Cours de Probabilités

expérimenté par Didier Piau

devant divers publics, dont des élèves de l'ÉNS de Lyon inscrits en magistère de mathématiques et applications et des étudiants des universités Claude Bernard et Joseph Fourier inscrits en maîtrise puis en master de mathématiques

Cours et exercices

Dernière révision substantielle : mars 2010

Table des matières

Chapitre 1

Outils probabilistes

Ce chapitre introduit les outils utilisés dans le reste du cours. Il est, dans sa plus grande partie, assez élémentaire. Certains passages sont des redites d'un cours d'Intégration.

Dans la section 1.1, le théorème lambda–pi 1.2 et son corollaire 1.3, dû à Dynkin, sont importants. On pourra omettre les théorèmes 1.4 de classe monotone et 1.5 de Carathéodory. Dans la section 1.2 consacrée à Radon–Nykodym, on peut choisir de tout omettre, ou bien se limiter aux définitions 1.11 et 1.13, à l'énoncé du théorème de Radon–Nykodym 1.12, ainsi qu'aux exemples 1.15. Dans la section 1.4, sur les probabilités produits, on peut se limiter à la définition 1.22 et à l'énoncé du résultat dans le cas indépendant (théorème 1.23). La section 1.5 est incontournable et distribuée à la classe. La section 1.7 sur les diverses notions de convergence de suites de variables aléatoires est plus difficile que le reste dans sa partie 1.7 sur l'uniforme intégrabilité. La première chose à omettre est la preuve du théorème 1.32 de Vitali. L'exercice 1.33 est aussi important que la définition 1.31. Enfin, le tableau de la section 1.7 est à savoir par cœur.

Avant de commencer

La topologie étudie les fonctions continues : une fonction f est continue si et seulement si l'image réciproque $f^{-1}(O)$ de tout ouvert O est un ouvert. La théorie de la mesure étudie les fonctions mesurables : une fonction f est mesurable si et seulement si l'image réciproque $f^{-1}(B)$ de tout ensemble mesurable B est un ensemble mesurable.

En topologie, la notion d'ensemble ouvert repose sur le fait que *toute* réunion d'ouverts est un ouvert et qu'une intersection *finie* d'ouverts est un ouvert. En théorie de la mesure, la notion d'ensemble mesurable repose, entre autres propriétés, sur le fait que toute réunion *dénombrable* d'ensembles mesurables est un ensemble mesurable et que toute intersection *dénombrable* d'ensembles mesurables est aussi un ensemble mesurable. Si l'on n'impose pas que *seules les opérations dénombrables sont licites*, la théorie de la mesure devient contradictoire.

Nous développons ces deux idées en montrant que la théorie de la mesure est un bon cadre pour les probabilistes.

La marche au hasard sur le réseau carré : retour en l'origine

Un marcheur sur $\mathbb{Z}^2$ se déplace de sommet en sommet en allant d'un sommet en un sommet voisin (les quatre sommets voisins de (x, y) sont $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$ et $(x, y - 1)$) choisi au hasard (et la probabilité de choisir chacun des quatre voisins possibles est $1/4$) en faisant des pas successifs indépendants (et un pas donné ne dépend pas des déplacements précédents). Le modèle classique pour un tel « marcheur » est celui d'un ivrogne qui titube dans les rues de Manhattan (ou des Brotteaux).

Notons τ le premier instant (aléatoire) où il repasse par son point de départ, si cet instant existe, et posons $\tau = +\infty$ si le marcheur ne repasse jamais par son point de départ. Si on souhaite évaluer $\mathbb{P}(\tau = 2n)$ pour $n \geqslant 1$ fixé, il suffit de compter des chemins de longueur $2n$ et de diviser par le nombre total de ces chemins, qui vaut 4^{2n} : c'est de la combinatoire finie. Par contre, pour évaluer une quantité comme

$$\mathbb{P}(\tau < +\infty),$$

nécessité de stabilité dénombrable : c'est l'idée de Lebesgue. Par contre, permettre « plus » est dangereux, comme le montre le paradoxe suivant.

Un paradoxe célèbre

Choisissons un point au hasard sur la sphère unité S^2 de $\mathbb{R}^3$. La probabilité que ce point soit dans un sous-ensemble A de S^2 est l'aire de A divisée par l'aire totale, 4π. Tout est normal. Mais Banach et Tarski ont montré, en utilisant l'axiome du choix, qu'il existe un sous-ensemble A de S^2 tel que S^2 est la réunion disjointe

$$S^2 = A \cup \varrho(A) \cup \varrho'(A),$$

où ϱ et ϱ' sont deux rotations. Soit. Donc l'aire de A est $4\pi/3$. Malheureusement, pour le même sous-ensemble A de S^2, on peut écrire S^2 comme la réunion disjointe

$$S^2 = A \cup \varsigma(A) \cup \varsigma'(A) \cup \varsigma''(A),$$

où ς, ς' et ς'' sont trois rotations. Donc l'aire de A est aussi $4\pi/4$. D'ailleurs S^2 est aussi la réunion de n copies de A pour tout $n \geqslant 3$ (mais pas pour $n = 2$) et même d'un nombre dénombrable de copies de A. La morale est que A n'est pas mesurable : il est si compliqué qu'on ne peut pas définir son aire (sa mesure).

Le même phénomène en dimension 1

Il existe une partition du cercle $S^1 = \mathbb{R}/\mathbb{Z}$ par des ensembles disjoints $A(q)$, $q \in \mathbb{Q}$, tels que chaque ensemble $A(q)$ est un translaté de $A(0)$. Donc, $1 = +\infty \times \text{mesure}(A(0))$, ce qui est impossible. La construction : sur S^1, on considère la relation d'équivalence $\sim$ définie par « $x \sim y$ si et seulement si $x - y \in \mathbb{Q}$ ». Soit A un sous-ensemble de S^1 possédant exactement un représentant de chaque classe d'équivalence (axiome du choix). Alors $A(q) = q + A$ convient.

1.1 Mesures de probabilité et classes monotones

L'objet de base de tout le cours est un espace de probabilité, c'est-à-dire un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ tel que :

- Ω est un ensemble quelconque ;
- $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une tribu, c'est-à-dire que $\mathcal{F}$ est non vide (ou contient Ω) et que $\mathcal{F}$ est stable par passage au complémentaire et par réunion dénombrable ;
- $\mathbb{P}$ est une probabilité, c'est-à-dire une mesure (positive) sur $\mathcal{F}$ de masse totale $\mathbb{P}(\Omega) = 1$.

On va dans un premier temps se concentrer sur la partie $(\Omega, \mathcal{F})$ du triplet, avant de faire intervenir la mesure $\mathbb{P}$.

Pour toute collection $\mathcal{C}$ de parties de Ω, on note $\sigma(\mathcal{C})$ la tribu engendrée par $\mathcal{C}$ c'est-à-dire l'intersection de toutes les tribus contenant $\mathcal{C}$. (Exercice : c'est une tribu.)

Si Ω est un espace topologique de topologie (= l'ensemble des ouverts) $\mathcal{T}$, on appelle $\mathcal{B}(\Omega) = \sigma(\mathcal{T})$ la tribu de Borel et les éléments de $\mathcal{B}(\Omega)$ des boréliens.

Remarque 1.1 La tribu $\mathcal{B}(\mathbb{R})$, tribu borélienne de $\mathbb{R}$, est engendrée par les ouverts de $\mathbb{R}$. Les éléments de $\mathcal{B}(\mathbb{R})$ peuvent être compliqués. Par contre, la famille

$$\pi(\mathbb{R}) = \{]-\infty, x]\,;\, x \in \mathbb{R}\}$$

est beaucoup plus manipulable et $\mathcal{B}(\mathbb{R})$ a le bon goût (exercice !) de vérifier la propriété fondamentale : $\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R}))$. On introduit ainsi les

Lambda-systèmes et pi–systèmes

Définition *Soit $\mathcal{C}$ une partie de $\mathcal{P}(\Omega)$. On dit que $\mathcal{C}$ est un* λ**-système** *si $\mathcal{C}$ contient Ω et si $\mathcal{C}$ est stable par différence stricte et par réunion croissante dénombrable. On dit que $\mathcal{C}$ est un* π**-système** *si $\mathcal{C}$ est stable par intersection finie.*

Exercice • λ-système + π-système = tribu.
• Si $\mathcal{C}$ est une collection de parties de Ω, on note $\lambda(\mathcal{C})$, resp. $\pi(\mathcal{C})$, le λ-système, resp. π-système, engendré par $\mathcal{C}$ c'est-à-dire le plus petit λ-système, resp. π-système, contenant $\mathcal{C}$. C'est l'intersection de tous les λ-systèmes, resp. π-systèmes, contenant $\mathcal{C}$. (Exercice : montrer que c'est un λ-système, resp. π-système.)

Exemple. Sur $S^1 = \mathbb{R}/\mathbb{Z}$, on considère la classe $\mathcal{C}$ formée des $[x, x + 1/2[$, $x \in S^1$. Trouver $\lambda(\mathcal{C})$ et $\pi(\mathcal{C})$. ■

Dans l'exemple de la remarque 1.1, on voit qu'une tribu peut être compliquée et un π-système être plus simple. Question : comment étudier des tribus en ne manipulant que des π-systèmes ? Réponse : grâce au

Théorème 1.2 (Théorème lambda-pi) *Un λ-système contenant un π-système contient aussi la tribu engendrée par ce π-système.*

Corollaire 1.3 (Dynkin) *Si $\mathbb{P}$ et $\mathbb{Q}$ sont deux probabilités sur $\sigma(\mathcal{C})$ qui coïncident sur le π-système $\mathcal{C}$, alors $\mathbb{P} = \mathbb{Q}$ (sur $\sigma(\mathcal{C})$ tout entier).*

Démonstration On montre que le λ-système engendré par un π-système est un π-système : c'est donc une tribu et il doit contenir la tribu engendrée par le π-système car celle-ci est la plus petite tribu contenant le π-système. *A fortiori*, tout λ-système contenant le π-système contient aussi le λ-système engendré par ce π-système donc il contient encore la tribu engendrée par le π-système.

Pour cela, considérons $\lambda(\mathcal{C})$ le λ-système engendré par $\mathcal{C}$ et, pour $A \in \lambda(\mathcal{C})$, notons

$$\mathcal{C}(A) = \{B \in \lambda(\mathcal{C})\,;\, A \cap B \in \lambda(\mathcal{C})\}.$$

Si A est dans $\mathcal{C}$, alors $\mathcal{C}(A)$ contient $\mathcal{C}$. De plus, $\mathcal{C}(A)$ est toujours un λ-système (preuve !) donc $\mathcal{C}(A)$ contient $\lambda(\mathcal{C})$ par minimalité de $\lambda(\mathcal{C})$.

En d'autres termes, on a montré que $A \cap B \in \lambda(\mathcal{C})$ pour tout $A \in \mathcal{C}$ et tout $B \in \lambda(\mathcal{C})$. Ainsi, pour tout $B \in \lambda(\mathcal{C})$, $\mathcal{C}(B)$ contient $\mathcal{C}$ donc (rebelote) il contient aussi $\lambda(\mathcal{C})$. Ceci termine la démonstration du fait que $\lambda(\mathcal{C})$ est un π-système. ■

Exemple Le π–système $\pi(\mathbb{R})$ engendre $\mathcal{B}(\mathbb{R})$ donc une probabilité $\mathbb{P}$ sur $\mathcal{B}(\mathbb{R})$ est uniquement déterminée par sa fonction de répartition F définie sur $\mathbb{R}$ par

$$F(x) = \mathbb{P}(]-\infty,\, x]).$$

Exercice 1.4 (Théorème de la classe monotone) Une collection $\mathcal{C} \subset \mathcal{P}(\Omega)$ est une classe monotone si $\mathcal{C}$ est stable par réunion croissante dénombrable et par intersection décroissante dénombrable.

Soit $\mathcal{C}$ une algèbre (mêmes axiomes qu'une tribu avec réunion finie au lieu de réunion dénombrable). Montrer que si une classe monotone contient $\mathcal{C}$, elle contient aussi la tribu engendrée $\sigma(\mathcal{C})$. (Remarque : classe monotone + algèbre = tribu.) ■

Une autre façon de construire une mesure est de partir d'une algèbre.

Théorème 1.5 (Carathéodory) *Soit $m_0 : \mathcal{F}_0 \to [0, +\infty]$ une application σ-additive définie sur une algèbre $\mathcal{F}_0$ d'un ensemble Ω. Alors, il existe une mesure m définie sur $\mathcal{F} = \sigma(\mathcal{F}_0)$ telle que $m = m_0$ sur $\mathcal{F}_0$. Si de plus $m_0(\Omega)$ est fini, alors cette extension est unique.*

Admis.

Exemple $\Omega = \mathbb{R}$ et m_0 mesure de longueur sur les réunions finies d'intervalles disjoints d'intervalles $]x, y]$, $x \leqslant y$, donnent la mesure de Borel sur $\mathcal{B}(\mathbb{R})$.

Événements

Des propriétés de base importantes. Soit $(\Omega, \mathcal{F}, m)$ un espace mesuré.

Lemme *Soit $A_n \in \mathcal{F}$. Si A_n tend en croissant ou en décroissant vers A, alors $A \in \mathcal{F}$. Si A_n tend en croissant vers A, $m(A_n)$ tend en croissant vers $m(A)$. Si A_n tend en décroissant vers A et si $m(A_1)$ est fini, alors $m(A_n)$ tend en décroissant vers $m(A)$.*

Définition 1.6 *Un ensemble N est m-négligeable s'il existe $A \in \mathcal{F}$ tel que $N \subset A$ et $m(A) = 0$.*

Une réunion dénombrable d'ensembles négligeables est négligeable.

Exemple 1.7 (Le jeu de Pile ou face) Pour modéliser la suite des résultats des jets successifs d'une pièce de monnaie, on peut choisir $\Omega = \{0, 1\}^{\mathbb{N}^*}$. Un élément ω de Ω est $\omega = (\omega_n)_{n \geqslant 1}$ avec $\omega_n = 0$ ou $\omega_n = 1$ selon que le lancer numéro n produit un « pile » ou un « face ». On veut pouvoir mesurer au moins le fait que « pile » sort au n–ième coup donc $\mathcal{F}$ contient tous les $\{\omega_n = 0\}$. Posons

$$\mathcal{F} = \sigma(\{\omega\,;\, \omega_n = r\} \,:\, n \geqslant 1,\, r \in \{0, 1\}).$$

Pour $n \geqslant 1$, posons

$$Z_n = n^{-1} \operatorname{Card}\{k \leqslant n\,;\, \omega_k = 1\}.$$

Une loi des grands nombres naïve ou des considérations de symétrie incitent à conjecturer que Z_n converge vers $1/2$ en un certain sens quand n devient grand. En effet, on démontrera plus loin :

$$\mathbb{P}(\lim_{n\to\infty} Z_n = 1/2) = 1.$$

Admettons ce résultat pour le moment et notons $\mathcal{A}$ l'ensemble des suites $\alpha = (\alpha(n))_n$ strictement croissantes d'entiers. Pour tout $\alpha \in \mathcal{A}$, on a également :

$$\mathbb{P}(\lim_{n\to\infty} Z_n^\alpha = 1/2) = 1,$$

avec $Z_n^\alpha = n^{-1}\,\mathrm{Card}\{k \leqslant n\,;\,\omega_{\alpha(k)} = 1\}$. En d'autres termes, $P(A_\alpha) = 1$ pour A_α ensemble de vérité de la loi des grands nombres pour la suite α : la loi des grands nombres est presque sûrement vraie pour une suite extraite α fixée. Pourtant,

$$\bigcap_{\alpha\in\mathcal{A}} A_\alpha = \emptyset.$$

En effet, soit la suite $(\omega_n)_n$ contient une infinité de 1 et on choisit pour α les indices n pour lesquels $\omega_n = 1$, soit elle ne contient qu'un nombre fini de 1 et on choisit $\alpha_n = n$. Dans les deux cas, p_n^α ne converge pas vers $1/2$: la loi des grands nombres n'est jamais vraie simultanément pour toutes les suites extraites α. ■

Exercice Définitions de $\limsup$ et $\liminf$ pour des ensembles. Des faits à connaître et à savoir démontrer :

- l'indicatrice de $\limsup A_n$ est la $\limsup$ des indicatrices des A_n.
- Si $A_n \in \mathcal{F}$, alors $\limsup A_n \in \mathcal{F}$.
- la partie triviale du lemme de Borel–Cantelli.

Exemples de mesures de probabilité

- Masse de Dirac : pour x dans Ω, $\delta_x(A) = 1$ pour tout ensemble mesurable A qui contient x et $\delta_x(A) = 0$ pour tout ensemble mesurable A qui ne contient pas x (la tribu est donc implicite mais on note toutes ces mesures de la même façon).
- Barycentre de probabilités.
- Soit f une fonction intégrable sur $\mathbb{R}^d$, positive, et d'intégrale $\ell(f) = 1$ par rapport à la mesure de Lebesgue ℓ. Alors $\mathbb{P}$ définie par $\mathbb{P}(A) = \ell(f\,\mathbf{1}_A)$ est la probabilité sur $\mathcal{B}(\mathbb{R})$ de densité f par rapport à la mesure de Lebesgue ℓ.
- Probabilité produit :
  On se donne deux espaces de probabilité $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ et on définit une tribu $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ sur $\Omega = \Omega_1 \times \Omega_2$ par $\mathcal{F} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ (c'est la tribu engendrée par les pavés). On veut construire une probabilité $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ sur $\mathcal{F}$. Pour $A \in \mathcal{F}$, notons

  $$A_{\omega_1} = \{\omega_2 \in \Omega_2\,;\,(\omega_1, \omega_2) \in A\}, \quad A^{\omega_2} = \{\omega_1 \in \Omega_1\,;\,(\omega_1, \omega_2) \in A\}.$$

  Alors $A_{\omega_1} \in \mathcal{F}_2$ pour tout $\omega_1 \in \Omega_1$, $A^{\omega_2} \in \mathcal{F}_1$ pour tout $\omega_2 \in \Omega_2$ et d'après le théorème de Fubini

  $$\int_{\Omega_1} \mathbb{P}_2(A_{\omega_1})\,\mathbb{P}_1(\mathrm{d}\omega_1) = \int_{\Omega_2} \mathbb{P}_1(A^{\omega_2})\,\mathbb{P}_2(\mathrm{d}\omega_2),$$

  donc on peut définir $\mathbb{P}(A)$ par cette quantité. Alors, $\mathbb{P}$ est la seule probabilité sur $\mathcal{F}$ qui coïncide avec le produit de $\mathbb{P}_1$ et $\mathbb{P}_2$ sur $\mathcal{F}_1 \times \mathcal{F}_2$ c'est-à-dire telle que l'on ait $\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1) \times \mathbb{P}_2(A_2)$ pour tous $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$.
- Soit Ω un ensemble fini ou dénombrable. Une probabilité $\mathbb{P}$ sur $\mathcal{P}(\Omega)$ est donnée par sa valeur sur les singletons. En effet,

  $$\mathbb{P} = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\,\delta_\omega.$$

- Probabilité uniforme sur un ensemble fini : on pose $\mathbb{P}(\{\omega\}) = 1/|\Omega|$ pour tout $\omega \in \Omega$ où $|\Omega|$ est le cardinal de Ω.

Exercice Il n'existe pas de probabilité uniforme sur $\mathbb{N}$. Une tentation naturelle est de vouloir considérer la collection $\mathcal{C}$ des parties A de $\mathbb{N}$ pour lesquelles la densité

$$d(A) = \lim_{n\to\infty} |A \cap [0, n]|/(n+1)$$

existe et de choisir $d(A)$ comme mesure de A si $A \in \mathcal{C}$. Alors, deux inconvénients : $d(\{n\}) = 0$ pour tout n et $d(\mathbb{N}) = 1$ donc d est additive mais non σ-additive ; et même avant ce problème « dénombrable », $\mathcal{C}$ n'est pas une algèbre.

Par exemple, on choisit $A = 2\mathbb{N} + 1$ et B l'ensemble des nombres impairs appartenant aux segments $[4^n, 2 \cdot 4^n]$ pour tout entier n et des nombres pairs appartenant aux segments $[2 \cdot 4^n, 4^{n+1}]$ pour tout entier n. Alors A et B admettent pour densité $d(A) = d(B) = 1/2$ donc $A \in \mathcal{C}$ et $B \in \mathcal{C}$ mais la suite des densités partielles de $A \cap B$ admet comme valeurs d'adhérence le segment $[1/3, 2/3]$ donc $A \cap B \notin \mathcal{C}$.

Pour la suite de ce sujet fascinant qu'est la théorie probabiliste des nombres, voir Elliott, *Probabilistic theory of numbers,* Springer.

1.2 Variables aléatoires

La définition

Contenu : une variable aléatoire est une fonction mesurable.

Soient $(\Omega, \mathcal{F})$ et $(E, \mathcal{E})$ deux ensembles mesurables et $f : \Omega \to E$ une fonction. Pour toute partie A de E, on note

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

Alors, f est mesurable pour les tribus $\mathcal{F}$ et $\mathcal{E}$ si et seulement si $f^{-1}(A) \in \mathcal{F}$ pour tout $A \in \mathcal{E}$. Dessin :

$$\begin{array}{c} \Omega \xrightarrow{f} E \\ \mathcal{F} \xleftarrow{f^{-1}} \mathcal{E} \end{array}$$

Quelques propriétés importantes :

- f^{-1} préserve les opérations ensemblistes.
- Une fonction continue est mesurable pour les tribus de Borel.
- Une fonction $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable si et seulement si $\{f \leqslant x\} \in \mathcal{F}$ pour tout $x \in \mathbb{R}$.
- La composition des fonctions préserve la mesurabilité : si $f : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ est $\mathcal{F}/\mathcal{E}$ mesurable et si $g : (E, \mathcal{E}) \to (G, \mathcal{G})$ est $\mathcal{E}/\mathcal{G}$ mesurable, alors $g \circ f$ est $\mathcal{F}/\mathcal{G}$ mesurable. Dessin :

$$\begin{array}{c} \Omega \xrightarrow{f} E \xrightarrow{g} G \\ \mathcal{F} \xleftarrow{f^{-1}} \mathcal{E} \xleftarrow{g^{-1}} \mathcal{G} \end{array}$$

- Les passages aux limites inférieures et aux limites supérieures préservent la mesurabilité.

Exercice Soit $f_n : \Omega \to \mathbb{R}$ une suite de fonctions $\mathcal{F}/\mathcal{B}(\mathbb{R})$ mesurables. Montrer que l'ensemble $C = \{\omega \in \Omega\,;\, (f_n(\omega))_n \text{ converge}\}$ est mesurable.

Définition *Soit* $(\Omega,\, \mathcal{F},\, \mathbb{P})$ *un espace de probabilité et* $(E, \mathcal{E})$ *un espace mesurable. Soit* $X : \Omega \to E$ *une fonction. La* **tribu engendrée par** X *est*

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

On dit que X *est une* **variable aléatoire à valeurs dans** E *si et seulement si* $\sigma(X) \subset \mathcal{F}$*, c'est-à-dire si et seulement si* X *est* $\mathcal{F}/\mathcal{E}$ *mesurable.*
La mesure image $X(\mathbb{P})$ *est la mesure sur* $\mathcal{E}$ *définie par* $X(\mathbb{P})(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{X \in B\})$ *pour* $B \in \mathcal{E}$ *est appelée* **loi de** X *et notée* $\mathbb{P}_X$ *ou* $X(\mathbb{P})$.

Dessin pour la loi de $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ mesurable pour $\mathcal{F}$ et $\mathcal{B}(\mathbb{R})$:

$$\begin{array}{ccccc} & & \Omega & \xrightarrow{X} & \mathbb{R} \\ [0,1] & \xleftarrow{P} & \mathcal{F} & \xleftarrow{X^{-1}} & \mathcal{B}(\mathbb{R}) \\ [0,1] & \xleftarrow{P} & \sigma(X) & \xleftarrow{X^{-1}} & \mathcal{B}(\mathbb{R}) \end{array}$$

Si X est à valeurs réelles, un π–système qui engendre $\sigma(X)$ est

$$\pi(X) = \{\{X \leqslant x\}\,;\, x \in \mathbb{R}\}.$$

Définition *Tribu engendrée par une collection de fonctions $\mathcal{X}$: c'est la plus petite tribu $\sigma(\mathcal{X})$ qui rend chaque $X \in \mathcal{X}$ mesurable.*

De façon évidente, $\sigma(\mathcal{X})$ est engendrée par les $\{X \in B\}$, $X \in \mathcal{X}$, $B \in \mathcal{E}$.

Exercice 1.8 (Dynkin) Soit $X : \Omega \to E$ une variable aléatoire et $Y : \Omega \to \mathbb{R}$ une fonction. Alors Y est $\sigma(X)/\mathcal{B}(\mathbb{R})$ mesurable si et seulement si il existe une fonction $f : E \to \mathbb{R}$ $\mathcal{E}/\mathcal{B}(\mathbb{R})$ mesurable telle que $Y = f \circ X$.

