étape précédente. Par unicité de la limite, c'est toute la suite qui converge.

Si $X_n \to X$ au sens $\mathcal{L}^1$ alors $X_n \to X$ en probabilité, et on utilise la deuxième étape. $\square$

Lemme 3.3. *Si $X_n \Rightarrow a$, $a \in \mathbb{R}$, alors $X_n \to a$ en probabilité.*

Il convient d'observer que si $X = a$ p.s., alors pour toute fonction mesurable $\varphi : E \to \mathbb{R}_+$, on a

$$\varphi(a) = \mathbf{E}(\varphi(a)) = \mathbf{E}(\varphi(X)) = \int_E \varphi(x) d\mathbf{P}^X(x),$$

ce qui signifie que $\mathbf{P}^X = \delta_a$.
Preuve du Lemme 3.3. Pour $\varepsilon > 0$, on définit la fonction $\varphi(x) = |x - a| \wedge \varepsilon$. On estime

$$\begin{aligned} \mathbf{P}\{|X_n - a| \geq \varepsilon\} &= \mathbf{P}\{|X_n - a| \wedge \varepsilon \geq \varepsilon\} \\ &\leq \frac{1}{\varepsilon}\mathbf{E}(|X_n - a| \wedge \varepsilon) \to \frac{1}{\varepsilon}\mathbf{E}(|a - a| \wedge \varepsilon) = 0, \end{aligned}$$

puisque $\varphi \in C_b(\mathbb{R})$. □

Théorème 3.4 (Loi faible des grands nombres)**.** *Soit (X_n) une suite de var iid $\mathcal{L}^1$. On a*

$$\bar{X}_n := \frac{S_n}{n} \;\Rightarrow\; \mathbf{E}(X_1).$$

Preuve du Théorème 3.4. Preuve dans le cas $\mathcal{L}^2$. Pour une suite (X_n) de var iid et $\mathcal{L}^2$, on a vu au paragraphe précédent que

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\mathrm{Var}(X_1).$$

Pour tout $\delta > 0$ et $n \in \mathbb{N}$, on a donc (ce n'est rien d'autre que l'inégalité de BT)

$$\begin{aligned} \mathbf{P}(\{|\bar{X}_n - \mathbf{E}(X_1)| > \delta\}) &= \mathbf{P}(\{|\bar{X}_n - \mathbf{E}(\bar{X}_n)|^2 > \delta^2\}) \\ &\leq \frac{1}{\delta^2}\mathbf{E}((\bar{X}_n - \mathbf{E}(\bar{X}_n))^2) \\ &= \frac{1}{\delta^2}\frac{1}{n}\mathrm{Var}(X_1). \end{aligned}$$

Cela prouve $\bar{X}_n \to \mathbf{E}(X_1)$ en probabilité, et donc en loi d'après le Lemme 3.2.
Preuve dans le cas général $\mathcal{L}^1$. Quitte à remplacer X_i par $X_i - \mathbf{E}(X_i)$, on peut supposer la moyenne nulle (les var centrées), soit donc $\mathbf{E}(X_1) = 0$. On calcule

$$\phi_{\bar{X}_n}(\xi) = \mathbf{E}(e^{i\bar{X}_n\xi}) = \prod_{i=1}^n \mathbf{E}(e^{iX_i\xi/n}) = \phi_{X_1}(\xi/n)^n.$$

La fonction ϕ_{X_1} est de classe C^1 puisque $X_1 \in \mathcal{L}^1$, et on a donc

$$\begin{aligned} \phi_{X_1}(s) &= \phi_{X_1}(0) + s\phi'_{X_1}(0) + s\varepsilon(s) \\ &= 1 + si\mathbf{E}(X_1) + s\varepsilon(s) = 1 + s\varepsilon(s), \end{aligned}$$

avec $\varepsilon(s) \to 0$ lorsque $s \to 0$. Pour $z \in \mathbb{C}$, on majore facilement

$$\big|(1+z)^n - 1\big| = \Big|\sum_{k=1}^n C_n^k z^k\Big| \leq \sum_{k=1}^n C_n^k |z|^k = e^{n\log(1+|z|)} - 1 \leq e^{n|z|} - 1.$$

On en déduit

$$|\phi_{\bar{X}_n}(\xi) - 1| \leq e^{|\xi|\varepsilon(|\xi|/n)} - 1 \underset{n\to\infty}{\longrightarrow} 0, \quad \forall\, \xi \in \mathbb{R}.$$

Cela prouve que $\mathcal{F}(P^{\bar{X}_n})(\xi) \to \mathcal{F}(\delta_0)(\xi)$ ponctuellement, donc $P^{\bar{X}_n} \to \delta_0$ d'après le Théorème de Lévy, et donc $\bar{X}_n \Rightarrow 0$ en loi. □