Preuve : on commence par $Y \geqslant 0$, Y simple, i.e. Y combinaison linéaire (finie) d'indicatrices d'éléments de $\sigma(X)$. Si Y est $\sum_i \alpha_i\,\mathbf{1}(A_i)$ avec $A_i = \{X \in B_i\}$, alors $f = \sum_i \alpha_i\,\mathbf{1}(B_i)$ convient. Ensuite, supremum de fonctions simples. Enfin, $Y = Y^+ - Y^-$. ∎

Exemple 1.9 (Le jeu de Pile ou face (suite)) Pour $n \geqslant 1$, posons $X_n(\omega) = \omega_n$. Alors X_n est une variable aléatoire, de même que $S_n = X_1 + \cdots + X_n$ et $Z_n = n^{-1}\,S_n$. Pour $z \in [0,1]$, on peut considérer

$$B_z = \{\omega\,;\, Z_n \to z\},$$

et énoncer la loi des grands nombres « naïve » : pour une pièce non truquée, $\mathbb{P}(B_{1/2}) = 1$. Exercice : montrer que $B_z \in \mathcal{F}$ pour tout z. ∎

Lois classiques

- Loi de Bernoulli : $b(p)$ (pile ou face). C'est la loi de X_1 dans l'exemple 1.9.
- Loi binômiale : $B(n,p)$ (pile ou face itéré n fois). C'est la loi de S_n dans l'exemple 1.9. Montrer que, pour tout entier $0 \leqslant k \leqslant n$,
$$\mathbb{P}(S_n = k) = \binom{n}{k}\,p^k\,(1-p)^{n-k}.$$
- Loi exponentielle : $\mathcal{E}(a)$. Désintégration de particules radioactives. Durée de vie V d'une ampoule :
$$\mathbb{P}(V > t + s \,|\, V > t) = \mathbb{P}(V > s).$$
- Loi de Poisson : $P(a)$. Nombre d'ampoules remplacées au temps t si chacune a une durée de vie en $\mathcal{E}(a/t)$. Preuve : pour $\omega : \mathbb{N}^* \to \mathbb{R}_+$, on pose $N_t(\omega) = k$ si $\omega = (\omega_n)_{n\geqslant 1}$ vérifie $\omega_1 + \cdots \omega_k \leqslant t < \omega_1 + \cdots \omega_{k+1}$ et on munit l'ensemble Ω des fonctions $\omega : \mathbb{N}^* \to \mathbb{R}_+$ de la loi produit des lois $\mathcal{E}(a/t)$.
- Loi uniforme sur $[0,1]$. Développement dyadique sur $[0,1]$. Nombres normaux.
- Loi de Cauchy : $\mathcal{C}(m,a)$, ordonnée du lieu d'atteinte de la droite $\{x = a\}$ avec $a \geqslant 0$ par le mouvement brownien plan issu de $(0,m)$, faire le dessin pour une marche au hasard sur $\mathbb{Z}^2$ et en déduire la loi de la somme de deux variables aléatoires de Cauchy indépendantes.
- Loi normale ou gaussienne : $\mathcal{N}(m,\sigma^2)$, loi des erreurs ; sablier de Pascal.
- Lois de statistiques classiques et quantiques : Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac.

Théorème de Radon-Nykodym et fonctions de répartition

On rappelle que la fonction de répartition F d'une probabilité P sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est définie par

$$F(t) = P(]-\infty, t]).$$

Il est facile de montrer qu'une telle fonction est càdlàg (continue à droite et admettant des limites à gauche), croissante, qu'elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$ et enfin que la mesure d'un singleton est donnée par

$$P(\{t\}) = F(t) - F(t-).$$

Réciproquement, rappelons qu'une fonction F continue à droite et croissante étant donnée, il existe une mesure unique m sur $\mathcal{B}(\mathbb{R})$ telle que $m(]t,s]) = F(s) - F(t)$ pour tous $-\infty < t < s < +\infty$. Si de plus F tend vers 0 en $-\infty$ et vers 1 en $+\infty$, alors il existe une probabilité unique P sur $\mathcal{B}(\mathbb{R})$ telle que $P(]-\infty, t]) = F(t)$.

Un problème important est de savoir si une fonction de répartition (= fonction càdlàg, croissante, de limites 1 et 0 en $\pm\infty$), correspond toujours à une v.a. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Exercice Toute loi sur $\mathbb{R}$ est réalisée. (Voir TD.)

Remarque 1.10 (Tribu engendrée et Tyché) On se donne un espace $(\Omega, \mathcal{F}, \mathbb{P})$. Des expériences sont réalisées, qui donnent les résultats X_i. À développer.

Définition 1.11 *Soit m et n deux mesures sur $(E, \mathcal{E})$. On dit que m est absolument continue par rapport à n, ce que l'on note $m \ll n$, si $n(B) = 0$ entraîne $m(B) = 0$ pour tout $B \in \mathcal{E}$.*

Théorème 1.12 (Radon-Nykodym) *Soit m et n deux mesures σ-finies. Alors $m \ll n$ si et seulement si il existe une fonction f mesurable positive telle que $m = f \cdot n$.*
Alors f est unique m-presque partout et n est finie si et seulement si $f \in L^1(m)$. On note $f = \mathrm{d}m/\mathrm{d}n$.

Définition 1.13 *m et n sont deux mesures étrangères l'une à l'autre sur $(E, \mathcal{E})$, ce que l'on écrit $m \perp n$, si il existe $B \in \mathcal{E}$ avec $m(B) = n(E \setminus B) = 0$.*

Théorème 1.14 (Décomposition d'une mesure) *Si m et n sont deux mesures σ–finies sur un espace mesurable $(E, \mathcal{E})$, il existe un unique couple (m_a, m_e) de mesures tel que $m = m_a + m_e$, $m_a \ll n$ et $m_e \perp n$.*
Soit P une mesure de probabilité sur $\mathcal{B}(\mathbb{R})$ et ℓ la mesure de Lebesgue. Il existe un unique triplet (P_a, P_e, P_d) de mesures tel que $P = P_a + P_e + P_d$, $P_a \ll \ell$, P_d discrète (donc $P_d \perp \ell$), $P_e \perp \ell$ et P_e diffuse.

Voir Rudin pour la démonstration des théorèmes 1.12 et 1.14. Rappelons que la propriété d'absolue continuité de P par rapport à ℓ se lit sur la fonction de répartition F de P par l'absolue continuité de F c'est-à-dire

$$\forall \epsilon > 0,\ \exists \alpha > 0,\ \forall [x_n, y_n[\text{ disjoints}, \quad \sum_n y_n - x_n \leqslant \alpha \Longrightarrow \sum_n |F(y_n) - F(x_n)| \leqslant \epsilon,$$

et que l'étape–clé dans la preuve du théorème 1.14 consiste à montrer que si m est finie et si n est σ–finie, alors le fait que $m \ll n$ est équivalent à la propriété suivante :

$$\forall \epsilon > 0,\ \exists \alpha > 0,\ \forall A \in \mathcal{F}, \quad n(A) \leqslant \alpha \Longrightarrow m(A) \leqslant \epsilon.$$

(Le résultat est faux si m n'est pas finie, par exemple $m(\mathrm{d}x) = x^{-1}\,\mathrm{d}x$ et $n(\mathrm{d}x) = \mathrm{d}x$ sur $\mathbb{R}^+$ muni de ses boréliens.)

Exemple 1.15 Appliquons enfin ce qui précède au cas des mesures boréliennes sur $\mathbb{R}$ c'est-à-dire des mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. **[Dessins.]**

- P discrète si et seulement si F en escaliers.
- P diffuse si et seulement si F continue.
- P densitable si et seulement si F est continue et dérivable presque partout de dérivée f (théorème de dérivation de Lebesgue, Rudin 8.6 et 8.8).

$\blacksquare$

1.3 Indépendance

Digression empirique : soit A une partie mesurable de Ω, de mesure $P(A)$ non nulle ; évaluer la mesure de B « sachant A » revient à se placer sur A au lieu de Ω. On remplace donc P par P restreinte à A et renormalisée d'où

$$P(B|A) = c\,P_A(B) = P(A \cap B)/P(A).$$

[Dessin.]
On remarque que $P(\cdot|A)$ est une mesure de probabilité (à suivre).
Affirmer que A et B sont indépendants, c'est dire que savoir que l'on est sur A ne modifie pas les chances de B de se produire (exemple : ? ?). Mais alors, le même fait est vrai pour A ou A^c et B ou B^c (exercice). D'où la

Définition *Deux événements A et B sont* **indépendants** *si*

$$P(A' \cap B') = P(A')\,P(B')$$

pour tout $A' \in \sigma(\{A\})$, $B' \in \sigma(\{B\})$. De même, n événements A_k, $1 \leqslant k \leqslant n$ sont **indépendants** *si $P(A'_1 \cap \cdots \cap A'_n) = P(A'_1) \cdots P(A'_n)$ pour tout $A'_k \in \sigma(\{A_k\})$.*

On voit facilement que A et B sont indépendants si et seulement si $P(A \cap B) = P(A)\,P(B)$. De même

Proposition *Les A_k, $1 \leqslant k \leqslant n$, sont indépendants si et seulement si, pour toute partie I de $\{1, \ldots, n\}$, on a :*

$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

Démonstration Si $A' \in \sigma(\{A\})$, alors $\mathbf{1}_{A'} = \alpha \mathbf{1}_A + \beta \mathbf{1}_{A^c}$. Ensuite, on intègre un produit d'indicatrices. $\blacksquare$

Notation : désormais $(*)$ désigne la propriété « la probabilité de l'intersection est le produit des probabilités » et $\perp$ désigne la propriété d'indépendance.

La bonne définition

On veut maintenant définir l'indépendance d'un nombre infini de familles, d'où un problème de produits infinis de probabilités, problème évité par la

Définition *La collection $(C_i)_{i \in I}$ de classes de parties de $\mathcal{F}$ est* **indépendante** *si et seulement si pour toute famille* **finie** *$J \subset I$ et tous $A_i \in C_i$, $i \in J$, on a $(*)$.*

Toujours dans la veine « les tribus c'est bien difficile ma bonne dame, heureusement que les π–systèmes c'est facile », on démontre le

Théorème *Soit $(C_i)_i$ des π–systèmes. Alors, les C_i sont $\perp$ si et seulement si les tribus engendrées $\sigma(C_i)$ sont $\perp$.*

Démonstration $\Leftarrow$ évident.
Pour $\Rightarrow$, on suppose $I = \{1, \ldots, n\}$ (exercice) et on introduit

$$L_1 = \{B_1 \in \sigma(C_1)\,;\, \forall A_i \in C_i, P(B_1 \cap A_2 \cap \cdots \cap A_n) = (*)\}.$$

Puis théorème λ–π itéré. ■

Corollaire 1.16 (Coalitions) *Soit $(\mathcal{F}_i)_{i\in I}$ des sous–tribus $\perp$ de $\mathcal{F}$ et soit $I = \bigcup_{a\in A} I_a$ une partition de I. Alors, les tribus $(\mathcal{G}_a)_{a\in A}$ sont $\perp$ avec*

$$\mathcal{G}_a = \sigma\Big(\bigcup_{i\in I_a} \mathcal{F}_i\Big).$$

Démonstration On introduit la classe $\mathcal{C}_a$ des intersections de $A_i \in \mathcal{F}_i$ pour $i \in J$, J partie finie de I_a. Alors, $\mathcal{C}_a$ est un π–système qui contient la réunion des $\mathcal{F}_i$ pour $i \in I_a$ donc la tribu engendrée par $\mathcal{C}_a$ est $\mathcal{G}_a$. On finit par λ–π. ■

Indépendance de variables aléatoires

Une application importante du théorème précédent concerne les lois de variables aléatoires.

Définition *Des variables aléatoires définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$ sont $\perp$ si et seulement si les tribus engendrées sont $\perp$.*

Par conséquent, si deux variables aléatoires X et Y vérifient

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

pour tous réels x et y, les π–systèmes $\pi(X)$ et $\pi(Y)$ sont $\perp$ donc les tribus $\sigma(X)$ et $\sigma(Y)$ aussi. De même, des variables aléatoires $X_1, \ldots, X_n$ sont $\perp$ si et seulement si on a

$$\mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) = \prod_i \mathbb{P}(X_i \leqslant x_i),$$

pour tous $x_i \in \mathbb{R}$. On peut aussi écrire cette propriété :

$$(X_i)_i(\mathbb{P}) = \otimes_i X_i(\mathbb{P}).$$

Dans le cas où chaque $X_i(\mathbb{P})$ admet une densité f_i par rapport à la mesure de Borel sur $\mathbb{R}$, la loi du vecteur aléatoire $(X_i)_i$ admet une densité f par rapport à la mesure de Borel sur $\mathbb{R}^n$ donnée par

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

Un cas particulier est le

Corollaire *Soit X_i des variables aléatoires indépendantes et $S = X_1 + \cdots + X_n$. Alors la loi de S est le produit de convolution*

$$S(\mathbb{P}) = X_1(\mathbb{P}) * \cdots * X_n(\mathbb{P}).$$

Démonstration On regarde $n = 2$. ■

Le corollaire 1.16 implique, quant à lui, le fait suivant : soit T un ensemble d'indices, $(X_t)_{t\in T}$ une collection de variables aléatoires indépendantes, $(\mathcal{T}_n)_n$ une partition de T et f_n des fonctions mesurables définies sur les espaces convenables. Alors les v.a. $(Y_n)_n$ sont $\perp$ avec

$$Y_n = f_n(X_t,\, t \in \mathcal{T}_n).$$

Démonstration $\sigma(Y_n) \subset \sigma(X_t\,;\, t \in \mathcal{T}_n)$ et coalitions. ■

Lemme de Borel–Cantelli

Soit $(A_n)_n$ une suite d'éléments de $\mathcal{F}$ et $A = \limsup_n A_n$.

Lemme 1.17 (Borel–Cantelli partie triviale) *Si la somme $\sum_n \mathbb{P}(A_n)$ converge, alors $\mathbb{P}(A) = 0$.*

Lemme 1.18 (Borel–Cantelli partie non triviale) *Si la somme $\sum_n \mathbb{P}(A_n)$ diverge et si les $(A_n)_n$ sont indépendants, alors $\mathbb{P}(A) = 1$.*

Contrexemple à la partie non triviale sans l'indépendance.

Démonstration 1) Puisque $A \subset \bigcup_{k\geqslant n} A_k$, on écrit $\mathbb{P}(\bigcup_{k\geqslant n} A_k) \leqslant \sum_{k\geqslant n} \mathbb{P}(A_k)$, qui est le reste d'une série convergente donc qui tend vers 0 quand n tend vers l'infini.

2) Par définition, $\mathbb{P}(A) = \inf_n \mathbb{P}(\bigcup_{k\geqslant n} A_k)$. Indépendance donc on veut une intersection (finie !) :

$$\mathbb{P}(\bigcap_k A_k^c) = \prod_k (1 - \mathbb{P}(A_k)) \leqslant \exp(-\sum_k \mathbb{P}(A_k)).$$

On l'écrit pour $n \leqslant k \leqslant m$, on fait tendre m vers l'infini, le produit tend vers 0 donc pour tout n,

$$\mathbb{P}(\bigcup_{k\geqslant n} A_k) = 1.$$

■

Exemple 1.19 (Singe dactylographe) (et nombres normaux, plus tard)
Soit $(x_k)_k$ i.i.d. uniforme sur l'ensemble des caractères (lettres + ponctuations), ensemble de cardinal N et s un mot de longueur L. Alors

$$\mathbb{P}((x_1, x_2, \ldots, x_L) = s) = N^{-L} = \mathbb{P}((x_{i+1}, x_{i+2}, \ldots, x_{i+L}) = s)$$

est strictement positif. Pas indépendance donc on regarde les

$$\{(x_{iL+1}, x_{iL+2}, \ldots, x_{(i+1)L}) = s\}$$

qui le sont. Alors, si on écrit s en un temps multiple de L, on a écrit s au moins une fois donc

$$\mathbb{P}(\limsup_n \{\text{on écrit } s \text{ au temps } n\}) = 1$$

c'est-à-dire que l'on écrit s une infinité de fois. Temps moyen de la première écriture : N^L lettres. La Bible et plusieurs milliers de lettres écrites par seconde donne beaucoup de fois l'âge de l'Univers. ■

Exercice Soit X_n des variables aléatoires indépendantes de loi commune exponentielle :

$$\mathbb{P}(X_n \geqslant x) = \exp(-x).$$

Montrer que

$$\mathbb{P}(X_n > a \log n \text{ pour une infinité de } n)$$

vaut 0 si $a > 1$ et 1 si $a \leqslant 1$. En déduire que $\limsup X_n / \log n$ vaut 1 et $\liminf n\, X_n$ vaut 0 presque sûrement.

Tribu asymptotique

Définition *Soit $\mathcal{F}_n$ des sous–tribus de $\mathcal{F}$. On note*

$$\mathcal{F}^n = \sigma(\bigcup_{k \geqslant n} \mathcal{F}_k), \quad \mathcal{F}^\infty = \bigcap_{n \geqslant 1} \mathcal{F}^n.$$

Exemples : choisir chaque $A_n \in \mathcal{F}_n$ donne $\limsup_n A_n \in \mathcal{F}^\infty$ (événement asymptotique).

Théorème 1.20 (Loi du zéro–un de Kolmogorov)
Soit $(\mathcal{F}_n)_n$ une suite de sous–tribus de $\mathcal{F}$ indépendantes. Alors $\mathcal{F}^\infty$ est triviale au sens où, pour tout $A \in \mathcal{F}^\infty$,

$$\mathbb{P}(A) \in \{0, 1\}.$$

Démonstration Pour tout n, soit $\mathcal{G}_n = \sigma\left(\bigcup_{k<n} \mathcal{F}_k\right)$. Les coalitions donnent l'indépendance des tribus $\mathcal{G}_n$ et $\mathcal{F}^n$, donc des tribus $\mathcal{G}_n$ et $\mathcal{F}^\infty$. Soit $\mathcal{C}$ la réunion des $\mathcal{G}_n$ pour $n \geqslant 1$. Alors $\mathcal{C}$ est un π–système $\perp$ de $\mathcal{F}^\infty$ et $\sigma(\mathcal{C}) = \sigma\left(\bigcup_k \mathcal{F}_k\right)$, qui contient $\mathcal{F}^\infty$, donc $\mathcal{F}^\infty \perp \mathcal{F}^\infty$. ■

Exemple 1.21 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. Préciser lesquels des événements suivants sont asymptotiques.

$$A = \{X_n = 0 \text{ une infinité de fois}\}\,; \qquad B = \{\text{la suite } (X_n)_n \text{ est bornée}\}\,;$$
$$C = \{\text{la suite } (X_n)_n \text{ converge}\}\,; \qquad D = \{\forall n, X_n \leqslant 3\}\,; \qquad E = \{\sup_{n \geqslant 1} X_n \geqslant 3\}\,;$$
$$F = \{X_n = 0 \text{ au moins deux fois}\}.$$

Pour un mouvement brownien $(B_t)_t$, l'événement suivant est asymptotique (et de probabilité 1) :

$$G = \left\{ \limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = +1,\ \liminf_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = -1 \right\}.$$

■

1.4 Probabilités produits

Questions : peut-on construire une suite $(X_n\,;\,n\geqslant 1)$ de v.a. indépendantes de lois fixées de fonctions de répartition F_n ? et une suite de variables aléatoires de lois non indépendantes mais fixées ?

On démontrera en TP que la réponse à la première question est oui, par les suites de v.a. de Bernoulli 0 ou 1 de probabilité 1/2, puis par sous–suites, puis par représentation de Skorokhod.

En cours, on traite le premier cas par application de Carathéodory puis le cas général.

Cas indépendant

Soit T un ensemble quelconque et pour $t\in T$, soit $(\Omega_t,\mathcal{F}_t,P_t)$ un espace de probabilité. On convient d'écrire $I\overset{\mathrm{F}}{\subset}T$ si $I\subset T$ et si I est fini. Pour $I\subset T$, soit Ω_I le produit $\prod_{t\in I}\Omega_t$ et π_I la projection canonique de Ω_T sur Ω_I. Pour tout $I\overset{\mathrm{F}}{\subset}T$, notons $P_I=\bigotimes_{t\in I}P_t$ la probabilité produit ordinaire sur $(\Omega_I,\mathcal{F}_I)$ où $\mathcal{F}_I=\bigotimes_{t\in I}\mathcal{F}_t$ est la tribu produit ordinaire.

Question : existe-t-il une tribu $\mathcal{F}$ sur $\Omega=\Omega_T$ et une probabilité P sur $\mathcal{F}$ telle que $\pi_I(P)=P_I$ pour tout $I\overset{\mathrm{F}}{\subset}T$?

Motivation : On dispose de $X_t:(\Omega_t,\mathcal{F}_t,P_t)\to\mathbb{R}$ de loi donnée. On cherche à construire Ω et $Y_t:\Omega\to\mathbb{R}$ de façon que Y_t suive la loi de X_t.
On va voir que la réponse consiste à choisir pour Ω le produit des Ω_t et à poser $Y_t(\omega)=X_t(\omega_t)$ pour $\omega=(\omega_t)_t$. Mais pour cela, il faut construire $\mathcal{F}$ et P convenablement.

Définition 1.22 *Soit $\mathcal{F}$ la tribu engendrée par les applications π_I pour $I\overset{\mathrm{F}}{\subset}T$. On note $\mathcal{F}=\bigotimes_{t\in T}\mathcal{F}_t$. Par ailleurs, pour $I\overset{\mathrm{F}}{\subset}T$, on appelle $\sigma(\pi_I)$ la tribu des cylindres de base dans I.*

Théorème 1.23 *Il existe une unique mesure de probabilité P sur*

$$(\Omega,\mathcal{F})=(\prod_{t\in T}\Omega_t,\bigotimes_{t\in T}\mathcal{F}_t)$$

telle que, pour tout $I\overset{\mathrm{F}}{\subset}T$, on a $\pi_I(P)=P_I$. On note $P=\bigotimes_{t\in T}P_t$.

Démonstration Soit $\mathcal{A}$ la réunion des $\sigma(\pi_I)$ pour tout $I\overset{\mathrm{F}}{\subset}T$. Alors $\mathcal{A}$ est une algèbre qui engendre $\mathcal{F}$. On définit une fonction d'ensembles Q sur $\mathcal{A}$ de la façon suivante : si $A\in\sigma(\pi_I)$, il existe un $B\in\mathcal{F}_I$ tel que $A=\pi_I^{-1}(B)$, posons $Q(A)=P_I(B)$.

La définition est cohérente : si $A\in\sigma(\pi_I)$, il existe un unique $B\in\mathcal{F}_I$ tel que $A=\pi_I^{-1}(B)$ car π_I est surjective ; d'autre part, si $I\subset J\overset{\mathrm{F}}{\subset}T$ et si $A\in\sigma(\pi_I)$, alors $A=\pi_I^{-1}(B)$ avec $B\in\mathcal{F}_I$ donc $A=\pi_J^{-1}(B\times\Omega_{J\setminus I})$ et, comme il se doit, $P_I(B)=P_J(B\times\Omega_{J\setminus I})$.

La définition est nécessaire : si on veut avoir $\pi_I(P)=P_I$, alors la mesure de $A=\pi_I^{-1}(B)$ doit être $P(A)=P(\pi_I^{-1}(B))=P_I(B)$.

On voit facilement que $0\leqslant Q\leqslant 1$, $Q(\Omega)=1$ et que Q est additive sur $\mathcal{A}$. Si Q est σ–additive sur $\mathcal{A}$, on aura terminé la démonstration en appliquant le théorème de Carathéodory. Tout se ramène donc à montrer que Q vérifie la contraposée de la σ–additivité, souvent appelée propriété de Carathéodory, et qui s'énonce comme suit :

$$\forall (A_n)_{n\geqslant 1} \subset \mathcal{A},\ (\exists \varepsilon > 0,\ \forall n \geqslant 1,\ A_{n+1} \subset A_n,\ Q(A_n) \geqslant \varepsilon) \implies \bigcap_{n\geqslant 1} A_n \neq \emptyset.$$

Carathéodory usuel : On se donne $m_0 : \mathcal{F}_0 \to \mathbb{R}^+$, fonction σ–additive sur une algèbre $\mathcal{F}_0$. Alors, il existe une mesure m sur $\sigma(\mathcal{F}_0)$ telle que la restriction de m à $\mathcal{F}_0$ coïncide avec m_0. Si $m_0(\Omega)$ st finie, cette extension est unique.

Exemple usuel : Sur $\Omega = \mathbb{R}$, l'ensemble des réunions finies d'intervalles bornés est une algèbre et m_0 est la fonction longueur. ■

Reprenons la démonstration du théorème et montrons que Q vérifie la propriété de Carathéodory.

Soit $(A_n)_n$ une suite d'éléments de $\mathcal{A}$ comme dans la propriété de Carathéodory. Chaque A_n fait intervenir un nombre fini de $t \in T$, donc on peut imposer $T = \mathbb{N}^*$ et, en répétant éventuellement les A_n, on peut supposer que $A_n \in \sigma(\pi_{[1,n]})$.

On va montrer qu'il existe $\omega^* \in \Omega$ tel que $\omega^* \in A_n$ pour tout $n \geqslant 1$. Pour cela, pour $n \geqslant 1$ et $I \overset{\mathrm{F}}{\subset}]n, +\infty[$, appelons π_I^n la projection canonique de $\Omega_{]n,+\infty[}$ sur Ω_I et notons $\mathcal{A}_n$ la réunion des $\sigma(\pi_I^n)$ pour $I \overset{\mathrm{F}}{\subset}]n, +\infty[$. Alors $\mathcal{A}_n$ est une algèbre et on peut définir une fonction d'ensemble Q_n sur $\mathcal{A}_n$ comme on avait défini Q sur $\mathcal{A}$. Pour $A \in \mathcal{A}$ et $\omega \in \Omega_{[1,n]}$, notons

$$A(\omega) = \{\omega' \in \Omega_{]n,+\infty[}\,;\, (\omega, \omega') \in A\}.$$

Alors $A(\omega) \in \mathcal{A}_n$ et, par Fubini, on a

$$Q(A) = \int dP_1(\omega_1) \dots dP_n(\omega_n)\, Q_n(A((\omega_k)_{1\leqslant k\leqslant n})).$$

Revenons aux A_n. Notons

$$B_1^n = \{\omega_1 \in \Omega_1\,;\, Q_1(A_n(\omega_1)) \geqslant \varepsilon/2\}.$$

Il vient

$$Q(A_n) = \int dP_1(\omega_1)\, Q_1(A_n(\omega_1)) = \int_{B_1^n} + \int_{\Omega\setminus B_1^n} \leqslant P_1(B_1^n) + \varepsilon/2,$$

donc $P_1(B_1^n) \geqslant \varepsilon/2$. Or, la suite $(B_1^n)_n$ décroît et P_1 est une mesure donc P_1 vérifie la condition de Carathéodory et il existe au moins un $\omega_1^* \in \Omega_1$ tel que $Q_1(A_n(\omega_1^*)) \geqslant \varepsilon/2$ pour tout n.

Étape suivante : on remplace A_n par $A_n(\omega_1^*)$, ε par $\varepsilon/2$ et Ω_1 par Ω_2. Il existe donc au moins un $\omega_2^* \in \Omega_2$ tel que $Q_2(A_n(\omega_1^*, \omega_2^*)) \geqslant \varepsilon/4$.
De proche en proche, il existe $\omega^* = (\omega_k^*)_k \in \Omega$ tel que pour tout n et tout k, on ait

$$Q_k(A_n(\omega_1^*, \omega_2^*, \dots, \omega_k^*)) \geqslant \varepsilon/2^k.$$

Montrons que ω^* est un élément de l'intersection des A_n. Fixons $n \geqslant 1$ et montrons que ω^* est un élément de A_n. On sait que $A_n \in \sigma(\pi_{[1,n]})$ et que

$$Q_n(A_n(\omega_1^*, \omega_2^*, \dots, \omega_n^*)) \geqslant \varepsilon/2^n > 0,$$

donc $A_n(\omega_1^*, \omega_2^*, \dots, \omega_n^*))$ n'est pas vide. Mais A_n est un $[1, n]$–cylindre donc

$$(\omega_1^*, \omega_2^*, \dots, \omega_n^*, \omega')$$

est un élément de A_n pour tout ω'. En particulier $\omega^* \in A_n$. ■

Corollaire *Soit $(E, \mathcal{E}, Q)$ un espace de probabilité et T un ensemble quelconque. Il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et des variables aléatoires indépendantes $X_t : \Omega \to E$ pour $t \in T$ telles que $X_t(\mathbb{P}) = Q$.*

Démonstration On choisit $(\Omega_t, \mathcal{F}_t, P_t) = (E, \mathcal{E}, Q)$ pour tout $t \in T$ et

$$(\Omega, \mathcal{F}, \mathbb{P}) = \left(\prod_{t \in T} \Omega_t, \bigotimes_{t \in T} \mathcal{F}_t, \bigotimes_{t \in T} P_t \right).$$

Alors, $X_t(\omega) = \omega_t$ pour $\omega = (\omega_t)_{t \in T}$ convient. ∎

Cas général sur $\mathbb{R}^T$

Rappelons que pour $I \subset J \subset T$, on note $\pi_I : \mathbb{R}^T \to \mathbb{R}^I$ et $\pi_{I,J} : \mathbb{R}^J \to \mathbb{R}^I$ les projections canoniques. On sait aussi que $\mathcal{B}(\mathbb{R}^I) = \mathcal{B}(\mathbb{R})^{\otimes I}$ si I est fini.

Définition *Un système cohérent de probabilités sur $\mathbb{R}^T$ est une collection*

$$\left\{ P_I \, ; \, I \overset{\mathrm{F}}{\subset} T \right\}$$

de mesures de probabilités P_I sur $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I))$ avec $P_I = \pi_{I,J}(P_J)$ pour tout $I \subset J \overset{\mathrm{F}}{\subset} T$.

Le résultat suivant figure dans le mémoire historique *Grundbegriffe der Wahrsheinlichkeitrechnung* de Kolmogorov paru en 1933.

Théorème 1.24 (Théorème d'extension de Kolmogorov) *Tout système cohérent de mesures de probabilité P_I sur $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I))$ pour $I \overset{\mathrm{F}}{\subset} T$ admet une unique extension, au sens où il existe une unique probabilité P sur $(\mathbb{R}^T, \mathcal{B}(\mathbb{R})^{\otimes T})$ telle que $P_I = \pi_I(P)$ pour tout $I \overset{\mathrm{F}}{\subset} T$.*

Démonstration Admis. ∎

Remarque On ne dispose pas d'une probabilité sur $\mathcal{B}(\mathbb{R}^T)$ mais seulement sur la tribu cylindrique

$$\mathcal{B}(\mathbb{R})^{\otimes T} \subset \mathcal{B}(\mathbb{R}^T).$$

Si T est dénombrable, $\mathbb{R}^T$ est séparable donc sa topologie est à base dénombrable et les deux tribus sont égales. Si T n'est pas dénombrable, les deux tribus sont différentes.

Un exemple où les deux tribus sont différentes est le suivant. On pose $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $T = [0, 1]$, et on définit $f : \mathbb{R} \to Y = \mathbb{R}^T$ par $f(\omega) = \mathbf{1}_{\{\omega\}}$. La topologie de Y est la topologie produit, définie par la convergence simple des fonctions : une suite $(g_n)_n$ de Y converge vers g si et seulement si, pour tout $x \in \mathbb{R}$, $g_n(x) \to g(x)$. Choisissons alors $A \subset \mathbb{R}$ et notons

$$Z = \{g \in Y \, ; \, \exists x \in A, \, g(x) > 1/2\}.$$

Comme Z est ouvert, Z est un borélien de Y. De plus, $f^{-1}(Z) = A$ donc il suffit de choisir A non Lebesgue mesurable pour en conclure que f n'est pas $\mathcal{B}(\mathbb{R})/\mathcal{B}(Y)$ mesurable.

Par contre, f est $\mathcal{B}(\mathbb{R})/\mathcal{G}$ mesurable avec $\mathcal{G} = \mathcal{B}(\mathbb{R})^{\otimes T}$. En effet, soit C un cylindre élémentaire de $\mathcal{G}$, donc $C = A_1 \times \cdots \times A_s \times \mathbb{R}^{T\setminus S}$ avec $S = \{i_1, \ldots, i_s\}$ fini et $A_i \in \mathcal{B}(\mathbb{R})$. Alors,

$$f^{-1}(C) = \{x \in \mathbb{R}\,;\, \forall i \in S,\ f(x)(i) \in A_i\}.$$

On voit que $f^{-1}(C)$ contient tous les points de $\mathbb{R} \setminus S$ ou aucun, selon que 0 est un élément de l'intersection des A_i ou non. Donc, $f^{-1}(C)$ est une partie de S ou bien contient $\mathbb{R} \setminus S$. En tout cas, $f^{-1}(C) \in \mathcal{B}(\mathbb{R})$, donc f est $\mathcal{B}(\mathbb{R})/\mathcal{G}$ mesurable.

Une conséquence est que $\mathcal{B}(\mathbb{R}^T) \neq \mathcal{B}(\mathbb{R})^{\otimes T}$ si $T = [0, 1]$. (Source : Dudley.)

1.5 Rappels de théorie de la mesure (d'après D. Williams)

Un résumé des faits à connaître absolument :

• Un π–système qui engendre la tribu $\mathcal{B}(\mathbb{R})$:

$$\pi(\mathbb{R}) = \{]-\infty, x]\,;\, x \in \mathbb{R}\}.$$

• Si $(\Omega, \mathcal{F}, m)$ est un espace mesuré, on note $\mathcal{E}^+ = \mathcal{E}^+(m)$ l'ensemble des fonctions "simples", c'est-à-dire des $h = \sum_{k=1}^n a_k \mathbf{1}_{A_k}$ pour $n \geqslant 1$, $a_k \geqslant 0$ et $A_k \in \mathcal{F}$. On note alors

$$m_0(h) = \sum_k a_k\, m(A_k).$$

Les fonctions f mesurables positives, ce que l'on note $f \in \mathcal{L}^+ = \mathcal{L}^+(m)$, sont les suprema de fonctions de $\mathcal{E}^+$ et on pose alors

$$m(f) = \sup\{m_0(h)\,;\, h \in \mathcal{E}^+,\ h \leqslant f\}.$$

On note indifféremment $m(f) = \displaystyle\int f\,\mathrm{d}m = \int f(\omega)\,m(\mathrm{d}\omega)$ et, pour $A \in \mathcal{F}$,

$$m(f : A) = \int_A f\,\mathrm{d}m = \int_A f(\omega)\,m(\mathrm{d}\omega) = m(f\,\mathbf{1}_A).$$

• Les cinq outils fondamentaux :

– Convergence monotone (MON) : si $f_n \in \mathcal{L}^+$ converge en croissant vers f, alors $f \in \mathcal{L}^+$ et $m(f_n)$ tend vers $m(f)$.

– Lemme de Fatou (FAT) : si $f_n \in \mathcal{L}^+$, alors

$$m(\liminf_n f_n) \leqslant \liminf_n m(f_n).$$

Deux exemples importants : $f_n = \mathbf{1}_{[n,+\infty[}$ et $f_n = n\,\mathbf{1}_{]0,1/n[}$.

– Convergence dominée (DOM) : si $f_n \in \mathcal{L}$ converge vers f et si $|f_n| \leqslant g$ avec $g \in \mathcal{L}^+$ et $m(g)$ finie, alors $m(f_n)$ tend vers $m(f)$.

– Lemme de Scheffé (SCH) : soit $f_n \in \mathcal{L}^+$ une suite qui converge vers f. Alors, f_n tend vers f dans $\mathcal{L}^1$ (c'est-à-dire $m(|f_n - f|)$ tend vers 0) si et seulement si $m(f_n)$ tend vers $m(f)$.
Note : les f_n sont positives. La preuve : $(f - f_n)^+ \leqslant f$ donc (DOM) puis $|x| = 2x^+ - x$.

– Procédé standard (STA) : $h = \mathbf{1}_A$ pour $A \in \mathcal{F}$ puis $h \in \mathcal{E}^+$ (en général par linéarité) puis $h \in \mathcal{L}^+$ (en général par (MON)) puis $h = h^+ - h^-$.

• Produits : soit $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(A_1 \times A_2\,;\, A_i \in \mathcal{F}_i)$. Il existe une mesure unique m notée $m = m_1 \otimes m_2$ sur $\mathcal{F}$ telle que $m(A_1 \times A_2) = m_1(A_1) \times m_2(A_2)$ pour tout $A_i \in \mathcal{F}_i$ si néammoins m_1 et m_2 sont σ–finies.

Alors toute $f \in \mathcal{L}^+(m)$ vérifie le THÉORÈME DE FUBINI (FUB) :

$$m(f) = \int_{\Omega_1} I_1^f(\omega_1)\, m_1(\mathrm{d}\omega_1) = \int_{\Omega_2} I_2^f(\omega_2)\, m_2(\mathrm{d}\omega_2),$$

si on pose

$$I_1^f(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2)\, \mathrm{d}m_2(\omega_2), \qquad I_2^f(\omega_2) = \int_{\Omega_1} f(\omega_1, \omega_2)\, \mathrm{d}m_1(\omega_1).$$

Contrexemple : $[0, 1]$ muni de ses boréliens ; m_1 mesure de Borel, m_2 mesure de comptage, f fonction indicatrice de $\{\omega_1 = \omega_2\}$.

Application : soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^+$ une variable aléatoire positive et m la mesure sur $\Omega \times \mathbb{R}^+$ produit de $\mathbb{P}$ et de la mesure de Borel. Notons $A = \{(\omega, x) ; 0 \leqslant x \leqslant X(\omega)\}$ et $f = \mathbf{1}_A$. Alors, $I_1^f(\omega) = X(\omega)$ et $I_2^f(x) = \mathbb{P}(X \geqslant x)$. On en déduit

$$m(A) = \int X(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_0^{+\infty} \mathbb{P}(X \geqslant x)\, \mathrm{d}x.$$

Dessin : l'aire en dessous du graphe est $\mathbb{E}(X) = \displaystyle\int X\, \mathrm{d}\mathbb{P}$.

1.6 Espérance

Définition *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^+$ une variable aléatoire positive. Son espérance est son intégrale relativement à $\mathbb{P}$:*

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

Analogue pour $X \in \mathcal{L}^1$, puis pour X à valeurs dans un espace vectoriel normé. Une notation importante : si $A \in \mathcal{F}$, on note

$$\mathbb{E}(X : A) = \mathbb{E}(X; A) = \mathbb{E}(X\, \mathbf{1}_A).$$

Proposition *Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire et $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne. Alors, si f est positive, on a*

$$\mathbb{E}(f(X)) = \int_\mathbb{R} f(x)\, \mathrm{d}m(x),$$

pour la mesure $m = X(\mathbb{P})$ sur $\mathcal{B}(\mathbb{R})$. D'autre part, $f \in \mathcal{L}^1(m)$ si et seulement si $f(X) \in \mathcal{L}^1(\mathbb{P})$ et $\mathbb{E}(f(X))$ est alors donnée par la même expression.

Démonstration Procédé standard (STA). ∎

Définition *Soit X une variable aléatoire à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$ et soit $n \geqslant 1$ un entier. Si $|X|^n$ est intégrable, le moment de X d'ordre n est $\mathbb{E}(X^n)$. La variance est le moment d'ordre 2 centré :*

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

Si X est une variable aléatoire à valeurs dans un espace vectoriel $\mathbb{R}^d$ ou $\mathbb{C}^d$ telle que $\|X\|^2$ est intégrable, on note $\mathrm{Cov}(X)$ la matrice

$$\mathrm{Cov}(X) = \mathbb{E}((X - E(X))(X - E(X))^*),$$

et on note $\mathrm{cov}(X_i, X_j) = \mathrm{Cov}(X)_{ij}$.

Exercice Montrer que $\mathrm{var}(X)$ est le minimum des $\mathbb{E}(X - t)^2$ pour t réel.

Indépendance et intégrabilité

Proposition *Soit X_i des variables aléatoires intégrables. Si les X_i sont $\perp$, alors $Y = X_1 \cdots X_n$ est intégrable et*

$$\mathbb{E}(Y) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

Si maintenant les X_i sont des variables aléatoires $\mathcal{L}^2$ et $\perp$ deux à deux, alors $\operatorname{Cov}((X_i)_i)$ est la matrice diagonale des $\operatorname{var}(X_i)$. Par exemple, on a

$$\operatorname{var}(\sum_i X_i) = \sum_i \operatorname{var}(X_i).$$

Démonstration Fubini. ■

En particulier, si X et Y sont deux variables aléatoires $\mathcal{L}^2$ indépendantes, on a

$$\operatorname{cov}(X, Y) = 0 \quad \text{et} \quad \operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

Exercice Soit X_i, $1 \leqslant i \leqslant n$, des variables aléatoires et f_i des fonctions mesurables telles que les variables aléatoires $f_i(X_i)$ sont intégrables. Montrer que si les X_i sont $\perp$, alors

$$\mathbb{E}\left(\prod_{i=1}^{n} f_i(X_i)\right) = \prod_{i=1}^{n} \mathbb{E}(f_i(X_i)).$$

(Ici aussi, $\perp$ = multiplier.) Trouver deux variables aléatoires X et Y non $\perp$ telles que $\operatorname{cov}(X, Y) = 0$.

Exercice 1.25 (LGN $\mathcal{L}^4$) Soit $(X_n)_n$ une suite de variables aléatoires de $\mathcal{L}^4$ indépendantes, centrées et bornées dans $\mathcal{L}^4$; il existe c fini tel que, pour tout n, $\mathbb{E}(X_n^4) \leqslant c$. Soit $S_n = \sum_{k=1}^{n} X_k$. Alors S_n/n tend vers 0 presque sûrement.

Démonstration En effet, $\mathbb{E}(S_n^4) = \sum_k \mathbb{E}(X_k^4) + 6 \sum_{k<i} \mathbb{E}(X_k^2)\,\mathbb{E}(X_i^2)$. Par Cauchy–Schwarz, une borne est $nc + 6cn(n-1)/2 \leqslant 3cn^2$. On en déduit que $\sum_n (S_n/n)^4$ est intégrable ! Donc cette série aléatoire converge presque sûrement donc son terme général tend vers 0 presque sûrement. ■

Extension au cas $\mathbb{E}(X_n) = m_n$. Cas particulier : $(X_n)_n$ i.i.d. et $X_1 \in L^4$. Autre cas particulier : $X_n = \mathbf{1}(A_n)$. ■

Inégalités

Et quand les variables aléatoires ne sont pas indépendantes ? Troncature...

Proposition 1.26 (Markov) *Soit X une variable aléatoire positive et $a > 0$. Alors,*

$$\mathbb{P}(X \geqslant a) \leqslant a^{-1}\, \mathbb{E}(X).$$

On en déduit :

- **Markov composé :** soit X une variable aléatoire et f une fonction positive croissante et mesurable. Alors, pour tout a réel,
$$\mathbb{P}(X \geqslant a) \leqslant \mathbb{E}(f(X))/f(a).$$
En particulier, pour tout a réel et tout $t \geqslant 0$,
$$\mathbb{P}(X \geqslant a) \leqslant \mathrm{e}^{-ta}\,\mathbb{E}(\mathrm{e}^{tX}).$$
- **Chebyshev :** si X est de carré intégrable, pour tout $a > 0$,
$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant a) \leqslant \mathrm{var}(X)/a^2.$$

...ou convexité :

Proposition 1.27 (Jensen) *Soit X une variable aléatoire intégrable à valeurs dans un intervalle G de $\mathbb{R}$ et soit c une fonction convexe définie sur G telle que $c(X)$ est intégrable. Alors,*
$$c(E(X)) \leqslant E(c(X)).$$

Démonstration Si c est convexe, c est un supremum de fonctions affines. ∎

Notation : $m(f)$ est l'intégrale de f par rapport à m. À partir de Jensen, on peut montrer

Proposition 1.28 (Hölder et Minkowski) *On se donne une mesure quelconque m et p un réel $p > 1$. On note q le réel conjugué de p : c'est le réel $q > 1$ tel que $1/p + 1/q = 1$.*
- *Soit $f \in \mathcal{L}^p$ et $g \in \mathcal{L}^q$. Alors fg est intégrable et $|m(fg)| \leqslant m(|f|^p)^{1/p}\, m(|g|^q)^{1/q}$.*
- *Soit $f \in \mathcal{L}^p$ et $g \in \mathcal{L}^p$. Alors $m(|f+g|^p)^{1/p} \leqslant m(|f|^p)^{1/p} + m(|g|^p)^{1/p}$.*

En particulier, inégalité de Cauchy–Schwarz : si X et Y sont de carré intégrable, XY est intégrable et
$$\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\,\mathbb{E}(Y^2).$$

Démonstration Pour Hölder : $c(t) = t^q$ est convexe, $P = f^p \cdot m/m(f^p)$ est une probabilité et $X = g\,\mathbf{1}_{f\neq 0}/f^{p-1}$ convient.
Pour Minkowski : $(f+g)^p = f\,(f+g)^{p-1} + g\,(f+g)^{p-1}$ donc par Hölder,
$$m((f+g)^p) = m(f\,(f+g)^{p-1}) + m(g\,(f+g)^{p-1}) \leqslant \|f\|_p\, A + \|g\|_p\, A,$$
avec $A = \|(f+g)^{p-1}\|_q = m((f+g)^p)^{1/q}$. ∎

Application de l'espérance (en TP) : Weierstrass par Bernstein. Pour la convergence ponctuelle, f bornée et continue au point considéré suffit. Pour la convergence uniforme de $B_n(f)$ vers f, il faut écrire que f est uniformément continue.

1.7 Convergences

Soit X et X_n, $n \geqslant 1$, des variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On note :
- $X_n \xrightarrow{p.s.} X$ si $\mathbb{P}(C) = 1$ avec $C = \{\lim_n X_n = X\}$,
- $X_n \xrightarrow{L^p} X$ si $\mathbb{E}(|X - X_n|^p)$ tend vers 0.

Convergence presque sûre

On sait que si $X_n \xrightarrow{p.s.} X$, il existe une suite $(X'_n)_n$ de variables aléatoires telle que $\mathbb{P}(X_n = X'_n) = 1$ pour tout n et telle que $X'_n(\omega) \to X(\omega)$ pour tout $\omega \in \Omega$. On a une caractérisation plus forte faisant intervenir la convergence uniforme grâce au

Théorème 1.29 (Egorov) *$X_n \xrightarrow{p.s.} X$ si et seulement si pour tout $\varepsilon > 0$, il existe $A \in \mathcal{F}$ tel que $\mathbb{P}(A) \geqslant 1-\varepsilon$ et*

$$\sup\{|X_n(\omega) - X(\omega)|\,;\,\omega \in A\} \longrightarrow 0.$$

On appelle aussi ce mode de convergence la convergence presqu'uniforme.

Démonstration Pour la réciproque, soit A_p l'ensemble associé à $\varepsilon_p = 1/p$ et A la réunion des A_p pour $p \geqslant 1$. Alors, $A \in \mathcal{F}$ et $\mathbb{P}(A) = 1$.

Pour $\omega \in A$, il existe un indice p tel que $\omega \in A_p$. Or, pour $n \geqslant n_p$, $|X_n - X| \leqslant \varepsilon_p$ sur A_p donc $|X_n(\omega) - X(\omega)| \leqslant \varepsilon_p = 1/p$.

Pour le sens direct, soit $\varepsilon > 0$ et $\alpha > 0$. Notons $Y_n = \sup_{k \geqslant n} |X_k - X|$. Alors,

$$0 = \mathbb{P}(\limsup_n |X_n - X| \geqslant \alpha) = \mathbb{P}(\inf_n Y_n \geqslant \alpha) = \mathbb{P}\left(\bigcap_n \{Y_n \geqslant \alpha\}\right) = \lim_n \mathbb{P}(Y_n \geqslant \alpha),$$

donc $\mathbb{P}(Y_n \geqslant \alpha)$ tend vers 0. En choisissant $\alpha = 2^{-p}$, on en déduit l'existence d'une suite croissante $(n_p)_p$ telle que pour tout $p \geqslant 1$,

$$\mathbb{P}(Y_{n_p} \geqslant 2^{-p}) \leqslant \varepsilon 2^{-p}.$$

Si A est l'intersection des $\{Y_{n_p} < 2^{-p}\}$, on en déduit que $\mathbb{P}(A^c) \leqslant \varepsilon \sum_p 2^{-p} = \varepsilon$ et que A convient car pour tout p et tout $n \geqslant n_p$, le supremum de $|X_n - X|$ sur A est au plus celui de Y_{n_p} sur A_p qui vaut au plus 2^{-p}. ■

Remarque Le sens réciproque est vrai pour toute mesure. Le sens direct est faux pour les mesures non bornées. Exemple : mesure de Lebesgue sur $\mathbb{R}$ et $X_n = \mathbf{1}_{[n,+\infty[}$.

Convergence en probabilité

Motivation : si $X_n(\mathbb{P}) = (1-1/n)\,\delta_0 + (1/n)\,\delta_1$ et si les X_n sont indépendantes, on voit bien que X_n ressemble de plus en plus à la variable aléatoire nulle. Pourtant, par Borel-Cantelli non trivial, $\mathbb{P}(\lim_n X_n = 0) = 0$. D'où la

Définition *$X_n \xrightarrow{P} X$ si pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$.*

Proposition *La convergence presque sûre implique la convergence P. La convergence L^p implique la convergence P.*

Démonstration Convergence presque sûre : d'après la preuve d'Egorov, pour $\varepsilon > 0$, on sait que la quantité $\mathbb{P}(Y_n \geqslant \varepsilon)$ tend vers 0. C'est plus fort que la convergence P. Convergence L^p : inégalité de Markov. ■

Remarque Pour une mesure non bornée, $\xrightarrow{p.s.}$ n'implique pas $\xrightarrow{P}$. Et il n'y a pas d'autres implications.

Lemme *Si la somme* $\sum_n P(|X_n - X| \geqslant \varepsilon)$ *est finie pour tout* $\varepsilon > 0$*, alors* $X_n \xrightarrow{p.s.} X$.

Démonstration Borel–Cantelli, partie triviale. ■

Théorème *Si* $X_n \xrightarrow{P} X$*, il existe une suite extraite* $(\phi(n))_n$ *telle que* $X_{\phi(n)} \xrightarrow{p.s.} X$.