Théorème 3.5 (limite centrale/central limit). *Soit (X_n) une suite de var iid $\mathcal{L}^2$ et soit $\sigma^2 :=$ $\mathrm{Var}(X_1)$. On a*

$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n - n\mathbf{E}[X_1]) \;\Rightarrow\; \mathcal{N}(0, \sigma^2).$$

Preuve du Théorème 3.5. Quite à remplacer X_n par $X_n - \mathbf{E}[X_n]$, on peut supposer $\mathbf{E}[X_1] = 0$, et on pose

$$Z_n := \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

On calcule

$$\phi_{Z_n}(\xi) = \mathbf{E}(e^{iZ_n\xi}) = \prod_{i=1}^{n} \mathbf{E}(e^{iX_i\xi/\sqrt{n}}) = \phi_{X_1}(\xi/\sqrt{n})^n.$$

L'hypothèse $X_1 \in \mathcal{L}^2$ implique $(1 + |x|^2)\, e^{ix\cdot\xi} \in \mathcal{L}^1(\mathbf{P}^{X_1})$, on peut ainsi appliquer (deux fois) le théorème de dérivation sous le signe somme et on obtient que la fonction ϕ_{X_1} est de classe C^2. Un développement limité à l'ordre 2 donne alors

$$\begin{aligned}
\phi_{X_1}(s) &= \phi_{X_1}(0) + s\phi'_{X_1}(0) + \frac{s^2}{2}\phi''_{X_1}(0) + s^2\varepsilon(s) \\
&= 1 + si\mathbf{E}(X_1) - \frac{s^2}{2}\mathbf{E}(X_1^2) + s^2\varepsilon(s) \\
&= 1 - \frac{1}{2}\sigma^2 s^2 + s^2\varepsilon(s),
\end{aligned}$$

avec $\varepsilon(s) \to 0$ lorsque $s \to 0$. Pour $\xi \in \mathbb{R}$ fixé, on en déduit

$$\begin{aligned}
\big[\phi_{X_1}(\frac{\xi}{\sqrt{n}})\big]^n &= \Big(1 - \frac{1}{2}\sigma^2\frac{\xi^2}{n} + \frac{1}{n}\varepsilon_\xi(n)\Big)^n \\
&= \exp\Big(n\ln\big(1 - \frac{1}{2}\sigma^2\frac{\xi^2}{n} + \frac{1}{n}\varepsilon_\xi(n)\big)\Big) \\
&= \exp\Big(-\frac{1}{2}\sigma^2\xi^2 + \tilde{\varepsilon}_\xi(n)\Big),
\end{aligned}$$

avec $\varepsilon_\xi(n), \tilde{\varepsilon}_\xi(n) \to 0$ lorsque $n \to \infty$. Finalement, on a

$$\forall\, \xi \in \mathbb{R}, \quad \phi_{Z_n}(\xi) \underset{n\to\infty}{\longrightarrow} \exp\big(-\frac{1}{2}\sigma^2\xi^2\big) =: \varphi(\xi). \tag{3}$$

Une façon de justifier ce calcul avec des nombres complexes est la suivante. Pour $\alpha \in [0, 1[$ et $\beta \in \mathbb{C}$, $\alpha + |\beta| \le 1$, on écrit

$$\begin{aligned}
\big|(1 - \alpha + \beta)^n - (1 - \alpha)^n\big| &= \Big|\sum_{k=0}^{n-1} C_n^k (1 - \alpha)^k \beta^{n-k}\Big| \\
&\le \sum_{k=0}^{n-1} C_n^k (1 - \alpha)^k |\beta|^{n-k} \\
&= (1 - \alpha + |\beta|)^n - (1 - \alpha)^n,
\end{aligned}$$

et on applique le raisonnement précédent avec $\alpha := \frac{1}{2}\sigma^2\frac{\xi^2}{n}$ et $\beta := \frac{\varepsilon_\xi(n)}{n}$. Cela donne

$$\Big|\big[\phi_{X_1}(\frac{\xi}{\sqrt{n}})\big]^n - \Big(1 - \frac{1}{2}\sigma^2\frac{\xi^2}{n}\Big)^n\Big| \le \Big(1 - \frac{1}{2}\sigma^2\frac{\xi^2}{n} + \frac{1}{n}|\varepsilon_\xi(n)|\Big)^n - \Big(1 - \frac{1}{2}\sigma^2\frac{\xi^2}{n}\Big)^n.$$

Comme les trois derniers termes convergent vers $\varphi(\xi)$, on en déduit (3). En observant que φ est la fonction caractéristique d'une loi $\mathcal{N}(0, \sigma^2)$, on conclut grâce au théorème de Lévy. □