Démonstration On choisit $\phi(n)$ suffisamment grand pour avoir

$$\mathbb{P}(|X_{\phi(n)} - X| \geqslant 2^{-n}) \leqslant 2^{-n}.$$

La somme des 2^{-n} converge donc Borel–Cantelli, partie triviale, indique que presque sûrement, $|X_{\phi(n)}(\omega) - X(\omega)| \leqslant 2^{-n}$ à partir d'un certain rang $n_0(\omega)$. Donc, $X_{\phi(n)}(\omega)$ converge vers $X(\omega)$. ■

Remarque Dans l'exemple donné en motivation avec $X_n(\mathbb{P}) = (1-1/n)\delta_0+(1/n)\delta_1$, on voit que $\phi(n) = n^2$ convient. Par ailleurs, si on choisit $X_n(\mathbb{P}) = (1-1/n)\delta_0 + (1/n)\delta_{2^n}$, on voit que $(X_n)_n$ ne converge vers 0 dans L^p pour aucun $p > 0$. Donc, nécessité de la notion de convergence P.

Remarque Les limites aux sens de la convergence presque sûre, P et L^p de variables aléatoires sont presque sûrement uniques au sens où deux limites X et X' vérifient $\mathbb{P}(X \neq X') = 0$. On quotiente comme suit : $\mathcal{L}^0(\mathbb{P})$ est l'ensemble des variables aléatoires $X : \Omega \to \mathbb{R}$ ou $\mathbb{C}$, $\mathcal{N}(\mathbb{P})$ est l'ensemble des $X \in \mathcal{L}^0(\mathbb{P})$ telles que $\mathbb{P}(X \neq 0) = 0$ et $L^0(\mathbb{P}) = \mathcal{L}^0(\mathbb{P})/\mathcal{N}(\mathbb{P})$. On identifie alors $X \in \mathcal{L}^0(\mathbb{P})$ et la classe de X, élément de $L^0(\mathbb{P})$.

Exercice (1) Pour toute suite $(X_n)_n$, il existe une suite de réels $(a_n)_n$, $a_n > 0$, telle que la suite $(X_n/a_n)_n$ converge p.s. vers 0.
(2) Bernoulli indépendantes de paramètre p_n. CNS portant sur la suite $(p_n)_n$ pour avoir convergence P vers 0. Idem pour la convergence p.s.
(3) Soit f continue et $X_n \xrightarrow{P} X$. Alors, $f(X_n) \xrightarrow{P} f(X)$. (Commencer par f uniformément continue.)

Uniforme intégrabilité

Rappels : on note $\mathbb{E}(X : A) = \mathbb{E}(X \cdot \mathbf{1}_A)$ si X est intégrable et $A \in \mathcal{F}$. Pour toute variable aléatoire X, $X\mathbf{1}(X \geqslant t) \xrightarrow{P} 0$. Si de plus X est intégrable, alors $X\mathbf{1}(X \geqslant t) \xrightarrow{L^1} 0$.

Motivation : on veut étendre le résultat suivant.

Proposition 1.30 *Soit* $X \geqslant 0$ *intégrable. Alors,* $\mathbb{E}(X : X \geqslant t) \to 0$ *quand* t *tend vers l'infini. De plus,* $\mathbb{E}(X : A) \to 0$ *quand* $\mathbb{P}(A) \to 0$*, au sens où :*

$$\forall \varepsilon > 0,\ \exists \alpha > 0,\ \forall A \in \mathcal{F},\ \mathbb{P}(A) \leqslant \alpha \Longrightarrow \mathbb{E}(X : A) \leqslant \varepsilon.$$

Démonstration Sinon, on obtient une suite $(A_n)_n$ avec $\mathbb{P}(A_n) \leqslant 2^{-n}$ telle que X vérifie $\mathbb{E}(X : A_n) > \varepsilon$. Alors, le lemme de Fatou donne $\mathbb{E}(X : \limsup_n A_n) \geqslant \varepsilon$, ce qui est impossible car $\mathbb{P}(\limsup_n A_n) = 0$ par Borel–Cantelli, partie triviale. Ensuite, $\mathbb{P}(X \geqslant t) \to 0$ dès que X est presque sûrement finie. ■

Définition 1.31 *Une famille* H *de variables aléatoires est uniformément intégrable (u.i.) si* $H \subset L^1$ *et si :*

$$\forall \varepsilon > 0,\ \exists t \geqslant 0,\ \forall X \in H, \quad \mathbb{E}(|X| : |X| \geqslant t) \leqslant \varepsilon.$$

On notera $t_H(\varepsilon)$ ou t_H le réel t dont l'existence est ainsi garantie. Bien remarquer l'ordre des quantificateurs. Autre formulation :

$$\lim_{t\to\infty} \sup_{X\in H} \mathbb{E}(|X| : |X| \geqslant t) = 0.$$

Exemple • H et K u.i. Alors $H \cup K$ u.i. Preuve : $t_{H\cup K} = \max(t_H, t_K)$.

• H partie finie de L^1 (fait).

• H et K u.i. Alors $H + K$ u.i. Preuve : $t_{H\cup K}(4\varepsilon) = 2\max(t_H(\varepsilon), t_K(\varepsilon))$.

On remarque, pour $X, Y \geqslant 0$, et en séparant $\{X \geqslant Y\}$ et $\{X < Y\}$:

$$\mathbb{E}(X + Y : X + Y \geqslant 2t) \leqslant 2\,\mathbb{E}(X : X \geqslant t) + 2\,\mathbb{E}(Y : Y \geqslant t). \qquad (*)$$

• H bornée dans L^p avec $p > 1$.

Preuve : par Hölder, $\mathbb{E}(|X| : |X| \geqslant t) \leqslant \mathbb{E}(|X|^p)^{1/p}\,\mathbb{P}(|X| \geqslant t)^{1/q}$ et par Markov, on sait majorer le dernier terme par $\mathbb{P}(|X| \geqslant t) \leqslant \mathbb{E}(|X|^p)/t^p$. Finalement,

$$\mathbb{E}(|X| : |X| \geqslant t) \leqslant \mathbb{E}(|X|^p)/t^{p-1}.$$

• $H = \{X \in L^1 ; |X| \leqslant Y\}$ avec $Y \in L^1$. En effet, $\{Y\}$ est u.i. et, pour $X \in H$,

$$\mathbb{E}(|X| : |X| \geqslant t) \leqslant \mathbb{E}(Y : Y \geqslant t).$$

• H u.i. Alors l'enveloppe convexe de H est u.i. (à faire avec la caractérisation de l'exercice 1.33).

• H u.i. Alors l'adhérence de H dans L^1 est u.i. (exercice 1.33).

Lemme *Si H est u.i., alors H est bornée dans L^1.*

Démonstration La définition de l'uniforme intégrabilité pour $\varepsilon = 1$ donne $\mathbb{E}(|X| : |X| \geqslant t) \leqslant 1$. Donc $\mathbb{E}|X| \leqslant t + 1$ pour tout $X \in H$. ■

Le résultat important reliant convergence presque sûre et convergence dans L^1 est le théorème suivant, dit théorème de l'uniforme intégrabilité ou théorème de Vitali.

Théorème 1.32 *Soit $p \geqslant 1$. Les deux assertions suivantes sont équivalentes :*
- $X_n \xrightarrow{L^p} X$ *et la suite* $(X_n)_n$ *est dans* L^p
- $X_n \xrightarrow{P} X$ *et* $\{|X_n|^p ; n \geqslant 1\}$ *est u.i.*

Remarque Pas d'hypothèse $X \in L^p$. De plus, c'est un raffinement du théorème (DOM).

Démonstration Sens direct pour $p = 1$ **:** $X_n \xrightarrow{L^1} X$ donc X_n et X sont dans L^1 et $X_n \xrightarrow{P} X$. Soit $\varepsilon > 0$. Pour $n \geqslant n_\varepsilon$, $\mathbb{E}|X_n - X| \leqslant \varepsilon$. La famille $\{X_n, n \leqslant n_\varepsilon\}$ est finie et dans L^1 donc u.i. Il existe t_ε tel que, pour tout n_ε et tout $t \geqslant t_\varepsilon$,

$$\mathbb{E}(|X_n| : |X_n| \geqslant t) \leqslant \varepsilon.$$

Si $n \geqslant n_\varepsilon$, on remarque que la suite $(X_n)_n$ est bornée dans L^1, par exemple par c, donc

$$\mathbb{P}(|X_n| \geqslant t) \leqslant \mathbb{E}|X_n|/t \leqslant c/t$$

pour tout $n \geqslant 1$. Par ailleurs, d'après le résultat démontré en introduction et appliqué à X, il existe $\alpha > 0$ tel que tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) \leqslant \alpha$ vérifie $\mathbb{E}(|X| : A) \leqslant \varepsilon$. En particulier, si $n \geqslant n_\varepsilon$,

$$\mathbb{E}(|X_n| : |X_n| \geqslant t) \leqslant \mathbb{E}(|X| : |X_n| \geqslant t) + \mathbb{E}|X_n - X| \leqslant 2\varepsilon,$$

dès que $t \geqslant c/\alpha$. Ainsi, $(X_n)_n$ est u.i.

Sens réciproque pour $p = 1$ **:** Montrons d'abord que $X \in L^1$. $X_n \xrightarrow{P} X$ donc une sous–suite $(X_{\phi(n)})_n$ converge presque sûrement vers X. Par Fatou,

$$\mathbb{E}|X| = \mathbb{E} \liminf_n |X_{\phi(n)}| \leqslant \liminf_n \mathbb{E}|X_{\phi(n)}|,$$

donc $\mathbb{E}|X|$ est finie car la suite $(X_n)_n$ est u.i. donc bornée dans L^1.

Montrons maintenant que $X_n \xrightarrow{L^1} X$. Soit $\varepsilon > 0$ et $t \geqslant \varepsilon$. Alors, $\mathbb{E}|X_n - X| \leqslant a_n + b_n + c_n$ avec

$$\begin{aligned} a_n &= \mathbb{E}(|X_n - X| : |X_n - X| \leqslant \varepsilon), \\ b_n &= \mathbb{E}(|X_n - X| : \varepsilon \leqslant |X_n - X| \leqslant t), \\ c_n &= \mathbb{E}(|X_n - X| : |X_n - X| \geqslant t). \end{aligned}$$

On voit que $a_n \leqslant \varepsilon$ et $b_n \leqslant t\,\mathbb{P}(|X_n - X| \geqslant \varepsilon)$. Pour c_n, on remarque que $(X_n - X)_n$ est u.i. comme sous-famille de $H + K$ avec $H = \{X_n, n \geqslant 1\}$ et $K = \{-X\}$. Il existe donc t_ε tel que $c_n \leqslant \varepsilon$ pour le choix $t = t_\varepsilon$.

Une fois t_ε fixé, on remarque que $X_n \xrightarrow{P} X$. Donc, pour $n \geqslant n_\varepsilon$ avec n_ε assez grand et dépendant de la valeur de t_ε, on sait que $\mathbb{P}(|X_n - X| \geqslant \varepsilon) \leqslant \varepsilon/t_\varepsilon$ donc $b_n \leqslant \varepsilon$.

Finalement, $\mathbb{E}|X_n - X| \leqslant 3\varepsilon$ pour tout $n \geqslant n_\varepsilon$. ■

Exercice Faire la preuve pour $p > 1$. Indication $|x \pm y|^p \leqslant 2^p\,(|x|^p + |y|^p)$.

Exercice 1.33 Une CNS pour que H soit u.i. est que H soit bornée dans L^1 et vérifie :

$$\forall \varepsilon > 0,\ \exists \alpha > 0,\ \forall A \in \mathcal{F},\ \mathbb{P}(A) \leqslant \alpha \Longrightarrow \forall X \in H,\ \mathbb{E}(|X| : A) \leqslant \varepsilon.$$

Démonstration Sens direct : si H est u.i., on sait que H est bornée dans L^1. De plus, pour $\varepsilon > 0$, il existe $t \geqslant 0$ tel que $\mathbb{E}(|X| : |X| \geqslant t) \leqslant \varepsilon$ pour tout $X \in H$. Si $\alpha = \varepsilon/t$ et $A \in \mathcal{F}$ avec $\mathbb{P}(A) \leqslant \alpha$, alors $X \in H$ vérifie

$$\mathbb{E}(|X| : A) \leqslant \mathbb{E}(|X| : |X| \geqslant t) + \mathbb{E}(|X| : A,\ |X| \leqslant t) \leqslant \varepsilon + t\,\mathbb{P}(A) \leqslant 2\varepsilon.$$

Sens réciproque : H est bornée dans L^1, disons par c, donc $\mathbb{P}(|X| \geqslant t) \leqslant c/t$. On voit que $t \geqslant c/\alpha$ convient. ■

Exercice 1.34 Si $\mathbb{P}$ est diffuse, montrer que la condition « H bornée dans L^1 » est inutile dans l'exercice 1.33. Si $\mathbb{P}$ n'est pas diffuse, construire un contrexemple montrant que la condition « H bornée dans L^1 » est indispensable. On pourra considérer des variables aléatoires nulles partout sauf sur un atome de $\mathbb{P}$. ■

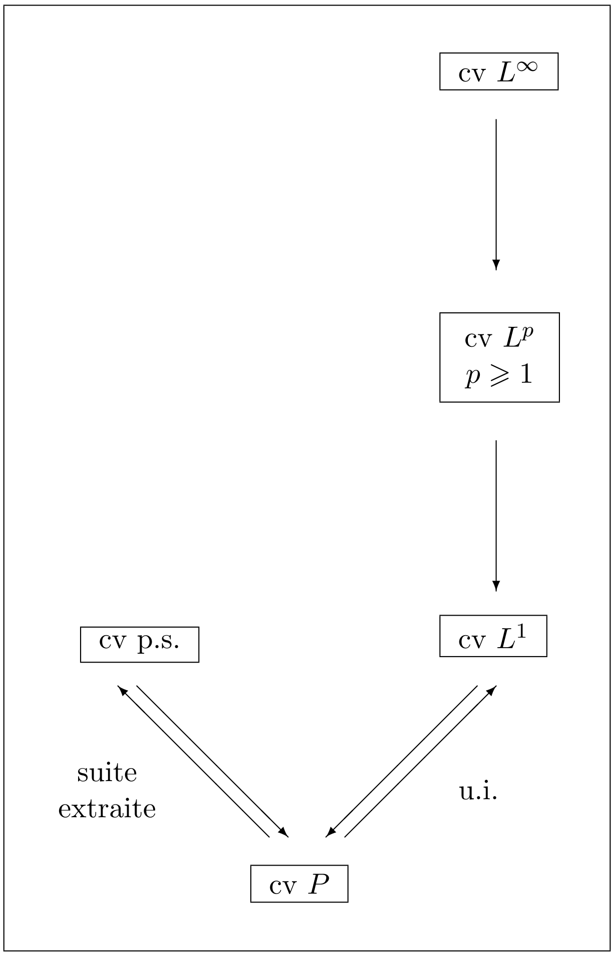
Exercice 1.35 Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction borélienne telle que $f(t)/t$ tend vers l'infini quand t tend vers l'infini. On suppose que l'ensemble des $\mathbb{E}(f(|X|))$, pour $X \in H$, est borné (on dit parfois que H est bornée dans L^f). Montrer que H est u.i.

Application : $f(t) = t^p$, $p > 1$, déjà fait ; $f(t) = t\,\log(t \vee 3)$.

Réciproquement : H est u.i. si et seulement s'il existe $f : \mathbb{R}_+ \to \mathbb{R}_+$ et c finie telles que $f(t)/t$ tend vers l'infini et $\mathbb{E}(f(|X|)) \leqslant c$ pour tout $X \in H$.

Démonstration On exhibe $(t_n)_n$ croissante avec $\mathbb{E}(|X|\,;\,|X| \geqslant t_n) \leqslant 2^{-n}$ pour tout $X \in H$. La fonction f définie par $f(t) = n\,t$ si $t_{n-1} \leqslant t < t_n$ convient. ■

Conclusion



Exercices

Classes

1 Pour $n \geqslant 0$, M_n est un « menteur ». M_0 émet un message ($+$ ou $-$) vers M_1 et chaque M_n, $n \geqslant 1$, transmet le message reçu de M_{n-1} vers M_{n+1} : il y a p chances, $0 \leqslant p \leqslant 1$, que M_n transmette le message reçu intact et $(1-p)$ chances qu'il le modifie. Trouver p_n, la probabilité que M_n transmette effectivement le message émis par M_0 et p_∞, la limite des p_n si elle existe.

2 Nabuchodonosor et Cléopâtre ont rendez-vous entre 5 et 7. Chacun a décidé d'attendre l'autre, si besoin est, pendant un quart d'heure. Calculer la probabilité que Nabuchodonosor et Cléopâtre réussissent leur rendez-vous.

3 Pour $n \geqslant 1$ on munit l'espace Σ_n des permutations de $\{1,\ldots,n\}$ de la probabilité uniforme P_n. Pour $\sigma \in \Sigma_n$, on note $F(\sigma)$ le nombre de points fixes de σ.

1) Exprimer $P_n(F=k)$ à l'aide de $p(i) = P_i(F=0)$. En déduire une relation entre les $p(i)$ puis la valeur de $p(i)$ (séries entières !). Trouver $\lim_{n\to\infty} P_n(F=k)$.

2) Établir la formule due à Poincaré :

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}(-1)^{k+1} \sum_{I\subset[1,n],|I|=k} P\left(\bigcap_{i\in I} A_i\right).$$

Indications : $P(A) = P(\mathbf{1}_A)$, $\mathbf{1}_{A\cap B} = \mathbf{1}_A\,\mathbf{1}_B$.

Retrouver directement les $p(i)$ à partir de cette formule.

4 Exhiber Ω et $\mathcal{C} \subset \mathcal{P}(\Omega)$ tel ques $\sigma(\mathcal{C}) = \mathcal{P}(\Omega)$ et deux probabilités P et Q distinctes sur $\mathcal{P}(\Omega)$ telles que $P = Q$ sur $\mathcal{C}$.

5 Exhiber $(\Omega, \mathcal{F})$ tel que $\mathcal{F}$ ne contient aucun singleton mais est en bijection avec $\mathcal{P}(\Omega)$.

6 Pour $A \subset \mathbb{N}^*$ et $n \geqslant 1$, on pose $d_n(A) = n^{-1}|A \cap [1,n]|$. Notons $\mathcal{A}$ l'ensemble des parties $A \subset \mathbb{N}^*$ telles que la « densité » $d(A) = \lim_n d_n(A)$ existe.

Exhiber $A \in \mathcal{A}$ et $B \in \mathcal{A}$ tels que $A \cap B \notin \mathcal{A}$. Donc $\mathcal{A}$ n'est pas un π–système.

Moralité : approcher une « mesure uniforme sur l'ensemble des entiers positifs » par la suite des mesures uniformes sur $[1,n]$ n'est pas la seule solution employée par les théoriciens des nombres. Pour tout $s > 1$, on considère la mesure de probabilité

$$m_s = \zeta(s)^{-1} \sum_{n\geqslant 1} n^{-s}\,\delta_n.$$

Outre des propriétés d'indépendance bien venues (voir l'exercice 20), montrer que $m_s(A) \to d(A)$ quand $s \to 1^+$, pour tout $A \in \mathcal{A}$.
Montrer que la convergence de $m_s(A)$ quand $s \to 1^+$ n'implique pas que $A \in \mathcal{A}$.

7 Soit $(A_n)_n$ une suite de parties de Ω. Montrer que

$$\mathbf{1}(\limsup A_n) = \limsup \mathbf{1}_{A_n}.$$

8 Soit $N : \Omega \to \mathbb{N}$ une fonction $\mathcal{F}/\mathcal{P}(\mathbb{N})$ mesurable et $X_n : \Omega \to \mathbb{R}$ des fonctions $\mathcal{F}/\mathcal{B}(\mathbb{R})$ mesurables. Montrer que $Y : \Omega \to \mathbb{R}$ définie par

$$Y(\omega) = X_{N(\omega)}(\omega),$$

est $\mathcal{F}/\mathcal{B}(\mathbb{R})$ mesurable.

9 (Atomes) On dit que $A \in \mathcal{F}$ est un atome de P si $P(A) > 0$ et si toute partie B de A mesurable vérifie $P(B) = 0$ ou $P(B) = P(A)$.
On note $A \sim B$ si et seulement si $P(A \triangle B) = 0$. La relation $\sim$ est une relation d'équivalence, on note A^* la classe de A.
1) $\mathcal{A} = \{A^* \,;\, A \in \mathcal{F},\ A \text{ atome}\}$ est au plus dénombrable.
2) Montrer que $d(A^*, B^*) = P(A \triangle B)$ définit une distance d sur l'ensemble des classes d'équivalence de la relation $\sim$.
3) Dans cette question $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et A est un atome. Montrer qu'il existe un réel x tel que $A^* = \{x\}^*$.
4) Dans cette question, P ne possède aucun atome. Soit $A \in \mathcal{F}$, $P(A) > 0$ et $0 < \varepsilon \leqslant P(A)$. On note

$$P_\varepsilon(A) = \sup\{P(B) \,;\, B \subset A,\ B \in \mathcal{F},\ P(B) \leqslant \varepsilon\}.$$

Montrer que $P_\varepsilon(A)$ n'est pas nul puis que $P_\varepsilon(A) \geqslant \varepsilon/2$. Pour montrer que $P_\varepsilon(A) \geqslant \varepsilon/2$, on pourra raisonner par l'absurde, montrer alors l'existence d'un événement $B \subset A$ tel que $P(B) = P_\varepsilon(A)$ et en déduire une contradiction.
5) Application : supposons qu'il existe un réel $0 \leqslant x \leqslant 1$ tel que P ne prend jamais la valeur x. Construire des B_n disjoints avec pour tout $n \geqslant 1$,

$$\sum_{k=1}^{n} 2^{-k} x \leqslant P(\bigcup_{k=1}^{n} B_k) \leqslant x.$$

Conclusion ?

10 On choisit deux points au hasard sur un segment de façon indépendante et uniforme. Calculer la probabilité que les trois portions du segment ainsi obtenues puissent former un triangle.

11 Cordes de Bertrand.

12 Soit $G = \langle a, b \rangle$ le groupe libre à deux générateurs. Probabilité de retour $1/3$ et longueur du mot réduit après n pas presque sûrement en $n/2 + o(n)$.

13 Soit $(\Omega, \mathcal{B}(\Omega))$ un espace métrique muni de ses boréliens et P une probabilité quelconque sur $\mathcal{B}(\Omega)$. Tout A borélien vérifie

$$P(A) = \inf\{P(O) \,;\, A \subset O,\ O \text{ ouvert}\} = \sup\{P(F) \,;\, F \subset A,\ F \text{ fermé}\}.$$

14 Soit $(\mathcal{F}_n)$ une suite croissante de tribus et $\mathcal{F}$ la tribu engendrée par leur réunion. Montrer que pour tous $\varepsilon > 0$ et $A \in \mathcal{F}$, il existe un indice n et un événement $A' \in \mathcal{F}_n$ tels que $P(A \triangle A') \leqslant \varepsilon$.

Variables aléatoires

15 Soit $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ une variable aléatoire et $\sigma(X) = X^{-1}(\mathcal{E})$. Vérifier que $\sigma(X)$ est bien « la plus petite tribu $\mathcal{G}$ rendant X mesurable pour $\mathcal{G}$ et $\mathcal{E}$ ».

Soit $\pi(X) = \{\{X \leqslant x\} \,;\, x \in \mathbb{R}\}$. Montrer que $\pi(X) = X^{-1}(\pi(\mathbb{R}))$ et que $\pi(X)$ est un π–système qui engendre $\sigma(X)$.

16 Soit $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ des variables aléatoires. Montrer que $\inf_n X_n$, $\liminf_n X_n$ et $\limsup_n X_n$ sont des variables aléatoires à valeurs dans $\overline{\mathbb{R}}$. Montrer que $\{\lim_n X_n \text{ existe}\}$ et $\{X_n \text{ est bornée}\}$ sont des éléments de $\mathcal{F}$.

17 (Représentation de Skorokhod) Soit $F : \mathbb{R} \to [0,1]$ une fonction de répartition.

1) On définit X sur $\Omega = [0,1]$ par $X(\omega) = \inf\{x\,;\, F(x) > \omega\}$. Montrer que

$$X(\omega) = \sup\{x\,;\, F(x) \leqslant \omega\}.$$

Montrer que X est $\mathcal{B}(\Omega)$ mesurable et que X admet pour fonction de répartition F par rapport à la mesure de Borel P.

2) Soit $Y(\omega) = \inf\{x\,;\, F(x) \geqslant \omega\}$. Montrer que $Y(\omega) = \sup\{x\,;\, F(x) < \omega\}$, que Y est $\mathcal{B}(\Omega)$ mesurable, que Y admet pour fonction de répartition F par rapport à P mesure de Borel. Montrer enfin que $P(X = Y) = 1$.

Indépendance

18 $\mathcal{C}_1$, $\mathcal{C}_2$ et $\mathcal{C}_3$ sont des π–systèmes avec $\mathcal{C}_i \subset \mathcal{F}$ et $\Omega \in \mathcal{C}_i$. Montrer que $(*)$ pour $A_i \in \mathcal{C}_i$, $i = 1, 2, 3$, entraîne $(*)$ pour $A_i \in \sigma(\mathcal{C}_i)$, $i = 1, 2, 3$. Pourquoi a-t-on imposé $\Omega \in \mathcal{C}_i$? Contrexemple ?