4 Compléments sur l'indépendance et l'asymptotique

4.1 Version $\mathcal{L}^4$ de la loi forte des grands nombres

Théorème 4.1. *Soit (X_n) une suite de var iid $\mathcal{L}^4$. Alors*

$$Y_n := \frac{1}{n}(X_1 + \cdots + X_n) \to \mathbf{E}(X_1) \quad p.s.$$

Pour aléger les notations on suppose $\mathbf{E}(X_1) = 0$. On calcule

$$\begin{aligned}
\mathbf{E}(Y_n^4) &= \frac{1}{n^4}\mathbf{E}\Big(\sum_{i_1,\dots,i_4} X_{i_1}X_{i_2}X_{i_3}X_{i_4}\Big) \\
&= \frac{1}{n^4}\Big(\sum_{i_1=I_2=i_3=i_4} + 3\sum_{i_1=i_2\neq i_3=i_4}\Big)\mathbf{E}X_{i_1}X_{i_2}X_{i_3}X_{i_4} \\
&= \frac{1}{n^4}\big(n^2\mathbf{E}X_1^4 + 3n(n-1)(\mathbf{E}X_1^2)^2\big) \le \frac{C}{n^2}.
\end{aligned}$$

On en déduit

$$\mathbf{E}\Big(\sum_n Y_n^4\Big) = \sum_n \mathbf{E}(Y_n^4) \le \sum_n \frac{C}{n^2} < \infty,$$

puis donc

$$\sum_n Y_n^4 < \infty \quad \text{p.s.}$$

On conclut que $Y_n^4 \to 0$ p.s. □

En particulier, si (A_n) est une suite d'évenements indépendants de même probabilité, alors

$$\frac{1}{n}(\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}) \to \mathbf{P}(A_1) \quad \text{p.s.}$$

4.2 Lemme de Borel-Cantelli

Pour une suite d'évenements (A_n), on définit

$$\limsup A_n := \bigcap_{n\ge 1}\bigcup_{k\ge n} A_k, \quad \liminf A_n := \bigcup_{n\ge 1}\bigcap_{k\ge n} A_k.$$

Lemme 4.2 (de Borel-Cantelli)**.** *Soit (A_n) une suite d'événements.*

(i) Si $\sum_{n\ge 1}\mathbf{P}(A_n) < \infty$, alors

$$\mathbf{P}(\limsup A_n) = 0,$$
soit donc, p.s. $\{n \in \mathbb{N};\ \omega \in A_n\}$ est fini.

(ii) Si $\sum_{n\ge 1}\mathbf{P}(A_n) = \infty$ et les A_n sont indépendants, alors

$$\mathbf{P}(\limsup A_n) = 1,$$
soit donc p.s. $\{n \in \mathbb{N};\ \omega \in A_n\}$ est infini.

Preuve du Lemme 4.2. (i) On écrit

$$\mathbf{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty}\mathbf{P}\Big(\bigcup_{k\ge n} A_k\Big) \le \lim_{n\to\infty}\sum_{k\ge n}\mathbf{P}(A_k) = 0,$$

où on utilise la propriété de limite monotone à la suite décroissante $\big(\bigcup_{k\ge n} A_k\big)$ dans la première égalité et on utilise la σ-sous-additivité pour avoir l'inégalité.

(ii) Pour tout $n \geq 1$, on a

$$\begin{aligned}\mathbf{P}\big(\bigcap_{k\geq n} A_k^c\big) &= \lim_{N\to\infty} \mathbf{P}\big(\bigcap_{k=n}^{N} A_k^c\big) = \lim_{N\to\infty} \prod_{k=n}^{N}(1-\mathbf{P}(A_k)) \\ &= e^{\sum_{k=n}^{\infty} \ln(1-\mathbf{P}(A_k))} \leq e^{-\sum_{k=n}^{\infty} \mathbf{P}(A_k)} = 0.\end{aligned}$$

D'où également $\mathbf{P}(\liminf A_n^c) = 0$. Et en passant au complémentaire $\mathbf{P}(\limsup A_n) = 1$. □

4.3 Théorème du 0-1 de Kolmogorov

Théorème 4.3 (Tout ou rien)**.** *Soit (X_n) une suite de var indépendantes. On définit les tribus $\mathcal{F}_n$ et $\mathcal{F}_\infty$ par*

$$\mathcal{F}_n := \sigma(X_k;\ k \geq n), \quad \mathcal{F}_\infty := \bigcap_{n\geq 1} \mathcal{F}_n.$$