19 (Fonction indicatrice d'Euler) On prend $\Omega = \{1, 2, \ldots, n\}$, P probabilité uniforme et on note $\mathcal{P}$ l'ensemble des entiers de Ω premiers et divisant n. Pour $p \in \mathcal{P}$, soit $A_p = \{q \in \Omega\,;\, p|q\}$. Calculer $P(A_p)$. Montrer que les $(A_p)_{p \in \mathcal{P}}$ sont indépendants. En déduire que la fonction d'Euler φ vérifie

$$\varphi(n) = n \prod_{p \in \mathcal{P}} (1 - 1/p).$$

20 (Fonction Zêta) Pour $s > 1$, on note $\zeta(s) = \sum_{n \geqslant 1} n^{-s}$, $\mathcal{P}$ l'ensemble des nombres premiers et on se donne une variable aléatoire X à valeurs dans $\mathbb{N}^*$ de loi $\mathbb{P}(X = n) = n^{-s}/\zeta(s)$.

1) On note $A_n = \{n|X\}$ pour tout entier n. Montrer que les A_p pour $p \in \mathcal{P}$ sont indépendants et en déduire une preuve probabiliste de l'égalité

$$1/\zeta(s) = \prod_{p \text{ premier}} (1 - 1/p^s).$$

2) Montrer que la probabilité qu'aucun carré ne divise X est $1/\zeta(2s)$.

3) Soit Y une variable aléatoire indépendante de X et de même loi que X. On note D le p.g.c.d. (aléatoire) de X et Y. Montrer que D est une variable aléatoire et que sa loi est donnée par

$$\mathbb{P}(D = n) = n^{-2s}/\zeta(2s).$$

On pourra introduire $\Omega_0 = \{n|D\}$ et la probabilité $P_0(\cdot) = \mathbb{P}(\cdot|\Omega_0)$.

21 (Records) Soit $(X_n)_n$ une suite de variables aléatoires i.i.d. de fonction de répartition commune F continue. Soit $R_1 = \Omega$ et pour $n \geqslant 2$, $R_n = \{\forall m < n, X_m < X_n\}$. R_n est l'événement « un record est battu au temps n ». Montrer que les R_n sont indépendants et que $P(R_n) = 1/n$. Conclusion ?

22 (Tribu asymptotique) 1) Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires. On note, pour $n \geqslant 1$, $\mathcal{F}_n = \sigma(X_n)$ et $\mathcal{F}^n = \sigma(X_k\,;\, k \geqslant n)$, et enfin $\mathcal{F}^\infty = \bigcap_n \mathcal{F}^n$ la tribu asymptotique. Préciser si chacun des ensembles suivants appartient à $\mathcal{F}^\infty$ ou non.

$$A = \{X_n \text{ converge}\}, \quad B = \left\{\sum_n X_n \text{ converge}\right\}, \quad C = \left\{n^{-1}\sum_{k=1}^n X_k \text{ converge}\right\},$$

$$D = \left\{\lim_n n^{-1}\sum_{k=1}^n X_k = a\right\}, \quad E = \left\{\sum_n X_n = b\right\}.$$

2) Supposons de plus que les X_n sont indépendantes et soit Y une variable aléatoire $\mathcal{F}^\infty$ mesurable. Montrer qu'il existe $a \in \mathbb{R}$ tel que $\mathbb{P}(Y = a) = 1$.

23 1) Soit $a > 0$. Montrer que X est intégrable si et seulement si $\sum_n \mathbb{P}(|X| \geqslant an)$ converge.

2) En déduire que si $(X_n)_n$ est une suite de variables aléatoires i.i.d. alors : si X_1 est intégrablee, X_n/n tend vers 0 presque sûrement ; sinon, X_n/n n'est pas bornée presque sûrement.

Suites de Bernoulli

24 (Symétrie, symétrie ...) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi diffuse et symétrique par rapport à l'origine. On pose $S_n = \sum_{k=1}^n X_k$.

1) Montrer que la loi de S_n est diffuse.

2) Montrer sans calcul que $\mathbb{P}(S_2 > 0 \,|\, S_1 > 0) = 3/4$.

25 (Retours en zéro) Soit $(X_n)_n$ une suite de variables aléatoires i.i.d. de Bernoulli de paramètre p de loi $b(p)$: $X_1(\mathbb{P}) = p\delta_1 + (1-p)\delta_{-1}$. On pose $S_0 = 0$ et $S_n = \sum_{k=1}^n X_k$ si $n \geqslant 1$. On rappelle que S_n suit la loi binômiale $B(n,p)$.

1) Soit $a \in \mathbb{Z}$. Montrer que $\mathbb{P}(S_n = a)$ n'est non nul que si $n - a$ est pair et positif et que quand n tend vers l'infini, $\mathbb{P}(S_{a+2n} = a)$ est équivalent à

$$(2p)^a\,(4p(1-p))^n/\sqrt{n\pi}.$$

2) Si $p \neq 1/2$, montrer que $\mathbb{P}(\exists n_0,\, \forall n \geqslant n_0,\, S_n \neq a) = 1$.

Préciser si on peut conclure si $p = 1/2$.

Désormais, $p = 1/2$. On pose $T = \inf\{n \geqslant 1\,;\, S_n = 0\}$ si cet ensemble n'est pas vide et $T = +\infty$ sinon. On va montrer

$$\mathbb{P}(T < +\infty) = 1 \quad \text{et} \quad \mathbb{E}(T) = +\infty.$$

3) Soit $\mathcal{C}_n$ l'ensemble des chemins possibles entre les temps 0 et n : $c \in \mathcal{C}_n$ si c est une fonction continue sur $[0,n]$, nulle en 0, affine sur chaque intervalle $[k, k+1]$ et telle que $f(k+1) - f(k) = \pm 1$. Montrer qu'il existe une bijection g_n de $\{-1,+1\}^n$ sur $\mathcal{C}_n$ telle que la mesure $g_n((X_1,\ldots,X_n)(\mathbb{P}))$ soit la probabilité uniforme sur $\mathcal{C}_n$, que l'on note désormais U_n.

4) Soit $a \geqslant 1$ et b deux entiers. Posons

$$F_b = \{c \in \mathcal{C}_n \,;\, c(n) = b\}, \quad F_b^a = \{c \in F_b \,;\, \exists k \geqslant 1,\, c(k) = a\}.$$

On suppose $b \geqslant a$. Alors, $F_b = F_b^a$. On suppose $b < a$. Alors, faire un dessin pour démontrer le principe de réflexion suivant :

$$U_n(F_b^a) = U_n(F_{2a-b}).$$

On pourra remarquer que $(2a - b)$ est le symétrique de b par rapport à a.

5) Déduire du 4 les égalités :

$$\begin{aligned} \mathbb{P}(S_1 \neq 0, \dots, S_{2n} \neq 0) &= 2\,\mathbb{P}(S_1 = -1, S_2 < 0, S_3 < 0, \dots, S_{2n} < 0) \\ &= \mathbb{P}(S_1 < 1, S_2 < 1, \dots, S_{2n} < 1) \\ &= \sum_{k=0}^{n-1} U_{2n-1}(F_{2k-(2n-1)} \setminus F^1_{2k-(2n-1)}) = 4^{-n} \binom{2n}{n} = a_n. \end{aligned}$$

6) Montrer que $\mathbb{P}(T = 2n) = a_{n-1} - a_n$ pour tout $n \geqslant 1$ et en déduire

$$\mathbb{P}(T < +\infty) = 1 \quad \text{et} \quad \mathbb{E}(T) = +\infty.$$

Pour la suite, voir Cottrell, Duhamel et Genon-Catalot.

26 (Le mur) Soit $(\varepsilon_n)_n$ une suite de variables aléatoires i.i.d. à valeurs complexes de loi $p\,\delta_1 + (1-p)\,\delta_i$ et $S_n = \sum_{k=1}^{n} \varepsilon_k$. On fixe deux entiers a et b positifs et on note M le mur

$$M = \{a + i\,k \,;\, 0 \leqslant k \leqslant b\}.$$

1) Calculer $m = \mathbb{P}(\exists n \geqslant 1,\, S_n \in M)$.

2) En remarquant que S_{a+b} est sur la droite $\{x+iy \in \mathbb{Z}+i\mathbb{Z} \,;\, x+y = a+b\}$, obtenir une autre expression de m. Vérifier que les deux expressions obtenues sont égales.

27 (Construction d'une suite de variables aléatoires i.i.d.) On munit $\{0,1\}$ de la tribu $\mathcal{P}(\{0,1\})$ et de la probabilité $b(p)$ donnée par

$$b(p) = p\,\delta_1 + (1-p)\,\delta_0.$$

Soit Ω l'ensemble des applications $\omega = (\omega_n)_n$ de $\mathbb{N}^*$ dans $\{0,1\}$ où on note $\omega_n = \omega(n)$. On identifie Ω à un produit infini de copies de $\{0,1\}$ et on appelle $\mathcal{F}$ la tribu cylindrique sur Ω et P la probabilité produit des $b(p)$; P est donc déterminée par

$$P(\{\omega \in \Omega,\ \omega_1 = a_1, \dots, \omega_n = a_n\}) = p^{|a|}\,(1-p)^{n-|a|},$$

pour tout $n \geqslant 1$ et tout n–uplet $a = (a_k)_{k \leqslant n}$ formé de 0 et de 1 et de « longueur » $|a| = \sum_k a_k$. On définit enfin des fonctions $\varepsilon_n : \Omega \to \{0,1\}$ par $\varepsilon_n(\omega) = \omega_n$.

A) 1) Montrer que ε_n est une variable aléatoire de loi $b(p)$ pour tout $n \geqslant 1$. Montrer que la suite $(\varepsilon_n)_n$ est indépendante.

2) Soit $X : \Omega \to \mathbb{R}$ donnée par $X = \sum_{n \geqslant 1} 2^{-n}\,\varepsilon_n$. Montrer que X est une variable aléatoire à valeurs dans $[0,1]$.

3) On suppose jusqu'à nouvel ordre que $p = 1/2$. Calculer la loi de X. Construire une variable aléatoire sur $(\Omega,\ \mathcal{F},\ \mathbb{P})$ de loi uniforme sur un intervalle $[a, b]$.

4) Si I est une partie infinie de $\mathbb{N}$ de la forme $I = \{i_n\,;\, n \geqslant 1\}$, montrer que X^I suit la même loi que X, avec
$$X^I = \sum_{n\geqslant 1} 2^{-n}\,\varepsilon_{i_n}.$$

5) Construire une suite $(X_n)_n$ de variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$.

B) 6) Soit Q une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Construire une variable aléatoire Y définie sur Ω et de loi $Y(\mathbb{P}) = Q$. On pourra chercher Y $\sigma(X)$–mesurable.

7) Soit $(Q_n)_n$ une suite de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Construire une suite de variables aléatoires $(Y_n)_n$ définies sur Ω, indépendantes et de lois $Y_n(\mathbb{P}) = Q_n$.

C) On revient au cas général $0 < p < 1$. On note $I = [0, 1[$, $\mathcal{I}$ sa tribu borélienne et ℓ la mesure de Borel sur $(I, \mathcal{I})$. Si $x \in I$, on note $0, x_1x_2x_3\ldots$ son développement dyadique qui ne se termine pas par des 1.

8) Exhiber une bijection bimesurable entre $(I, \mathcal{I})$ et $(\Omega_0, \mathcal{F}_0)$ pour un espace Ω_0 de la forme $\Omega_0 = \Omega \setminus N$ avec $\mathbb{P}(N) = 0$, et pour la tribu $\mathcal{F}_0$ induite.

9) En déduire l'existence d'une famille de mesures de probabilités $(m_p)_{p\in]0,1[}$ sur I, étrangères deux à deux, diffuses et telles que $\ell = m_{1/2}$: on admettra une loi des grands nombres pour le jeu de pile ou face, c'est-à-dire que la suite $n^{-1}\displaystyle\sum_{k\leqslant n}\varepsilon_k$ converge $\mathbb{P}$ presque sûrement vers p et on pourra construire m_p sur $(I, \mathcal{I})$ de sorte que la suite $(\alpha_n)_n$ de variables aléatoires définies sur $(I, \mathcal{I})$ par $\alpha_n(x) = x_n$ suive, sous m_p, la même loi que la suite $(\varepsilon_n)_n$ sous P.

Lois

28 Soit X et Y deux variables aléatoires indépendantes de lois de Poisson $\mathcal{P}(a)$ et $\mathcal{P}(b)$. Loi de $X + Y$? Idem pour des lois exponentielles de même paramètre.

29 Le couple (X, Y) admet pour loi $2\,\mathbf{1}_D(x, y)\,\mathrm{d}\ell_2(x, y)$ avec ℓ_2 mesure de Borel sur $\mathbb{R}^2$ et D le triangle de $\mathbb{R}^2$ de sommets $(0, 0)$, $(1, 0)$ et $(0, 1)$. Donner la loi de X ; Y ; $X + Y$; $X - Y$; $Y/(1 - X)$ (solution sans calcul !). Montrer que X et $Y/(1 - X)$ sont indépendantes.

30 X admet une loi diffuse et $Y \perp X$. Alors, $X \neq Y$ presque sûrement.

31 Trouver deux variables aléatoires X et Y non $\perp$ avec $X^2 \perp Y^2$.

32 $\Omega = \{0, 1\}^2$, P uniforme. Trouver trois événements $\perp$ deux à deux mais non dans leur ensemble.

33 $\Omega = \{0, 1, \ldots, n\}$, $P(\{0\}) = p$, $P(\{k\}) = (1 - p)/n$ pour $k \geqslant 1$. Trouver p tel que deux événements quelconques (distincts de $\emptyset$ et Ω) ne sont jamais indépendants.

Simulation

34 (Rejet) On veut simuler une variable aléatoire de loi de densité f donnée par rapport à la mesure de Borel sur $\mathbb{R}$ et on suppose que l'on sait simuler une variable aléatoire de loi de densité g avec $f \leqslant a\,g$ pour $a \geqslant 0$ (et donc obligatoirement $a \geqslant 1$).

On se donne une suite $(U_n)_n$ de variables aléatoires i.i.d. de loi uniforme sur $[0,1]$ et une suite $(Y_n)_n$ de variables aléatoires i.i.d. de loi g. On pose

$$T = \inf\{n \geqslant 1\,;\, f(Y_n) > a\,U_n\,g(Y_n)\}.$$

1) Montrer que T est fini presque sûrement.

2) On pose $X = Y_T$ si T est fini, $X = 0$ sinon. Montrer que X est une variable aléatoire et que sa loi admet pour densité f.

3) Calculer $E(T)$.

35 (Loi normale) 1) Soit X et Y deux variables aléatoires i.i.d. de loi normale centrée réduite $\mathcal{N}(0,1)$. On pose $(X,Y) = (R\cos\Theta, R\sin\Theta)$ pour $R \geqslant 0$ et $\Theta \in \mathbb{R}/2\pi\mathbb{Z}$. Déterminer la loi du couple (R,Θ).

2) Soit U et V deux variables aléatoires i.i.d. de loi uniforme sur $[0,1]$. Construire explicitement à partir de U et V un couple de même loi que (R,Θ).

3) En déduire un procédé de simulation d'une loi normale $\mathcal{N}(0,1)$.

Intégration

36 Soit X une variable aléatoire positive et Y la partie entière de X. Alors,

$$Y = \sum_{n\geqslant 1} \mathbf{1}(X \geqslant n), \quad \sum_{n\geqslant 1} \mathbb{P}(X \geqslant n) \leqslant \mathbb{E}(X) \leqslant 1 + \sum_{n\geqslant 1} \mathbb{P}(X \geqslant n).$$

Montrer $\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > x)\,\mathrm{d}x = \int_0^{+\infty} \mathbb{P}(X \geqslant x)\,\mathrm{d}x$.

37 (Difficile.) Soit X positive avec $\mathbb{E}(X) = a$ et $\mathbb{E}(X^2) = 1$. Alors, pour tout $0 \leqslant t \leqslant 1$,

$$\mathbb{P}(X \geqslant t\,a) \geqslant (1-t)^2\,a^2.$$

38 Soit F une fonction de répartition et a un réel : $\displaystyle\int_{-\infty}^{+\infty} (F(x+a) - F(x))\,\mathrm{d}x = a.$

39 Soit X positive et $r > 1$. Montrer que $\displaystyle\int_0^{+\infty} t^{-r}\,\mathbb{E}(X \wedge t^r)\,\mathrm{d}t = a\,\mathbb{E}(X^b)$ pour deux nombres a et b dépendant de r, dont on précisera la valeur.

40 Exhiber deux variables aléatoires X et Y de même loi uniforme sur $[0,1]$, telles que $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$ et X, Y non indépendantes.

41 (Stone–Weierstrass par les polynômes de Bernstein) On veut montrer que si $f : [0,1] \to \mathbb{R}$ est une fonction continue, on peut l'approcher uniformément par des polynômes. Pour cela, on introduit une suite $(\varepsilon_n)_n$ de variables aléatoires indépendantes de loi de Bernoulli de paramètre p : $\mathbb{P}(\varepsilon_n = 1) = p = 1 - \mathbb{P}(\varepsilon_n = 0)$. On pose comme d'habitude $S_n = \displaystyle\sum_{k=1}^{n} \varepsilon_k$.

1) Montrer que $\mathbb{E}(f(n^{-1}\,S_n))$ est un polynôme en p, que l'on notera $B_n(f)$, et donner une expression de $B_n(f)(p)$.

2) Utiliser le fait que f est bornée et uniformément continue pour montrer que $B_n(f)(p)$ tend vers $f(p)$ uniformément en p quand n tend vers l'infini.

42 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes positives de même loi et intégrables. Soit $M_n = \max(X_k\,;\, k \leqslant n)$. Utiliser l'exercice 36 pour montrer que $\mathbb{E}(M_n/n) \to 0$.

43 (Voir l'exercice 48.) Soit X_n des variables aléatoires indépendantes de même loi. Utiliser l'exercice 36 pour montrer que $\mathbb{E}|X_1|$ est fini si et seulement si

$$\mathbb{P}(|X_n| \geqslant n \text{ une infinité de fois}) = 0.$$

44 (Dyadiques) Soit $\Omega = [0,1]$, $\mathcal{F}$ boréliens et P mesure de Borel. Montrer qu'il existe $A \in \mathcal{F}$ et $B \in \mathcal{F}$ négligeables tels que $\Omega \subset A + B$.

Une somme de nombres normaux est-elle un nombre normal ?

Convergences

45 Soit $(X_n)_n$ une suite de variables aléatoires *quelconque*. Exhiber une suite de réels a_n non nuls telle que $X_n/a_n \to 0$ presque sûrement.

46 1) Soit $X_n : ([0,1], \mathcal{B}([0,1]), \ell) \to \mathbb{R}$ donné par $X_n = n\mathbf{1}_{]0,1/n[}$. Montrer que $X_n \longrightarrow 0$ et $\mathbb{E}(X_n) = 1$. Dessiner $Y = \sup_n X_n$ et préciser si Y est intégrable ou non.

2) Soit $Y_n = \mathbf{1}_{]0,1/n[}$. Ordonnons lexicographiquement l'ensemble des couples (b,a) de $\mathbb{N} \times \mathbb{N}$ tels que $1 \leqslant a \leqslant b$ et posons $Z_n = \mathbf{1}_{](a-1)/b, a/b]}$ si (b,a) est le n–ième couple. Convergence en probabilité, presque sûre et dans les espaces L^p des suites de terme général

$$Y_n, \quad n^2\, Y_n, \quad Z_n, \quad \sqrt{a}\, Z_n, \quad 2^n\, Z_n.$$

47 Soit $f : \mathbb{R} \to \mathbb{R}$ continue et $(X_n)_n$ une suite de variables aléatoires qui converge en probabilité vers X. Montrer que $f(X_n)$ converge en probabilité vers $f(X)$. Préciser ce qui se passe si f n'est pas continue. Préciser ce qui se passe avec la convergence presque sûre.

48 (Voir l'exercice 43.) Soit $(X_n)_n$ une suite de variables aléatoires i.i.d. On suppose que la suite des moyennes S_n/n converge p.s. vers une variable aléatoire Y. Montrer que Y est presque sûrement constante et que X_1 est intégrable (on pourra utiliser l'exercice 36).

Utiliser le même argument pour montrer que si $(X_n)_n$ est une suite de variables aléatoires i.i.d. avec $\mathbb{E}|X_1| = +\infty$, alors $\limsup |S_n|/n = +\infty$ presque sûrement.

49 Soit $\Omega = \{\omega_n\,;\, n \geqslant 1\}$ un ensemble fini ou dénombrable et $P = \sum_n p_n\, \delta_{\omega_n}$ une mesure sur $\mathcal{P}(\Omega)$. Montrer que la convergence P équivaut à la convergence presque sûre (et donc que la convergence L^p implique la convergence presque sûre). Montrer qu'une CNS pour que convergence presque sûre et convergence dans L^p soient équivalentes est que le support de P soit fini.

50 1) Pour $n \geqslant 1$, soit X_n une variable aléatoire de loi

$$P(X_n \leqslant x) = (1 - 1/(x+n))\, \mathbf{1}_{x \geqslant 0}.$$

On suppose que les X_n sont indépendantes. Montrer que X_n tend en probabilité vers 0 (facile) mais que la suite des S_n/n ne converge pas en probabilité (difficile).

2) En déduire si la topologie de la convergence P peut être définie par une norme ou non.

3) Préciser si la suite X_n converge presque sûrement ou dans L^p.

51 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi non presque sûrement constante. Montrer que la suite $(X_n(\omega))_n$ diverge presque sûrement.

52 Exhiber une suite $(X_n)_n$ de variables aléatoires telle que $\limsup_n X_n = +\infty$ presque sûrement mais telle qu'il n'existe pas de sous–suite $(\phi(n))_n$ et d'ensemble $A \in \mathcal{F}$ tels que $\mathbb{P}(A) > 0$ et que $(X_{\phi(n)})_n$ tend vers l'infini sur A.

53 Utiliser une suite de variables aléatoires convergeant P mais pas presque sûrement pour montrer que la topologie de la convergence presque sûre n'est pas métrisable, par exemple sur l'espace $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$.

54 (u.i.) 1) Si $(X_n)_n$ est u.i., la suite $(S_n/n)_n$ est u.i. (Indication : quelle caractérisation de l'u.i. utiliser ?) 2) Application à la LGN.

55 Soit X_n des variables aléatoires de Bernoulli indépendantes avec

$$\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p_n.$$

Trouver une CNS sur $(p_n)_n$ pour avoir $X_n \xrightarrow{P} 0$. Même question pour avoir $X_n \xrightarrow{p.s.} 0$.

56 On se donne n variables aléatoires indépendantes X_k de loi $\mathbb{P}(X_k \geqslant x) = \mathrm{e}^{-ax}$ pour $x \geqslant 0$ et on note $I = \inf\{X_k\,;\, k \leqslant n\}$. Calculer la loi de I. Posons

$$N_t = \mathrm{Card}\,\{k \leqslant n\,;\, X_k \geqslant t\}.$$

Loi, espérance, variance de N_t. Si les X_k modélisent les durées de vie aléatoires de n individus nés au temps $t = 0$, le nombre de survivants au temps t est N_t.

Solutions

10 Si x et y sont les points, on peut faire un triangle si $x \wedge y \leqslant 1/2 \leqslant x \vee y$ et si $x \vee y - x \wedge y \leqslant 1/2$. On dessine la zone favorable dans $[0,1]^2$ et on trouve $p = 1/4$.

17 À faire avec X seulement.

18 Sur $\Omega = \{0,1,2,3\}$, on pose $\mathcal{C}_i = \{\{0,j\}\,;\, j \neq 0,\, j \neq i\}$ pour tout $i \in \Omega$, $i \neq 0$. Si $P(\{i\}) = p$ pour tout $i \neq 0$, et si p est solution de $1 - 3p = (1-2p)^3$, ça marche.

Autre exemple : $\mathcal{C}_3 = \{\emptyset\}$.

20 Les A_{p^2} pour $p \in \mathcal{P}$ sont indépendants donc (2) est la probabilité de l'intersection des $\Omega \setminus A_{p^2}$.
On change d'espace : $\Omega_0 = \{n|D\}$ et $P_0 = \mathbb{P}(\ |\Omega_0)$. Alors, les A_{pn} pour $p \in \mathcal{P}$ sont indépendants sous P_0 donc $P_0(n = D) = 1/\zeta(2s)$.

24 X et Y symétriques diffuses et indépendantes de même loi donne : $\mathbb{P}(X + Y > 0 | X > 0) = 3/4$. Ensuite, $\mathbb{P}(S_{2^n} > 0\,|\,S_1 > 0) = 1/2 + 1/2^{n+1}$. (Non, c'est faux !)

50 Si S_n/n converge en probabilité, une sous–suite $S_{\phi(n)}/\phi(n)$ converge presque sûrement vers S. Quitte à extraire à nouveau une sous–suite on peut supposer que $2\phi(n) \leqslant \phi(n+1)$ et on pose

$$A_n = \{\exists k \in]n/2, n],\, X_k \geqslant c\,k\}.$$

Alors les $A_{\phi(n)}$ sont indépendants et $\mathbb{P}(A_{\phi(n)})$ est minoré par $p_c > 0$ donc BC–non trivial indique que $A_{\phi(n)}$ est réalisé une infinité de fois. Mais si $A_{\phi(n)}$ est réalisé, on a $S_{\phi(n)}/\phi(n) \geqslant c/2$. La constante c est arbitraire donc la suite $S_{\phi(n)}/\phi(n)$ n'est pas bornée.

50 Si Y_n converge en proba vers Y et si a_n tend vers 1, alors $a_n Y_n$ converge en proba vers Y (lemme A). Si Y_n et Z_n convergent en proba vers Y, alors $Y_n - Z_n$ converge en proba vers 0 (lemme B).

52 $X_n = n\varepsilon_n$ avec ε_n Bernoulli 0 ou 1 i.i.d.

Chapitre 2

Lois des grands nombres

La section 2.1 est incontournable. Dans la section 2.2, la proposition 2.3 est facultative. L'exercice 2.6 peut être omis. La suite de résultats portant sur le sens réciproque (convergence presque sûre vers convergence L^2) peut être supprimé : on omettra alors 2.8, 2.9 et le sens direct du corollaire 2.10. En ce cas, le théorème 2.11 des trois séries est réduit à son sens réciproque et, pour le sens direct, à la convergence de s_1 seule.

Dans la section 2.3, la partie 2.3 est indispensable. Dans la partie 2.3, les résultats 2.16, 2.17 et 2.18 peuvent être omis. On en restera alors au résultat 2.15. L'appendice 2.3 est facultatif.

Enfin, la partie 2.4 des applications statistiques peut être réduite aux résultats de 2.4.

Dans tout le chapitre, on se donne une suite $(X_n)_n$ de variables aléatoires et on note $(S_n)_n$ la suite de ses sommes partielles, donc

$$S_n = \sum_{k=1}^{n} X_k.$$

2.1 Rappels sur l'indépendance

On sait que si les X_n sont des variables aléatoires indépendantes et si $\mathcal{F}_n$ désigne la tribu engendrée par X_n, la tribu asymptotique des $\mathcal{F}_n$ est indépendante d'elle-même donc ses éléments sont de probabilité 0 ou 1. En particulier, l'ensemble des ω tels que $X_n(\omega) \geqslant 1$ une infinité de fois ; tels que la suite $(X_n(\omega))_n$ converge ; tels que la suite $(X_n(\omega))_n$ est bornée ; tels que la suite $(S_n(\omega))_n$ converge, sont tous asymptotiques.

Proposition 2.1 *Supposons que les X_n sont des variables aléatoires indépendantes. Alors, la suite S_n converge presque sûrement ou diverge presque sûrement. Soit $(a_n)_n$ une suite de réels $a_n > 0$ qui tend vers l'infini. Alors, S_n/a_n converge presque sûrement vers une variable aléatoire presque sûrement constante ou diverge presque sûrement.*

Démonstration $\{S_n/a_n \text{ converge}\}$ est asymptotique et $L = \lim_n S_n/a_n$ est mesurable pour la tribu asymptotique. ∎

On va donc tout d'abord chercher dans quels cas on peut trouver des constantes a_n telles que X_n/a_n ou S_n/a_n convergent presque sûrement. On sait par ailleurs que l'on peut toujours trouver une suite positive $(a_n)_n$ telle que X_n/a_n converge presque sûrement vers 0. On cherchera donc aussi les « bonnes » suites $(a_n)_n$.

On commence par considérer les diverses convergences possibles de la suite $(S_n)_n$ (convergence des séries) puis on s'attache aux convergences possibles de S_n/n (lois des grands nombres), avant de donner quelques applications statistiques de ces théorèmes limites.

2.2 La convergence des séries indépendantes

Un outil utile permettant de caractériser la convergence presque sûre est le lemme suivant.

Lemme *Soit $(\varepsilon_n)_n$ une suite de réels $\varepsilon_n \geqslant 0$ dont la somme $\sum_n \varepsilon_n$ converge.*

Si $\sum_n P(|X_{n+1} - X_n| \geqslant \varepsilon_n)$ converge, la suite $(X_n)_n$ converge presque sûrement.

Démonstration Par BC trivial, la suite $(X_n(\omega))_n$ est presque sûrement une suite de Cauchy. ■

On peut retrouver directement grâce à ce lemme le fait que si $(X_n)_n$ converge P, une sous–suite $(X_{\phi(n)})_n$ converge presque sûrement.

Des critères de Cauchy pour les convergences P et presque sûre des suites

Proposition 2.2 *Pour $n \geqslant 1$, notons*

$$Y_n = \sup_{p,q \geqslant n} |X_p - X_q|, \qquad Z_n = \sup_{p \geqslant n} |X_p - X_n|.$$

Alors, $(X_n)_n$ converge presque sûrement si et seulement si $Y_n \xrightarrow{P} 0$ si et seulement si $Z_n \xrightarrow{P} 0$.

Démonstration Clairement, $Z_n \leqslant Y_n \leqslant 2Z_n$. On traite donc seulement le cas de Y_n. Or Y_n est une suite décroissante de fonctions positives donc Y_n converge presque sûrement. Ainsi, $Y_n \xrightarrow{P} 0$ si et seulement si $Y_n \xrightarrow{p.s.} 0$ si et seulement siX_n est presque sûrement une suite de Cauchy si et seulement si X_n converge presque sûrement. ■

Proposition 2.3 *Pour $\varepsilon > 0$ et $n \geqslant 1$, notons*

$$a_n(\varepsilon) = \sup_{p,q \geqslant n} P(|X_p - X_q| \geqslant \varepsilon), \qquad b_n(\varepsilon) = \sup_{p \geqslant n} P(|X_p - X_n| \geqslant \varepsilon).$$

Alors, $(X_n)_n$ converge P si et seulement si $a_n(\varepsilon) \to 0$ pour tout $\varepsilon > 0$ si et seulement si $b_n(\varepsilon) \to 0$ pour tout $\varepsilon > 0$.

Démonstration Clairement, $b_n(\varepsilon) \leqslant a_n(\varepsilon) \leqslant b_n(\varepsilon/2)$ donc on traite le cas de $a_n(\varepsilon)$.

Si $X_n \xrightarrow{P} X$, alors $a_n(\varepsilon) \leqslant 2 \sup_{p \geqslant n} \mathbb{P}(|X_p - X| \geqslant \varepsilon/2)$ donc $a_n(\varepsilon) \to 0$.

Réciproquement, si $a_n(\varepsilon) \to 0$ pour tout $\varepsilon > 0$, pour chaque $k \geqslant 1$, on peut exhiber un indice n_k tel que

$$a_{n_k}(2^{-k}) \leqslant 2^{-k}.$$

Posons $D_k = X_{n_{k+1}} - X_{n_k}$. Alors $\mathbb{P}(D_k \geqslant 2^{-k}) \leqslant 2^{-k}$ donc $\sum_k D_k$ converge presque sûrement vers une limite X. Soit $\varepsilon > 0$ et $k \geqslant 1$ tels que $2^{-k} \leqslant \varepsilon/2$. Pour tout $n \geqslant n_k$, on peut majorer $\mathbb{P}(|X_n - X| \geqslant \varepsilon)$ comme suit :

$$\begin{aligned}\mathbb{P}(|X_n - X| \geqslant \varepsilon) &\leqslant \mathbb{P}(|X_n - X_{n_k}| \geqslant \varepsilon/2) + \mathbb{P}(|X_{n_k} - X| \geqslant \varepsilon/2) \\ &\leqslant a_{n_k}(\varepsilon/2) + \sum_{i \geqslant k} \mathbb{P}(D_i \geqslant 2^{-i}) \leqslant a_{n_k}(2^{-k}) + \sum_{i \geqslant k} 2^{-i} \leqslant 3 \cdot 2^{-k},\end{aligned}$$

d'où la conclusion. ∎

Comparaison des convergences P et presque sûre des séries

Le théorème 2.4 suivant, dû à Lévy, devrait constituer une surprise.

Théorème 2.4 (P. Lévy) *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. Alors, $(S_n)_n$ converge P si et seulement si $(S_n)_n$ converge presque sûrement.*

Démonstration On suppose que $S_n \xrightarrow{P} S$ et on veut montrer la convergence presque sûre. D'après la proposition 2.2, il suffit de montrer que $Z_n = \sup_{p \geqslant n} |S_p - S_n|$ tend P vers 0. Notons

$$\tau_n = \inf\{p \geqslant n\,;\, |S_p - S_n| \geqslant 2\varepsilon)$$

avec la convention $\inf \emptyset = +\infty$. On veut donc montrer que $\mathbb{P}(\tau_n < +\infty) \to 0$ quand n tend vers l'infini.

On sait déjà que $S_n \xrightarrow{P} S$ donc $\mathbb{P}(|S_n - S| \geqslant \varepsilon/2) \to 0$. Mais si $m \geqslant n$ et si $|S_n - S_m| \geqslant \varepsilon$, l'un des deux termes $|S_n - S|$ ou $|S_m - S|$ est supérieur à $\varepsilon/2$. Donc $\mathbb{P}(|S_n - S_m| \geqslant \varepsilon) \to 0$ quand n et m tendent vers l'infini. Fixons $\alpha > 0$ quelconque et notons n_α un indice tel que tous les $m \geqslant n \geqslant n_\alpha$ vérifient

$$\alpha \geqslant \mathbb{P}(|S_n - S_m| \geqslant \varepsilon).$$

Décomposons l'événement $\{|S_n - S_m| \geqslant \varepsilon\}$ selon les valeurs de τ_n. Pour p compris entre $n+1$ et m, le fait que $\tau_n = p$ et que $|S_m - S_p| < \varepsilon$ implique que $|S_m - S_n| \geqslant \varepsilon$. Il vient

$$\alpha \geqslant \sum_{p=n+1}^{m} \mathbb{P}(\tau_n = p\,;\, |S_m - S_p| < \varepsilon).$$

Mais $\{\tau_n = p\}$ est mesurable par rapport à $(X_k)_{k \leqslant p}$ et $\{|S_m - S_p| < \varepsilon\}$ est mesurable par rapport à $(X_k)_{k \geqslant p+1}$ donc ces deux événements sont indépendants. De plus, la probabilité de $\{|S_m - S_p| \geqslant \varepsilon\}$ est toujours inférieure à α donc la probabilité du complémentaire vaut au moins $(1-\alpha)$, donc

$$\alpha \geqslant (1-\alpha) \sum_{p=n+1}^{m} \mathbb{P}(\tau_n = p) = (1-\alpha)\, \mathbb{P}(\tau_n \leqslant m).$$

En faisant tendre m vers l'infini, on a montré que $\mathbb{P}(\tau_n < +\infty) \leqslant \alpha/(1-\alpha)$ pour tout $n \geqslant n_\alpha$, ce qui conclut la démonstration. ∎

Comparaison des convergences L^2 et presque sûre pour les séries

Quand les variables aléatoires X_n sont de carré intégrable, on peut relier leur distribution à leur moment d'ordre 2. Dans le cas indépendant, ce contrôle est particulièrement performant grâce au

Lemme 2.5 (inégalité de Kolmogorov) *Soient X_1, X_2, …, X_n des variables aléatoires indépendantes centrées de carré intégrable. Alors,*

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| \geqslant a) \leqslant \mathbb{E}(S_n^2)/a^2.$$

Remarque Amélioration (brutale quand n est grand) de Chebyshev.

Démonstration Comme dans la preuve de 2.4, on introduit le plus petit instant $\tau = k$ pour lequel $|S_k| \geqslant a$ avec la convention $\inf \emptyset = +\infty$. Il vient :

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| \geqslant a) = \mathbb{P}(\tau \leqslant n) = \sum_{k=1}^{n} \mathbb{P}(\tau = k).$$

À présent, on imite Chebyshev :

$$\mathbb{E}(S_n^2) \geqslant \sum_{k=1}^{n} \mathbb{E}(S_n^2\,\mathbf{1}(\tau = k)) \geqslant \sum_{k=1}^{n} \mathbb{E}([S_k^2 + 2S_k(S_n - S_k)]\,\mathbf{1}(\tau = k)).$$

L'événement $\{\tau = k\}$ et la variable aléatoire S_k dépendent des k premiers X_i, la variable aléatoire $S_n - S_k$ des autres X_i ; de plus $S_n - S_k$ est centrée donc ces termes disparaissent et par définition de τ, sur $\{\tau = k\}$ on a $S_k^2 \geqslant a^2$. Il reste

$$\mathbb{E}(S_n^2) \geqslant \sum_{k=1}^{n} \mathbb{E}(S_k^2\,\mathbf{1}(\tau = k)) \geqslant \sum_{k=1}^{n} \mathbb{E}(a^2\,\mathbf{1}(\tau = k)) = a^2\,\mathbb{P}(\tau \leqslant n).$$

∎

Exercice 2.6 Montrer que pour toute fonction ϕ convexe positive et croissante,

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} S_k \geqslant a) \leqslant \mathbb{E}\phi(S_n)/\phi(a).$$

∎

Le lemme 2.5 permet d'obtenir immédiatement la comparaison suivante.

Théorème 2.7 *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes centrées de carré intégrable.*

(i) La suite $(S_n)_n$ converge dans L^2 si et seulement si $\Sigma = \sum_n E(X_n^2)$ converge.

(ii) Sous cette condition, la suite $(S_n)_n$ converge aussi presque sûrement vers S_∞ et $\mathbb{E}(S_\infty^2) = \Sigma$.

Remarque On l'appliquera à des variables aléatoires non centrées. La condition portera alors sur les variances $\mathrm{var}(X_n)$.

Démonstration En temps fini, on peut appliquer le lemme 2.5, donc

$$\mathbb{P}(\max_{1\leqslant k\leqslant m} |S_{n+k} - S_n| \geqslant \varepsilon) \leqslant \varepsilon^{-2}\,\mathbb{E}(S_{n+m} - S_n)^2 \leqslant \varepsilon^{-2} \sum_{k\geqslant n} \mathbb{E}(X_k^2).$$

Par limite croissante, la même majoration s'applique à

$$\mathbb{P}(\sup_{k\geqslant n} |S_k - S_n| \geqslant \varepsilon),$$

qui tend donc vers 0 car $\sum_{k\geqslant n} \mathbb{E}(X_k^2)$ est le reste d'une série convergente. On finit par la proposition 2.2. ■

On va maintenant démontrer une réciproque, valable uniquement pour des variables aléatoires bornées.

Lemme 2.8 (inégalité de Kolmogorov inverse) *Soit $(X_k)_{1\leqslant k\leqslant n}$ des variables aléatoires indépendantes centrées uniformément bornées par c. Alors*

$$\mathbb{P}(\max_{1\leqslant k\leqslant n} |S_k| \geqslant a) \geqslant 1 - (a+c)^2/\mathbb{E}(S_n^2).$$

Démonstration La variable aléatoire $S_n - S_k$ est centrée donc

$$\mathbb{E}(S_n^2\,\mathbf{1}(\tau \leqslant n)) = \sum_{k=1}^n \mathbb{E}(S_n^2\,\mathbf{1}(\tau = k)) = \sum_{k=1}^n \mathbb{E}(S_k^2\,\mathbf{1}(\tau = k)) + \mathbb{E}((S_n - S_k)^2\,\mathbf{1}(\tau = k)).$$

Sur $\{\tau = k\}$, on sait que $|S_{k-1}| < a$ et $|X_k| \leqslant c$ donc $|S_k| \leqslant (a+c)$. Par ailleurs, la variable aléatoire $S_n - S_k$ et l'événement $\{\tau = k\}$ sont indépendants. Il vient

$$\begin{aligned}\mathbb{E}(S_n^2\,\mathbf{1}(\tau \leqslant n)) &\leqslant (a+c)^2 \sum_{k=1}^n \mathbb{P}(\tau = k) + \sum_{k=1}^n \mathbb{P}(\tau = k) \sum_{i=k+1}^n \mathbb{E}(X_i^2)\\ &\leqslant (a+c)^2\,\mathbb{P}(\tau \leqslant n) + \mathbb{P}(\tau \leqslant n)\,\mathbb{E}(S_n^2).\end{aligned}$$

À présent, sur le complémentaire de $\{\tau \leqslant n\}$, on sait que $|S_n| < a$ donc

$$\mathbb{E}(S_n^2\,\mathbf{1}(\tau \leqslant n)) = \mathbb{E}(S_n^2) - \mathbb{E}(S_n^2\,\mathbf{1}(\tau > n)) \geqslant \mathbb{E}(S_n^2) - a^2\,(1 - \mathbb{P}(\tau \leqslant n)).$$

Après simplifications, il vient

$$\mathbb{P}(\tau \leqslant n) \geqslant 1 - (a+c)^2/(\mathbb{E}(S_n^2) + (a+c)^2 - a^2) \geqslant 1 - (a+c)^2/\mathbb{E}(S_n^2).$$

■

Théorème 2.9 *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes centrées uniformément bornées (donc de carré intégrable). Si S_n converge presque sûrement, alors $\sum_n \mathbb{E}(X_n^2)$ converge.*

Démonstration Pour tous $n \geqslant 1$ et $m \geqslant 1$, le lemme donne

$$\mathbb{P}(\max_{1\leqslant k\leqslant m} |S_{n+k} - S_n| \geqslant 1) \geqslant 1 - (c+1)^2/\sum_{k=n+1}^{n+m} \mathbb{E}(X_k^2).$$

Par convergence presque sûre, on sait qu'il existe un indice n tel que

$$\mathbb{P}(\sup_{k\geqslant n}|S_k - S_n| \geqslant 1) \leqslant 1/2.$$

On en déduit

$$1/2 \geqslant 1-(c+1)^2/\sum_{k\geqslant n+1}\mathbb{E}(X_k^2),$$

soit une majoration de $\sum_{k\geqslant n+1}\mathbb{E}(X_k^2)$ par un nombre fini. ■

Corollaire 2.10 (sommes de signes) *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes centrées de même loi (non identiquement nulle) et bornées et soit $(a_n)_n$ une suite réelle bornée. Alors, $\sum_n a_n X_n$ converge presque sûrement si et seulement si la série $\sum_n a_n^2$ converge.*

Application aux signes : $X_n = \pm 1$ avec probabilité 1/2. Alors $\sum_n X_n/n^\alpha$ converge presque sûrement si et seulement si $\alpha > 1/2$.

Un critère de convergence presque sûre

Théorème 2.11 (critère des trois séries) *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. Pour $n \geqslant 1$, on pose*

$$S_n = \sum_{k=1}^n X_k, \qquad Y_n = X_n\,\mathbf{1}(|X_n|\leqslant 1).$$

Alors S_n converge presque sûrement si et seulement si les trois séries numériques suivantes convergent :

$$s_1 = \sum_n \mathbb{P}(|X_n|>1), \quad s_2 = \sum_n \mathbb{E}(Y_n), \quad s_3 = \sum_n \mathrm{var}(Y_n).$$

Démonstration Si les trois séries convergent, s_3 alliée au théorème 2.7 montre que $\sum_n Y_n - E(Y_n)$ converge presque sûrement. De plus, s_2 converge donc $\sum_n Y_n$ converge presque sûrement. Enfin, s_1 converge donc BC–trivial donne $\mathbb{P}(\limsup\{X_n \neq Y_n\}) = 0$ donc S_n converge presque sûrement.

Si S_n converge presque sûrement, X_n tend vers 0 presque sûrement donc

$$\mathbb{P}(\limsup\{X_n > 1\}) = 0.$$

Par BC–non trivial, s_1 converge donc, or $s_1 = \sum_n \mathbb{P}(X_n \neq Y_n)$, donc BC–trivial montre que $\sum_n Y_n$ converge.

Supposons qu'il existe une suite $(Y'_n)_n$ indépendante de la suite $(Y_n)_n$ et de même loi. Alors $\sum_n Y'_n$ converge presque sûrement, ainsi que $\sum_n Z_n$ pour $Z_n = Y_n - Y'_n$. L'idée–clé de cette construction est que la suite $(Z_n)_n$ est formée de variables aléatoires indépendantes centrées et bornées uniformément puisque la loi de Z_n est symétrique et que $|Z_n| \leqslant 2$ presque sûrement. Par le théorème 2.9, la somme $\sum \mathbb{E}(Z_n^2)$ doit converger. Or, $\mathbb{E}(Z_n^2) = 2\,\mathrm{var}(Y_n)$ donc s_3 converge. À présent, s_3 converge donc le théorème 2.7 appliqué aux variables aléatoires centrées $Y_n - \mathbb{E}(Y_n)$ montre que $\sum_n Y_n - E(Y_n)$ converge presque sûrement. Ainsi, s_2 converge.

Il reste à justifier la construction de la suite $(Y'_n)_n$. Pour cela, il suffit de se donner une suite $(X'_n)_n$ indépendante de la suite $(X_n)_n$ et de même loi et à poser $Y'_n = X'_n\,\mathbf{1}(|X'_n| \leqslant 1)$. Or l'existence de la suite $(X'_n)_n$ découle d'une construction tout à fait générale.

En effet, soit $(X_n)_n$ une suite de variables aléatoires sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Alors il existe un espace de probabilité $(\Psi, \mathcal{G}, \mathbb{Q})$ et deux suites de variables aléatoires $(\xi_n)_n$ et $(\xi'_n)_n$ sur Ψ, chacune de même loi sous $\mathbb{Q}$ que $(X_n)_n$ sous $\mathbb{P}$, et indépendantes l'une de l'autre. Pour le montrer, on peut choisir $\Psi = \Omega \times \Omega$, $\mathcal{G} = \mathcal{F} \otimes \mathcal{F}$ et définir la mesure $\mathbb{Q}$ de manière unique en imposant que $\mathbb{Q}(A \times A') = \mathbb{P}(A) \cdot \mathbb{P}(A')$ pour tous A et A' dans $\mathcal{F}$. Il reste à poser $\xi_n(\omega, \omega') = X_n(\omega)$ et $\xi'_n(\omega, \omega') = X_n(\omega')$, pour tout indice n et tout ω dans Ω. ■

2.3 Lois des grands nombres

Le cas L^2

Lemme 2.12 (Kronecker) *Soit $\sum_n x_n$ une série convergente à valeurs dans un espace vectoriel normé et soit $(a_n)_n$ une suite réelle positive croissante tendant vers l'infini. Alors*

$$\lim a_n^{-1} \sum_{k=1}^{n} a_k\, x_k = 0.$$

Démonstration Par transformation d'Abel, on sait que

$$a_n^{-1} \sum_{k=1}^{n} a_k\, x_k = s_n - \sum_{k=0}^{n-1} a_n^{-1}\,(a_{k+1} - a_k)\, s_k,$$

où s_k est la somme des k premiers x_i à condition de poser $s_0 = a_0 = 0$. Césaro ou une démonstration directe permet de montrer que le deuxième terme tend vers $s = \lim_n s_n$ car la somme des poids vaut 1. ■

On en déduit facilement le

Théorème 2.13 *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de carré intégrable et $(a_n)_n$ une suite réelle positive croissante tendant vers l'infini. Alors la suite $a_n^{-1}\,(S_n - E(S_n)) \xrightarrow{p.s.} 0$ et dans L^2 dès que la série $\sum_n \mathrm{var}(X_n)/a_n^2$ converge.*

Corollaire *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de carré intégrable et centrées. On suppose que $E(X_n^2) \leqslant c$ pour tout n. Alors $S_n/n^\alpha \xrightarrow{p.s.} 0$ et dans L^2 pour tout $\alpha > 1/2$.*
En particulier si les X_n sont des variables aléatoires indépendantes de même loi de carré intégrable, la suite $S_n/n \xrightarrow{p.s.} E(X_1)$ et dans L^2.

Application : nombres normaux.
Exemple : $0,12345678910111213141516\ldots$ est normal en base 10.
Pas de nombre normal dans toutes les bases connu.

Exercice
Soit $(X_n)_n$ une suite de variables aléatoires de carré intégrable telle que $E(X_n^2) \leqslant c$ pour tout n. On suppose que $S_{2^n}/2^n$ converge presque sûrement et que S_n/n converge P. Montrer que S_n/n converge presque sûrement.

La LGN L^1

Un outil essentiel introduit par Kolmogorov (encore lui !) pour passer d'un cas L^2 (en fait, L^∞) à un cas L^1 consiste à tronquer les variables aléatoires. Si $(X_n)_n$ est une suite de variables aléatoires, on note

$$Y_n = X_n \mathbf{1}(|X_n| \leqslant n) \quad \text{et} \quad T_n = \sum_{k=1}^{n} Y_k.$$

Lemme 2.14 (troncature) *Soit $(X_n)_n$ une suite de variables aléatoires de même loi intégrable. On a alors les propriétés suivantes :*
(i) $E(Y_n)$ tend vers $E(X_1)$ et $E(T_n/n)$ aussi ;
(ii) les séries $\sum_n n^{-2} \operatorname{var}(Y_n)$ et $\sum_n n^{-2} E(Y_n^2)$ convergent ;
(iii) $P(\liminf\{Y_n = X_n\}) = 1$.

Démonstration (i) On sait que $E(Y_n) = E(X_1 \,;\, |X_1| \leqslant n)$; puis Césaro.
(ii) On considère $E(Y_n^2) = E(X_1^2 \,;\, |X_1| \leqslant n)$ donc

$$\begin{aligned}
\sum_n E(Y_n^2)/n^2 &= \sum_n n^{-2} \sum_{m \leqslant n} E(X_1^2 \,;\, m-1 < |X_1| \leqslant m) \\
&\leqslant \sum_m mE(|X_1| \,;\, m-1 < |X_1| \leqslant m) \sum_{n \geqslant m} n^{-2} \\
&\leqslant 2 \sum_m E(|X_1| \,;\, m-1 < |X_1| \leqslant m) = 2E|X_1|,
\end{aligned}$$

en majorant X_1^2 par $m\,|X_1|$ quand on le peut et le reste $\sum_{n \geqslant m} n^{-2}$ par $2/m$.
(iii) BC–trivial sur la suite $\{X_n \neq Y_n\} = \{|X_n| > n\}$. ∎

On en déduit aussitôt une forme intermédiaire de LGN.