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

Preuve du Théorème 4.3. Par définition, pour tout $n > m$, on a

$$\mathcal{F}_n \perp\!\!\!\perp \mathcal{G}_m := \sigma(X_k;\ k \leq m).$$

On a donc également

$$\mathcal{F}_\infty \perp\!\!\!\perp \mathcal{G}_m.$$

Comme $\cup_{m=1}^{\infty}\mathcal{G}_m$ est stable par intersections finies et passages au complémentaire, cette famille de parties est une algèbre, et d'après le Lemme VII-2.5, on a

$$\mathcal{F}_\infty \perp\!\!\!\perp \sigma(\mathcal{G}_m,\, m \geq 1) = \mathcal{F}_1 \supset \mathcal{F}_\infty.$$

Pour tout $B \in \mathcal{F}_\infty$, on a donc $\mathbf{P}(B) = \mathbf{P}(B \cap B) = \mathbf{P}(B)\mathbf{P}(B)$, et donc $\mathbf{P}(B) = 0$ ou 1. □

4.4 Marche aléatoire

Théorème 4.4. *Soit (X_n) une suite de var indépendantes de même loi définie par $\mathbf{P}(X_n = 1) = \mathbf{P}(X_n = -1) = 1/2$. Pour tout $n \geq 1$, on pose*

$$S_n := X_1 + \cdots + X_n.$$

Alors

$$p.s. \qquad \sup_{n\geq 1} S_n = +\infty \quad et \quad \inf_{n\geq 1} S_n = -\infty.$$

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

La preuve du Théorème 4.3 est présentée sous forme d'exercice.

1. Pour deux entiers $p \geq 1$ et $k > 2p$ fixés, et pour tout $j \geq 0$, on définit

   $$A_j := \{X_{jk+1} = X_{jk+2} = \cdots = X_{jk+k} = 1\}.$$

   Calculer $\mathbf{P}(A_j)$ et montrer que les (A_j) sont indépendants. A l'aide du Théorème de Borel-Cantelli, en déduire que

   $$\mathbf{P}\Big(\bigcup_{j=0}^{\infty} A_j\Big) = 1,$$

   puis que

   $$\mathbf{P}\Big(\big\{\inf S_n \leq -p\big\} \cup \big\{\sup_n S_n \geq p\big\}\Big) = 1.$$

2. On définit $\mathcal{B}_n := \sigma(X_\ell,\, \ell \geq n\}$, puis la tribu asymptotique $\mathcal{B}_\infty := \lim \mathcal{B}_n$. Montrer que

$$\mathbf{P}\big(\{\inf_n S_n = -\infty\}\big) + \mathbf{P}\big(\{\sup_n S_n = +\infty\}\big) \geq 1.$$

(Ind. On pourra montrer que

$$\mathbf{P}\big(\{-\infty < \inf_n S_n\} \cap \{\sup_n S_n < +\infty\}\big) \leq \sum_{p=1}^{\infty} \mathbf{P}\big(\{\forall\, n,\ S_n \in [-p,p]\}\big)).$$

Montrer que S_n a même loi que $-S_n$ et en déduire que

$$\mathbf{P}\big(\{\inf_n S_n = -\infty\}\big) = \mathbf{P}\big(\{\sup_n S_n = +\infty\}\big).$$

En observant que pour tout $k \geq 1$, on a

$$\{\inf_n S_n = -\infty\} = \{\inf_n S_n - S_k = -\infty\} \in \mathcal{B}_k,$$

montrer que

$$\{\inf_n S_n = -\infty\} \in \mathcal{B}_\infty.$$

Conclure que

$$\mathbf{P}\big(\{\inf_n S_n = -\infty\}\big) = \mathbf{P}\big(\{\sup_n S_n = +\infty\}\big) = 1.$$

4.5 Version $\mathcal{L}^2$ de loi forte des grands nombres

Théorème 4.5. *Soit (X_n) une suite de var iid $\mathcal{L}^2$. Alors*

$$Y_n := \frac{1}{n}(X_1 + \cdots + X_n) \to \mathbf{E}(X_1) \quad p.s.$$

La preuve est laissée en execice (cf. examen 2018-2019, deuxième session (Juin 2019)).

4.6 Version $\mathcal{L}^1$ de loi forte des grands nombres

Théorème 4.6. *Soit (X_n) une suite de var iid $\mathcal{L}^1$. Alors*

$$Y_n := \frac{1}{n}(X_1 + \cdots + X_n) \to \mathbf{E}(X_1) \quad p.s.$$

La preuve est laissée en execice (cf. examen 2018-2019, première session (Janvier 2019)).

5 Lexique

Petit lexique de correspondances :

- une partie mesurable $A \in \mathcal{A}$ est un *événement* ;

- une fonction mesurable $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ est une *variable aléatoire* ;

- une propriété presque partout (p.p.) vraie devient une propriété *presque sûrement (p.s.)* vraie.

- la convergence en mesure des suites de fonctions devient la convergence en probabilité des suites de variables aléatoires.