Théorème 2.15 *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi intégrable. La suite S_n/n converge presque sûrement vers $E(X_1)$.*

Démonstration Le lemme 2.14 (ii) montre que l'on peut appliquer le théorème 2.13 à la suite des Y : ainsi, $n^{-1}(T_n - E(T_n))$ tend vers 0 presque sûrement. Le (i) du lemme montre que $n^{-1} E(T_n)$ tend vers $E(X_1)$ et le (iii) que la suite X_n coïncide avec la suite Y_n à partir d'un certain rang. ∎

Corollaire *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi intégrable. Si $E|X_n|$ est infini, S_n/n diverge presque sûrement.*
Si EX_1^+ est infini et EX_1^- fini, alors S_n/n tend presque sûrement vers $+\infty$.

Exercice Trouver un exemple avec $\limsup S_n/n = +\infty$ et $\liminf S_n/n = -\infty$.

Rappelons un outil essentiel reliant la convergence p.s. et la convergence L^1 des variables aléatoires positives.

Lemme 2.16 (Scheffé)
Soit X_n et X des variables aléatoires intégrables telles que $X_n \xrightarrow{p.s.} X$ et $E|X_n| \to E|X|$. Alors $X_n \xrightarrow{L^1} X$ c'est-à-dire que $E|X_n - X|$ tend vers 0.

Exercice Montrer que la convergence du théorème 2.15 a lieu dans L^1.
Premier pas : $X_n = X_n^+ + X_n^-$. Ensuite, Scheffé.

La version définitive (pour ce cours) de la LGN est donnée par le

Théorème 2.17 *Soit $(X_n)_n$ une suite de variables aléatoires de même loi intégrable. Supposons que les variables aléatoires X_n sont deux à deux indépendantes.*
Alors $S_n/n \xrightarrow{p.s.} E(X_1)$ et dans L^1.

Démonstration (1) En décomposant en parties positives et négatives, il suffit de démontrer la convergence presque sûre pour des variables aléatoires positives. La convergence L^1 se déduira de Scheffé puisque $E(S_n/n) = E(X_1)$.

On va tronquer les X_n pour obtenir une convergence rapide d'une sous–suite de S_n/n puis utiliser la monotonie de la suite S_n.

(2) Soit $a > 1$ et a_n la partie entière de a^n. Posons

$$Z_n = a_n^{-1}(T_{a_n} - E(T_{a_n})).$$

Alors la série $\sum_n E(Z_n^2)$ converge. En effet, par indépendance deux à deux des Y, on a

$$E(Z_n^2) = a_n^{-2} \sum_{k=1}^{a_n} \mathrm{var}(Y_k) \leqslant a_n^{-2} \sum_{k=1}^{a_n} E(Y_k^2),$$

d'où par sommation,

$$\sum_n E(Z_n^2) \leqslant \sum_k E(Y_k^2) \sum_{n\,:\,a_n \geqslant k} a_n^{-2} \leqslant c \sum_k E(Y_k^2)/k^2$$

qui est fini par le lemme 2.14 (ii). On en déduit que $\sum_n Z_n^2$ est presque sûrement finie donc que Z_n tend presque sûrement vers 0.

(3) Par le lemme 2.14 (i), on sait que $a_n^{-1}E(T_{a_n})$ tend vers $E(X_1)$. Donc $a_n^{-1}\, T_{a_n} \xrightarrow{p.s.} E(X_1)$. On en déduit que $a_n^{-1}\, S_{a_n} \xrightarrow{p.s.} E(X_1)$ car $(S_{a_n} - T_{a_n})/a_n$ tend presque sûrement vers 0 d'après le lemme 2.14 (iii).

(4) Il reste à repasser de la suite S_{a_n}/a_n à la suite S_n/n. Pour $n \geqslant 1$, soit $k(n)$ l'indice k tel que $a_k \leqslant n < a_{k+1}$. Alors,

$$S_n/n \leqslant (a_{k(n)}/a_{k(n)+1})\, S_{a_{k(n)+1}}/a_{k(n)+1}$$

d'où $\limsup S_n/n \leqslant \limsup(a_{k(n)}/a_{k(n)+1}) \cdot E(X_1) = a\, E(X_1)$. De même,

$$\liminf S_n/n \geqslant E(X_1)/a.$$

Le réel $a > 1$ est arbitraire donc on a terminé. ∎

On a utilisé les résultats auxiliaires suivants laissés en exercice au lecteur :

Lemme 2.18 *Soit $a > 1$ et a_n la partie entière de a^n. Alors a_{n+1}/a_n tend vers a.*
Il existe une constante c indépendante de k (mais dépendant de a) telle que

$$\sum_{n\,:\,a_n \geqslant k} 1/a_n^2 \leqslant c/k^2.$$

Exercice Utiliser la notion d'uniforme intégrabilité et l'exercice 2 pour montrer que l'on peut déduire la convergence L^1 de la convergence p.s. dans les LGN qui précèdent.

Appendice : la loi des grands nombres pour une suite échangeable

D'après L. Pratelli, Séminaire XXIII, Lecture Notes 1372, 527–530 (1989).
Soit $(X_n)_{n\geqslant 1}$ une suite échangeable de variables aléatoires et notons

$$S_n = \sum_{k=1}^{n} X_k, \qquad U_n = S_n/n.$$

Rappelons qu'une famille est échangeable si sa loi est invariante sous l'effet de toute permutation de support fini c'est-à-dire si pour $n \geqslant 1$ et s permutation de $\{1, \dots, n\}$ quelconques, $X^s = (X_{s(k)})_{k\leqslant n}$ admet la même loi que $(X_k)_{k\leqslant n}$. Quelques propriétés :

– échangeable implique de même loi ;
– i.i.d. implique échangeable ;
– si $(X_n)_n$ est échangeable, f est borélienne et Z est indépendante de $(X_n)_n$, alors la suite des $Y_n = f(X_n, Z)$ est échangeable ;
– si $(X_n)_n$ est échangeable de carré intégrable, $E(X_n) = E(X_1)$, $\mathrm{var}(X_n) = \mathrm{var}(X_1)$. De plus, on a par exemple $E(X_n\, X_m) = E(X_1\, X_2)$ pour tout $n \neq m$.

Théorème *Soit $(X_n)_n$ une suite échangeable. Notons $\mathcal{A}$ la tribu asymptotique des X_n et supposons que X_1 est intégrable. Alors la suite $(U_n)_n$ converge presque sûrement et dans L^1 vers une variable aléatoire U $\mathcal{A}$–mesurable. On peut caractériser U comme suit : toute variable aléatoire Z $\mathcal{A}$–mesurable et bornée vérifie*

$$E(Z\, X_1) = E(Z\, U).$$

On verra plus tard que U est l'espérance conditionnelle de X_1 par rapport à $\mathcal{A}$.

Corollaire *Dans le cas i.i.d., la tribu $\mathcal{A}$ est triviale donc U est constante presque sûrement et $U = E(X_1)$.*

Démonstration La preuve se fait en plusieurs étapes.

1. Soit f une fonction borélienne bornée sur $\mathbb{R}$ et posons

$$Y_n = f(X_n), \qquad T_n = \sum_{k=1}^{n} Y_k, \qquad V_n = T_n/n.$$

   On a alors $E|U_n - V_n| \leqslant E|X_1 - Y_1|$. (Inégalité triangulaire.)

2. Supposons que X_1 est de carré intégrable et notons $a = E(X_1^2)$ et $b = E(X_1\, X_2)$. Alors, pour $m \leqslant n$, on peut calculer

$$E(U_n\, U_m) = (a + (n-1)\, b)/n = E(U_n^2),$$

   donc $E((U_m - U_n)^2) = E(U_m^2) - E(U_n^2)$.
   Par conséquent, la suite des $E(U_n^2)$ est décroissante et la suite U_n est de Cauchy dans L^2.

3. Supposons X_1 intégrable. Alors, pour $\varepsilon > 0$, on a :

$$\varepsilon\, P(\max_{m\leqslant i\leqslant n} |U_n - U_i| \geqslant \varepsilon) \leqslant E|U_n - U_m|.$$

   Pour le démontrer, on pose

$$\tau = \sup\{i\, ;\, m \leqslant i < n,\, |U_n - U_i| \geqslant \varepsilon\},$$

   et on évalue $\varepsilon\, P(\tau = i)$ en séparant la partie où $U_i > U_n$ et celle où $U_i \leqslant U_n$. Il vient

$$\varepsilon\, P(\tau \geqslant m) \leqslant E(U_n - U_m : A_-) + E(U_m - U_n : A_+),$$

   où A_- et A_+ sont deux ensembles disjoints ; le tout est donc majoré par $E|U_n - U_m|$.

4. On déduit de ce qui précède que la suite $(U_n)_n$ est de Cauchy dans L^1 : on se ramène au cas L^2 grâce à 1 et on applique alors le 2. Par conséquent, le 3 montre que la suite $(U_n)_n$ converge presque sûrement.
5. Posons alors $U = \limsup U_n$. La variable aléatoire U est bien $\mathcal{A}$–mesurable et, si Z est $\mathcal{A}$–mesurable bornée, on a $E(U_n\, Z) = E(X_1\, Z)$ pour tout n, d'où le résultat.

■

Remarque La variable aléatoire U est la seule telle que $E(U : A) = E(X_1 : A)$ pour tout $A \in \mathcal{A}$.

2.4 Applications statistiques

Moyenne, variance, méthode de Monte–Carlo

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi de carré intégrable. On note

$$M_n = S_n/n, \qquad V_n = (n-1)^{-1} \sum_{k=1}^{n} (X_k - M_n)^2.$$

En conséquence directe de la LGN, M_n est un estimateur consistant de $E(X_1)$ et V_n de $\mathrm{var}(X_1)$ au sens où l'on a :

$$\begin{array}{lll} M_n \xrightarrow{p.s.} E(X_1) & \text{avec} & E(M_n) = E(X_1), \\ V_n \xrightarrow{p.s.} \mathrm{var}(X_1) & \text{avec} & E(V_n) = \mathrm{var}(X_1). \end{array}$$

Laissé en exercice.

Voici une méthode robuste pour calculer des intégrales, conséquence directe de la LGN : soit X_n des variables aléatoires indépendantes de même loi uniforme sur un borélien B de $\mathbb{R}^d$ et $f : B \to \mathbb{R}$ mesurable bornée. Alors,

$$(1/n) \sum_{k=1}^{n} f(X_k) \longrightarrow (1/\mathrm{mes}(B)) \int_B f(x)\, dx.$$

En particulier, si les X_n sont uniforme sur $[0,1]^d$ et si A est un borélien de $[0,1]^d$, on peut calculer la mesure de A par

$$(1/n) \sum_{k=1}^{n} \mathbf{1}(X_k \in A).$$

Exemple Quart de cercle centré en $(0,0)$ de rayon 1, $(X_n)_n$ i.i.d. uniformes sur $[0,1]$.
Alors $n^{-1} \sum_{k \leqslant n} \sqrt{1 - X_k^2}$ tend presque sûrement vers $\pi/4$. La variance est décrite par les différences d'ordonnées de points de la courbe donc on améliore la convergence en regardant

$$(1/2n) \sum_{k \leqslant n} \sqrt{1 - X_k^2} + \sqrt{1 - (1 - X_k)^2}.$$

Ordre et rang

Soit $(X_k)_{k \leqslant n}$ des variables aléatoires indépendantes de même loi μ *continue*. Notons X le vecteur $X = (X_1, \ldots, X_n)$ et

$$s(k) = \sum_{l=1}^{n} \mathbf{1}(X_l \leqslant X_k).$$

Théorème *Presque sûrement, s est une permutation de $\{1, 2, \dots, n\}$. De plus, si on introduit l'échantillon ordonné*

$$X^* = (X_{s^{-1}(1)}, X_{s^{-1}(2)}, \dots, X_{s^{-1}(n)}),$$

alors s et X^* sont indépendants, la loi de s est uniforme et la loi de X^* est la trace de $n!\, \mu^{\otimes n}$ sur

$$I = \{x \in \mathbb{R}^n \,;\, x_1 < x_2 < \dots < x_n\}.$$

Lemme *Soit x et y deux variables aléatoires indépendantes avec x de loi continue. Alors $P(x = y) = 0$.*

Démonstration Pour la preuve du théorème, soit t une permutation. Alors

$$X^t = (X_{t(1)}, X_{t(2)}, \dots, X_{t(n)})$$

suit la loi de X donc

$$P(s(X) = r) = P(s(X^t) = r) = P(s(X) \circ t = r) = P(s(X) = r \circ t^{-1}).$$

On a montré que $P(s(X) = r) = 1/n!$ pour toute permutation r. Soit $A \subset I$ un borélien. Alors,

$$P(X^* \in A,\, s(X) = r) = P(X^{r^{-1}} \in A) = P(X \in A) = n!\, \mu^{\otimes n}(A)\, (1/n!),$$

donc X^* et $s(X)$ sont indépendantes et

$$P(X^* \in A) = \sum_r P(X^* \in A,\, s(X) = r) = n!\, \mu^{\otimes n}(A).$$

■

Le théorème fondamental de la statistique

Définition 2.19 *La fonction de répartition empirique du tirage $(X_k)_{k \geqslant 1}$ au temps $n \geqslant 1$ (ou bien du tirage $(X_k)_{1 \leqslant k \leqslant n}$) est la variable aléatoire*

$$F_n(x) = n^{-1} \sum_{k=1}^{n} \mathbf{1}(X_k \leqslant x).$$

La LGN donne, pour un x réel fixé, la convergence presque sûre de la variable aléatoire $F_n(x)$ vers le réel $F(x)$. En effet, $F_n(x)$ est une moyenne de variables aléatoires indépendantes bornées et de même loi telles que

$$E(\mathbf{1}(X_k \leqslant x)) = P(X_k \leqslant x) = F(x).$$

En fait, on a une convergence beaucoup plus forte, et cruciale dans les applications.

Théorème 2.20 (Glivenko–Cantelli) *Si la suite $(X_n)_n$ est une suite de tirages indépendants de loi X et si F est la fonction de répartition de X, la suite de variables aléatoires*

$$\sup\{|F_n(x) - F(x)| \,;\, x \in \mathbb{R}\}$$

tend presque sûrement vers 0.

Démonstration Soit S l'ensemble des sauts de F et Q un ensemble dénombrable dense dans $\mathbb{R}$ et disjoint de S. Par la loi des grands nombres appliquée aux $\mathbf{1}(X_n \leqslant x)$, si $x \in Q$, $F_n(x)$ tend vers $F(x)$ sur $\Omega \setminus N(x)$. Par la loi des grands nombres appliquée aux $\mathbf{1}(X_n = x)$, si $x \in S$, $F_n(x) - F_n(x-)$ tend vers $F(x) - F(x-)$ sur $\Omega \setminus N(x)$. Soit N la réunion des $N(x)$ pour $x \in S$ et $x \in Q$. Alors N est négligeable et sur $\Omega \setminus N$, toutes les convergences ponctuelles ci-dessus sont réalisées. Supposons que F_n ne tend pas uniformément vers F en un $\omega \in \Omega \setminus N$. Il existe $\varepsilon > 0$, une suite extraite $\phi(n)$ et une suite de réels x_n tels que

$$\forall n, \quad |F_{\phi(n)}(x_n) - F(x_n)| \geqslant \varepsilon.$$

Quelles sont les limites possibles de la suite $(x_n)_n$? En tout cas, pas $\pm\infty$. En effet, si x_n tend vers $-\infty$ en décroissant, on exhibe p tel que $F(x_n) \leqslant \varepsilon/3$ pour $n \geqslant p$. Alors, $F_{\phi(n)}(x_n) \geqslant \varepsilon - \varepsilon/3$ donc $F_{\phi(n)}(x_p) \geqslant F_{\phi(n)}(x_n) \geqslant 2\varepsilon/3$ ne peut converger vers $F(x_p) \leqslant \varepsilon/3$.

Supposons que x_n converge vers x fini, en croissant par exemple et avec $F_{\phi(n)}(x_n) - F(x_n) \geqslant \varepsilon$. Soit r et s des éléments de Q avec $r < x \leqslant s$. Pour $n \geqslant p$ avec p fixé, $x_n \geqslant r$ donc

$$\varepsilon \leqslant F_{\phi(n)}(x-) - F(r) \leqslant F_{\phi(n)}(x-) - F_{\phi(n)}(x) + F_{\phi(n)}(s) - F(r).$$

Si n tend vers l'infini, on obtient : $\varepsilon \leqslant F(x-) - F(x) + F(s) - F(r)$. On fait tendre r vers x en croissant puis s vers x en décroissant : il vient $\varepsilon \leqslant 0$. Les autres cas se traitent de même.

La suite $(x_n)_n$ est donc une suite réelle dont on ne peut extraire aucune sous–suite convergente. C'est absurde. ∎

Remarque Pour construire Q, on choisit $r(k,n)$ dans $]k/n, (k+1)/n[\setminus S$ et on pose

$$Q = \{r(k,n)\,;\, n \geqslant 1,\, |k| \leqslant n^2\}.$$

Théorème 2.21 (Kolmogorov–Smirnov) *Soit F une fonction de répartition continue. Alors la loi de* $\sup\{|F_n(x) - F(x)|\,;\, x \in \mathbb{R}\}$ *ne dépend pas de F. De même pour les lois de*

$$\sup\{F_n(x) - F(x)\,;\, x \in \mathbb{R}\} \quad \textit{et} \quad \inf\{F_n(x) - F(x)\,;\, x \in \mathbb{R}\}.$$

Démonstration Soit $(U_n)_n$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$ et $Y_n = G(U_n)$ pour

$$G(t) = \sup\{x\,;\, F(x) \leqslant t\}.$$

On sait que la suite $(Y_n)_n$ admet la même loi que la suite $(X_n)_n$ et que $F(x) \leqslant t$ est "équivalent" à $x \leqslant G(t)$. Plus précisément,

$$x < G(t) \Rightarrow F(x) \leqslant t \Rightarrow x \leqslant G(t).$$

Notons u_n la n–ième fonction de répartition empirique basée sur la suite $(U_n)_n$. Alors, $\mathbf{1}(Y_k < x) = \mathbf{1}(G(U_k) < x) = \mathbf{1}(0 \leqslant U_k < F(x))$ donc

$$F_n(x-) \overset{loi}{=} n^{-1} \sum_{k=1}^{n} \mathbf{1}(Y_k < x) = u_n(F(x)-).$$

A présent, le supremum de $|u_n(y-) - y|$ pour $0 \leqslant y \leqslant 1$ est aussi le supremum de $|u_n(y-) - y|$ pour $0 < y < 1$ qui est aussi le supremum de $|u_n(F(x)-) - F(x)|$ pour $x \in \mathbb{R}$ car F est continue donc $]0,1[\subset F(\mathbb{R})$, qui coïncide en loi avec le supremum de $|F_n(x-) - F(x)|$ pour $x \in \mathbb{R}$, et donc également avec le supremum de $|F_n(x) - F(x)|$ pour $x \in \mathbb{R}$ car F est continue.

Les autres démonstrations sont analogues. ∎

Complétons ce survol de quelques résultats statistiques en précisant les lois limites obtenues pour les quantités considérées quand n tend vers l'infini.

Théorème *Soit $(X_k)_{1\leqslant k\leqslant n}$ et $(Y_k)_{1\leqslant k\leqslant n}$ deux échantillons de la même loi continue et soit F_n et G_n les deux fonctions de répartition empiriques. Pour tout $x \geqslant 0$, on a*

$$\begin{aligned} \lim_{n\to\infty} P(\sup\{\sqrt{n}\,(F_n(t)-G_n(t))\,;\,t\in\mathbb{R}\}\leqslant x) &= 1-\exp(-x^2), \\ \lim_{n\to\infty} P(\sup\{\sqrt{n}\,|F_n(t)-G_n(t)|\,;\,t\in\mathbb{R}\}\leqslant x) &= \sum_{k=-\infty}^{+\infty}(-1)^k\exp(-k^2\,x^2). \end{aligned}$$

Démonstration On va simplement compter des chemins. Pour cela, ordonnons tous les X_k et Y_l en une suite croissante $(Z_i)_{i\leqslant 2n}$. Notons

$$x_i = \mathbf{1}(Z_i\in\{X_k\}) - \mathbf{1}(Z_i\in\{Y_l\}), \qquad s_k = \sum_1^k x_l.$$

Alors, $n\,(F_n(x)-G_n(x))$ est simplement le nombre de X_k inférieurs à x moins le nombre de Y_l inférieurs à x, donc $n\,(F_n-G_n)$ admet les mêmes variations sur $\mathbb{R}$ que $(s_k)_{k\leqslant 2n}$. En particulier,

$$n\sup\{F_n(t)-G_n(t)\,;\,t\in\mathbb{R}\} = \max\{s_k\,;\,1\leqslant k\leqslant 2n\}.$$

On doit donc (donc, car l'ordre est de loi uniforme) compter les trajectoires issues de $(0,0)$, passant par $(2n,0)$ et touchant la droite d'ordonnée h. Par principe de réflexion, la probabilité que $(s_k)_{k\leqslant 2n}$ touche h est le nombre de chemins de $(0,0)$ à $(2n,2h)$ divisé par le nombre de chemins de $(0,0)$ à $(2n,0)$, soit C_{2n}^{n-h}/C_{2n}^{n}. Enfin, Stirling. ■

Remarque Même formule avec $F_n - F$ si on remplace $\sqrt{2x}$ par $\sqrt{x}$.

Remarque Pour le problème des ex aequo, voir les questions de rang.

Sujet de devoir : LGN modérée, voir (article DiP 1996).

Aiguille de Buffon.

Exercices

1 Soit X_n des variables aléatoires indépendantes de même loi de Cauchy $\mathcal{C}(a)$.
1) Quelle est la loi de S_n/n ?
2) Montrer que la suite S_n/n ne converge pas P, ni d'ailleurs aucune de ses sous–suites.

2 Si $(X_n)_n$ est u.i., la suite $(S_n/n)_n$ est u.i. (Quelle caractérisation utiliser ?)

3 Illustrer par un exemple chacune des situations suivantes :
1) $\sum X_n/n$ converge presque sûrement et $\sum |X_n|/n$ diverge presque sûrement.
2) $\sum X_n$ converge presque sûrement et $\sum E(X_n)$ et $\sum E(X_n^2)$ divergent.
3) $\sum X_n$ diverge presque sûrement et $\sum E(X_n^2)$ converge.
4) $\sum X_n$ diverge presque sûrement et $\sum E(X_n^2)$ converge et $E(X_n) = 0$ pour tout n.

4 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi $\mu \neq \delta_0$ et soit $(a_n)_n$ une suite de réels.
1) On suppose que a_n ne tend pas vers 0. Montrer que $\sum a_n X_n$ diverge presque sûrement.
2) On suppose que $X_n \geqslant 0$ et $a_n \geqslant 0$ pour tout n, que a_n tend vers 0 et que X_n est de carré intégrable. Montrer que $\sum a_n X_n$ converge presque sûrement si et seulement si $\sum a_n$ converge. (Le sens direct utilise le théorème des trois séries.)

5 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes positives.
1) On suppose que $0 \leqslant X_n \leqslant c$ pour tout n. Montrer que $\sum X_n$ converge presque sûrement si et seulement si $\sum E(X_n)$ converge. (Le sens direct utilise le théorème des trois séries.)
2) Montrer que $\sum X_n$ converge presque sûrement si et seulement si $\sum E(X_n/(1+X_n))$ converge.

6 Pour $n \geqslant 1$, soit $r_n : [0,1[\to \{-1,+1\}$ défini par $r_n(t) = \operatorname{sgn} \sin(2^n \pi t)$ où on a décidé par convention de poser $\operatorname{sgn}(0) = 1$.
1) Montrer que $(r_n)_n$ est une suite de variables aléatoires indépendantes centrées et de même loi (que l'on calculera) dans $([0,1[, \mathcal{B}([0,1[), dx)$.
2) Soit $(\mu_n)_n$ une suite de lois sur $\mathbb{R}$. Déduire du (1) qu'il existe une suite de variables aléatoires X_n indépendantes, de lois respectives μ_n et définies sur $([0,1[, \mathcal{B}([0,1[), dx)$.

7 (un jeu presque équitable) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de lois $P(X_n = n^2 - 1) = n^{-2} = 1 - P(X_n = -1)$. Montrer que les X_n sont centrées mais que $n^{-1} \sum_{k=1}^n X_k$ tend p.s. vers -1.

8 (une application de la LGN L^4) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de loi $X_n(P) = \mathcal{P}(a)$ et $f : \mathbb{R}^+ \to \mathbb{R}$ une fonction continue et bornée.
1) Montrer que $S_n(P) = \mathcal{P}(n\,a)$.
2) Evaluer $E(f(x + S_n/n))$ pour montrer

$$f(x+a) = \lim_{n\to\infty} \sum_{k=0}^{+\infty} f(x + k/n)\, e^{-n\,a}\, (n\,a)^k/k!.$$

3) Soit μ une mesure de probabilité sur $\mathbb{R}$ intégrable, de moyenne m et μ^{*n} sa puissance n–ième de convolution. Montrer que, pour tout réel a, on a :

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f(a + (x/n))\, d\mu^{*n}(x) = f(a+m).$$

9 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi, positives et intégrables, et soit $f : \mathbb{R}^+ \to \mathbb{R}$ telle que $|f(x)| \leqslant c \cdot x$ pour tout réel x. Déterminer

$$\lim_{n\to\infty} E\left(\frac{\sum_{k=1}^n f(X_k)}{S_n}\right).$$

10 (Bernstein–suite) Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. On note $B_n(f)$ les polynômes de Bernstein associés :

$$B_n(f)(x) = \sum_{k=0}^n C_n^k\, x^k\,(1-x)^{n-k}\, f(k/n).$$

1) Fixons $x \in [0,1]$. Montrer qu'il existe une suite de variables aléatoires $(\varepsilon_n)_n$ i.i.d. avec

$$B_n(f)(x) = E(f(n^{-1}\sum_{k=1}^n \varepsilon_k)).$$

En déduire que $B_n(f)$ converge uniformément vers f. (On a redémontré le théorème de Stone–Weierstrass.)
2) Supposons que f est de plus hölderienne d'ordre $0 < a < 1$:

$$\forall x, y, \quad |f(x) - f(y)| \leqslant C\,|x-y|^a.$$

Montrer que dans ce cas, on a :

$$|B_n(f)(x) - f(x)| \leqslant C\, E|n^{-1}\sum_{k=1}^n \varepsilon_k - x|^a.$$

En utilisant la concavité de la fonction $t \mapsto t^{a/2}$ sur $t \geqslant 0$, démontrer le taux de convergence suivant :

$$\forall n \geqslant 1, \forall x \in [0,1], \quad |B_n(f)(x) - f(x)| \leqslant C/n^{a/2}.$$

11 (LGN L^2) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de carré intégrable et centrées. On sait que si $\sum_n \mathrm{var}(X_n)/n^2$ converge, la suite S_n/n converge presque sûrement vers 0.
1) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de lois $P(X_n = +1) = P(X_n = -1) = (1-2^{-n})/2$ et $P(X_n = +2^n) = P(X_n = -2^n) = 2^{-n}/2$. Montrer que S_n/n converge presque sûrement et que $\sum_n \mathrm{var}(X_n)/n^2$ diverge.
2) Soit $(v_n)_n$ une suite de réels $v_n \geqslant 0$ telle que $\sum_n v_n^2/n^2$ diverge. Exhiber une suite $(X_n)_n$ de variables aléatoires indépendantes de carré intégrable et centrées de variances $\mathrm{var}(X_n) = v_n^2$ telle que S_n/n diverge presque sûrement.
On cherchera X_n à valeurs dans $\{-n, 0, +n\}$ si $v_n \leqslant n$ et dans $\{-v_n, +v_n\}$ sinon, de telle sorte que la somme des $P(|X_n| \geqslant \varepsilon n)$ diverge pour un certain $\varepsilon > 0$.

12 (estimateurs de la moyenne) Soit $(X_k)_{k\leqslant n}$ des variables aléatoires indépendantes de loi uniforme sur $[a,b]$. Les bornes a et b sont inconnues et on veut estimer $(a+b)/2$ à partir de l'échantillon observé. Deux estimateurs sont proposés :

$$M_n = S_n/n \quad \text{et} \quad T_n = (\inf X_k + \sup X_k)/2.$$

Calculer $E(M_n)$, $E(T_n)$, $\mathrm{var}(M_n)$ et $\mathrm{var}(T_n)$ et en déduire que M_n et T_n sont deux estimateurs "consistants" (c'est-à-dire qu'ils convergent bien vers $(a+b)/2$ pour n grand), mais que T_n est meilleur que M_n. On pourra se ramener au cas $(a,b) = (0,1)$ pour simplifier les calculs, puis, pour T_n, remarquer que $I = \inf X_k$ suit la loi de $1 - S = 1 - \sup X_k$. On calculera ensuite $P(x \leqslant I, S \leqslant y)$, $P(S \in dy)$, $E(S)$ et $E(S^2)$. On montrera ensuite :

$$E((S-I)^2) = \frac{n(n-1)}{(n+2)(n+1)}, \quad \mathrm{var}(S+I) = E((2S-1)^2) - E((S-I)^2).$$

13 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi de carré intégrable. Posons

$$V_n = \sum_{k=1}^{n} (X_k - (S_n/n))^2.$$

1) Calculer $E(S_n)$ et $E(V_n)$.
2) Si la loi de X_1 est $\mathcal{N}(m, \sigma^2)$, montrer que S_n et V_n sont indépendants et trouver leurs lois.
3) On fixe $n \geqslant 2$ et on suppose que S_n et V_n sont indépendantes. Calculer de deux façons différentes $E(V_n \exp(itS_n))$ en fonction de la fonction $\phi : s \mapsto E(\exp(isX_1))$. Trouver une équation différentielle satisfaite par ϕ et en déduire que la loi de X_1 est gaussienne.

14 (Transformée de Laplace) Soit f une fonction continue bornée sur $\mathbb{R}_+$ et $\lambda > 0$. La transformée de Laplace de f évaluée en λ est

$$Lf(\lambda) = \int_0^{+\infty} e^{-\lambda x} f(x)\, dx.$$

Soit $((X_n)_n$ une suite de variables aléatoires indépendantes de loi $\mathcal{E}(\lambda)$ c'est-à-dire

$$P(X_n \geqslant x) = e^{-\lambda x}$$

pour $x \geqslant 0$. Montrer que

$$E[f(X_1 + \cdots + X_n)] = (-1)^{n-1} \frac{\lambda^n\, (Lf)^{(n-1)}(\lambda)}{(n-1)!}.$$

En déduire que l'on peut retrouver f à partir de Lf de la façon suivante :

$$f(x) = \lim_{n \to \infty} (-1)^{n-1} \frac{(n/x)^n\, (Lf)^{(n-1)}(n/x)}{(n-1)!}.$$

15 (Gauss et les sphères de dimension infinie) 1) Soit X un vecteur de $\mathbb{R}^d$ dont les composantes sont indépendantes de loi normale $\mathcal{N}(0,1)$ et soit O une transformation orthogonale de $\mathbb{R}^d$. Montrer que OX suit la même loi que X. En déduire que $X/\|X\|$ suit une loi uniforme sur la sphère unité S^{d-1} de $\mathbb{R}^d$.
2) Soit $(Z_n)_n$ une suite de variables aléatoires indépendantes de même loi normale $\mathcal{N}(0,1)$ et notons

$$R_n = (Z_1^2 + \cdots + Z_n^2)^{1/2}.$$

Montrer que $R_n/\sqrt{n}$ tend presque sûrement vers 1 quand n tend vers l'infini.
3) En déduire que si pour chaque $d \geqslant 1$, le point $Y^{(d)} = (Y_k^{(d)})_{1 \leqslant k \leqslant d}$ est choisi uniformément sur la sphère $\sqrt{d} S^{d-1}$ de $\mathbb{R}^d$, alors

$$P(Y_1^{(d)} \leqslant x) \longrightarrow \phi(x) = \int_{-\infty}^{x} e^{-y^2/2}\, dy,$$
$$P(Y_1^{(d)} \leqslant x_1,\, Y_2^{(d)} \leqslant x_2) \longrightarrow \phi(x_1)\, \phi(x_2).$$

Commenter.

16 (suites échangeables)
1) Soit $(X_k)_{k \leqslant n}$ des variables aléatoires échangeables de carré intégrable. Calculer $\mathrm{var}(S_n)$.
2) Soit $(X_k)_k$ une suite de variables aléatoires échangeables de carré intégrable. Montrer que

$$\mathrm{cov}(X_1, X_2) \geqslant 0.$$

3) Montrer qu'il existe n variables aléatoires échangeables $(Y_k)_{k\leqslant n}$ non plongeables dans une famille de $n+1$ variables aléatoires échangeables. On pourra introduire des variables aléatoires indépendantes de même loi X_k et poser $Y_k = X_k - S_n/n$.
4) Soit $(X_k)_{k\leqslant n}$ des variables aléatoires échangeables intégrables. Montrer que

$$E\,|S_n/n| \leqslant E\,|S_{n-1}/(n-1)|\,.$$

5) Soit S_f l'ensemble des permutations de $\mathbb{N}^*$ de support fini et $(X_n)_n$ une suite de variables aléatoires. On note $\mathcal{A}$ la tribu asymptotique et $\mathcal{E}$ la tribu des échangeables : $A \in \mathcal{E}$ si A est mesurable par rapport à $(X_n)_n$ et si $P(s(A)\triangle A) = 0$ pour tout $s \in S_f$.
Montrer que $\mathcal{A} \subset \mathcal{E}$ et donner un exemple pour lequel $\mathcal{A} \neq \mathcal{E}$.

17 (LLI pour des variables aléatoires gaussiennes) Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de loi normale $\mathcal{N}(0,1)$ et $S_n = X_1 + \cdots + X_n$. On notera $h(t) = \sqrt{2t\log\log t}$ pour $t \geqslant 3$. On va montrer que presque sûrement

$$\limsup_n \frac{S_n}{h(n)} = 1.$$

1) Montrer que $E(e^{t\,S_n}) = e^{n\,t^2/2}$.
2) Soit $t \geqslant 0$. Adapter la preuve de l'inégalité de Kolmogorov à la fonction $s \mapsto \exp(st)$ pour montrer, pour tout $c \geqslant 0$, que l'on a :

$$P(\sup_{k\leqslant n} S_k \geqslant c) \leqslant e^{-c\,t}\,E(e^{t\,S_n}).$$

En déduire la majoration suivante :

$$P(\sup_{k\leqslant n} S_k \geqslant c) \leqslant e^{-c^2/(2n)}.$$

3) Soit $a > 1$ un réel fixé et $a_n = a\,h(a^{n-1})$. Utiliser BC–trivial pour montrer que presque sûrement,

$$\sup_{k\leqslant a^n} S_k \leqslant a_n$$

à partir d'un certain rang n. En déduire que la limite supérieure de $S_n/h(n)$ est presque sûrement inférieure à a et enfin que la limite supérieure de $S_n/h(n)$ est presque sûrement inférieure à 1.
4) Soit p un entier $p \geqslant 2$ et ε un réel $0 < \varepsilon < 1$. Estimer la probabilité de l'événement

$$A_n = \{S_{p^{n+1}} - S_{p^n} \geqslant (1-\varepsilon)\,h(p^{n+1} - p^n)\}$$

à l'aide de la fonction

$$\Phi(x) = \int_x^{+\infty} \phi(y)\,dy \quad \text{avec} \quad \phi(y) = \exp(-y^2/2)/\sqrt{2\pi}.$$

5) Pour $x > 0$, montrer la double inégalité suivante :

$$\phi(x)/x \geqslant \Phi(x) \geqslant \phi(x)/(x + 1/x).$$

On pourra calculer la dérivée de $\phi(y)$ et celle de $\phi(y)/y$.
6) Appliquer BC–non trivial pour montrer que presque sûrement, une infinité de A_n ont lieu. Combiner l'estimation ainsi obtenue avec la conclusion du (3) pour en déduire

$$\limsup_n S_{p^{n+1}}/h(p^{n+1}) \geqslant (1-\varepsilon)\,\sqrt{1 - 1/p} - 2\sqrt{1/p}.$$

Conclure.

18 Soit $(X_n)_n$ une suite de variables aléatoires et N une variable aléatoire entière $N \geqslant 1$. On note

$$S_N : \omega \mapsto \sum_{k=1}^{N(\omega)} X_k(\omega).$$

1) Rappeler pourquoi S_N est une variable aléatoire.
2) On suppose que les X_n sont i.i.d. et intégrables. Soit $(N_k)_k$ une suite de variables aléatoires entières avec $N_k \xrightarrow{p.s.} +\infty$. Montrer que S_{N_k}/N_k tend presque sûrement vers $E(X_1)$.
2) On suppose que les X_n sont i.i.d. et de carré intégrable et que N est indépendante de la suite $(X_n)_n$ et de carré intégrable. Calculer $E(S_N)$ et $\mathrm{var}(S_N)$ en fonction de $E(X_1)$, $E(N)$, $\mathrm{var}(X_1)$ et $\mathrm{var}(N)$.
3) La fonction génératrice d'une variable aléatoire X positive est $\phi_X(t) = E(t^X)$ pour $t > 0$. Calculer ϕ_{S_N} en fonction de ϕ_{X_1} et de ϕ_N.
Application : un poisson pond $\mathcal{P}(a)$ œufs. Chaque œuf survit jusqu'à l'âge adulte avec une probabilité p. Nombre d'adultes ?

Solutions

1 La fonction $(a/\pi)/(a^2 + x^2)$ est $1/(2\pi)$ la transformée de Fourier de $e^{-a|t|}$ donc si X est $\mathcal{C}(a)$, on a $E(e^{itX}) = e^{-a|t|}$. Si Y est $\mathcal{C}(b)$ et indépendante de X, on voit que $X + Y$ est $\mathcal{C}(a + b)$.
Cv P ou sous–suite cv P donne une sous–suite qui cv presque sûrement, loi du 0–1 donc vers une constante S. Or $P(|S_n/n - S| < 1)$ ne tend pas vers 1 pour une loi de Cauchy.

3 1) ± 1 ind. 2) $(1 - n^{-2})\delta_0 + n^{-2}\delta_{n^2}$ ind. 3) $1/n$. 4) X_1/n avec $X_1 = \pm 1$.

4 1) Ne dv pas presque sûrement donne cv presque sûrement donne $a_n X_n$ cv presque sûrement vers 0 donne cv P vers 0. Pour $a_n \geqslant a > 0$, on doit avoir $P(|X_n| \geqslant \varepsilon/a) \to 0$ pour tout ε. Absurde.
2) $\sum a_n$ cv donne $\sum a_n\, E(X_n)$ fini donc $\sum a_n\, X_n$ fini presque sûrement.
$\sum a_n\, X_n$ cv donc (3 séries) $\sum a_n\, E(X_n : X_n \leqslant 1/a_n)$ cv. Il existe t avec $E(X : X \leqslant t) \geqslant E(X)/2$ et il existe n_0 tel que $a_n \leqslant 1/t$ après n_0 donc $\sum a_n\, E(X_n : X_n \leqslant 1/a_n)$ minorée pour les termes après n_0 par $\sum a_n\, E(X)/2$ donc $\sum a_n$ cv.

5 1) $\sum E(X_n)$ cv donc $\sum X_n$ fini presque sûrement.
$\sum X_n$ cv donc (3 séries, s_2) $\sum E(X_n)$ cv.
2) Soit $Y_n = X_n/(1 + X_n)$.

9 $E(f(X))/E(X)$. L'hypothèse sert à assurer $f(X)$ intégrable.

Chapitre 3

Conditionnement

La base du chapitre va jusqu'à la section 3.4, qui pourra être distribuée à la classe. On peut restreindre le problème des versions régulières (section 3.5) au minimum (partie 3.5). On pourra quand même énoncer le théorème de Jirina (3.1) pour amuser la foule. La section 3.6 est difficile et peut être omise. On traitera la section 3.7 (à l'exception de la partie 3.7) seulement si on souhaite aborder l'étude des variables gaussiennes. L'étude des files d'attente (section 3.8) est une application élémentaire, au moins dans son début, du reste du chapitre.

• • •

Notation : pour $A \in \mathcal{F}$ et X intégrable, on écrit $E(X : A)$ ou $E(X; A)$ pour $E(X\,\mathbf{1}_A)$.

3.1 Motivation

Rappel d'un exemple : un élève vient en cours une fois sur deux. Quand il vient, il est en retard une fois sur trois. Le cours commence et l'élève n'est pas là. Quelle est la probabilité pour qu'il vienne ?
Réponse : un dé équitable donne 1, 2 ou 3 s'il vient et 4, 5, 6 sinon. Le retard est le 1. Le cours commence sans l'élève donc 1, 4, 5 ou 6. Il vient si 1. Donc $p = 1/4$.
[Dessin.]

D'où la notation : pour A, $B \in \mathcal{F}$, $P(B) > 0$, $P(A|B) = P(A \cap B)/P(B)$.

Interprétation : on sait que B est réalisé ($\omega \in B$) donc on remplace $P(A)$ par $P(A|B)$ qui décrit mieux A ; si on savait que $\omega \notin B$, on remplacerait $P(A)$ par $P(A|B^c)$ pour décrire A.

En résumé : dans le cas normal, on veut mesurer la taille de A. Réponse $P(A)$. Ici, on veut mesurer la taille de A avec l'information "on est dans B". Réponse $P(A|B)$. On peut aussi chercher à mesurer la taille d'une variable aléatoire X. Réponse $E(X)$. Si on se donne l'information contenue dans une autre variable aléatoire Z, on obtient...

Une extension : X et Z sont des variables aléatoires qui prennent un nombre fini de valeurs, les x_i, $i \leqslant n$, pour X et les z_i, $i \leqslant m$, pour Z. Par exemple, X est la partie entière de la température exprimée en degrés Celsius au sommet du mont Aigoual à midi et Z est le numéro du jour de l'année. Comment décrire la moyenne de la température X selon le jour de l'année Z par un nombre de la même façon que l'on remplace X dans le cas classique par sa moyenne ?

On veut donc mesurer la moyenne de X (soit $E(X)$ dans le cas normal) avec l'information apportée par Z. Réponse : on fait la moyenne de X sur le sous–ensemble considéré. Il vient :

$$P(X = x_i | Z = z_j) = P(X = x_i,\, Z = z_j)/P(Z = z_j),$$

$$E(X|Z=z_j) = \sum x_i\, P(X=x_i|Z=z_j) = E(X : Z=z_j)/P(Z=z_j).$$

On appelle *espérance conditionnelle de X sachant Z* et on note $E(X|Z)$ une variable aléatoire $Y : \Omega \to \mathbb{R}$ telle que

$$Z(\omega) = z_j \quad \Rightarrow \quad Y(\omega) = E(X|Z=z_j).$$

Interprétation : on sait sur quel ensemble $\{Z = z_j\}$ on se trouve et on remplace X par sa moyenne sur cet ensemble.

Dessin : partition d'un patatoïde.
On appelle Z–atome, un atome de $\mathcal{G} = \sigma(Z)$.

Une description de Y en termes de tribus : en fait, Y ne dépend que de la tribu $\mathcal{G}$ engendrée par Z. Deux propriétés de Y :

(i) Y est constante sur les Z–atomes donc Y est $\mathcal{G}$–mesurable ;
(ii) soit $\{Z = z_j\}$ un Z–atome :

$$E(Y : Z=z_j) = y_j\, P(Z=z_j) = E(X : Z=z_j),$$

où y_j est la valeur de $E(X|Z=z_j)$, donc de Y, sur tout $\{Z = z_j\}$. Enfin tout événement de $\mathcal{G}$ est réunion de Z–atomes donc $E(X : A) = E(Y : A)$ pour tout $A \in \mathcal{G}$.

3.2 Le théorème fondamental

Théorème *Soit $\mathcal{G}$ une sous–tribu de $\mathcal{F}$ et X une variable aléatoire intégrable. Il existe une variable aléatoire intégrable Y telle que :*

(c1) Y est $\mathcal{G}$–mesurable ;
(c2) Pour tout $A \in \mathcal{G}$, $E(Y : A) = E(X : A)$.

De plus, si Y' est une autre variable aléatoire intégrable vérifiant (c1) et (c2), alors $Y = Y'$ presque sûrement.

Toute variable aléatoire intégrable vérifiant (c1) et (c2) est appelée une version de l'espérance conditionnelle de X sachant $\mathcal{G}$ et notée $E(X|\mathcal{G})$.

Notation : $E(X|Z) = E(X|\sigma(Z))$. Dans ce cas, on peut remplacer (c1) et (c2) par

(c1)' Il existe g borélienne intégrable pour P_Z telle que $Y = g(Z)$;
(c2)' Pour toute h borélienne bornée, $E(h(Z)\,Y) = E(h(Z)\,X)$.

Interprétation : $\mathcal{G}$ est l'information dont on dispose donc on connaît $Z(\omega)$ pour toute variable aléatoire Z $\mathcal{G}$–mesurable. Alors $E(X|\mathcal{G})(\omega)$ est la moyenne de X, toutes ces valeurs de $Z(\omega)$ étant données.
Un cas particulier : si $\mathcal{G} = \sigma(Z_0)$, toute variable aléatoire Z $\mathcal{G}$–mesurable est une fonction de Z_0. Se donner toutes les valeurs $Z(\omega)$ revient donc à se donner seulement la valeur de $Z_0(\omega)$.

Quelques exemples :

- **Dessin :** fibres. T variable aléatoire sur $\mathbb{R}^2$ et $\mathcal{G}$ tribu engendrée par la première coordonnée.
- si $\mathcal{G}$ est finie et engendrée par une partition $(B_i)_{i\geqslant 0}$ de Ω avec $P(B_0) = 0$ et $P(B_i) > 0$ pour $i \geqslant 1$, alors Y convient pour tout y réel :

  $$Y = y\,\mathbf{1}_{B_0} + \sum_{i\geqslant 1} E(X|B_i)\,\mathbf{1}_{B_i}.$$

  Le "presque sûrement" de la définition est donc inévitable.
- Si X est $\mathcal{G}$–mesurable, $Y = X$ convient ; si $\mathcal{G} = \{\emptyset, \Omega\}$, $Y = E(X)$ convient.
  On remarque que $Y = E(X)$ vérifie toujours (c1) et $Y = X$ vérifie toujours (c2). Dans le cas général, $E(X|\mathcal{G})$ est un compromis entre les deux.

$E(\cdot|\mathcal{G})$ comme projection de L^2

Notons $L^2(\mathcal{F}) = L^2(\Omega, \mathcal{F}, P)$ et

$$L^2(\mathcal{G}) = \{X \in L^2(\mathcal{F})\,;\, X\text{est } \mathcal{G}\text{–mesurable}\}.$$

Alors $L^2(\mathcal{F})$ est un espace de Hilbert pour le produit scalaire $(X, Y) = E(X\overline{Y})$ et $L^2(\mathcal{G})$ est un sous–espace fermé de $L^2(\mathcal{F})$ (à faire). Soit $X \in L^2(\mathcal{F})$ et Y la projection orthogonale de X sur $L^2(\mathcal{G})$. Alors, $Y = E(X|\mathcal{G})$ presque sûrement.

En effet, Y est caractérisée par : $Y \in L^2(\mathcal{G})$ et $Y - X$ orthogonal à $L^2(\mathcal{G})$, ce qui entraîne respectivement (c1) et (c2). (Remarque : $A \in \mathcal{G}$ donne $\mathbf{1}_A \in L^2(\mathcal{G})$.)

La construction

– L'unicité presque sûre : pour $A \in \mathcal{G}$, $E(Y - Y' : A) = 0$. On choisit $A = \{Y \geqslant Y' + \varepsilon\}$.
– L'existence pour X de carré intégrable : la projection orthogonale de X.
– L'existence pour X intégrable :
on écrit X comme $X^+ - X^-$ donc on suppose $X \geqslant 0$. Soit $X_n = X \wedge n$. Alors X_n est de carré intégrable et $0 \leqslant X_n$ tend en croissant vers X. Soit $Y_n = E(X_n|\mathcal{G})$. Supposons que les Y_n sont positives et forment une suite croissante et posons alors $Y = \sup Y_n$. On va vérifier que Y convient : (c1) et (c2) par convergence monotone. Enfin Y est intégrable car Y positive et $E(X) = E(Y)$.

Lemme *Soit X une variable aléatoire positive bornée. Alors $E(X|\mathcal{G}) \geqslant 0$ presque sûrement.*

Démonstration Soit $\varepsilon > 0$ et $A = \{E(X|\mathcal{G}) \leqslant -\varepsilon\}$. Alors $A \in \mathcal{G}$ et $P(A) = 0$. En effet,

$$0 \leqslant E(X : A) = E(E(X|\mathcal{G}) : A) \leqslant -\varepsilon\, P(A) \leqslant 0.$$

■

Une remarque importante

On peut obtenir directement le théorème fondamental à l'aide d'un résultat un peu plus sophistiqué que notre construction à la main : le théorème de Radon–Nikodym. Nous expliquons comment procéder : soit X positive et intégrable et $Q : \mathcal{G} \to \mathbb{R}_+$ définie par $Q(A) = E(X : A)$. Alors Q est une mesure finie sur $\mathcal{G}$. Soit P_0 la restriction de P à $\mathcal{G}$. Le point important est que Q est absolument continue par rapport à P_0. Soit $Y = dQ/dP_0$ la densité (la dérivée de Radon–Nikodym) de Q par rapport à P_0 sur $\mathcal{G}$. Alors Y est $\mathcal{G}$–mesurable et pour tout $A \in \mathcal{G}$,

$$E(Y : A) = \int_A \frac{dQ}{dP_0}\, dP = Q(A) = E(X : A).$$

Finalement, $Y = E(X|\mathcal{G})$ presque sûrement.

Remarque Soit $\mathcal{C}$ un π–système qui engendre $\mathcal{G}$ avec $\Omega \in \mathcal{C}$. On peut remplacer (c2) par la propriété (c2") plus faible :
(c2") Pour tout $A \in \mathcal{C}$, $E(Y : A) = E(X : A)$.
La condition $\Omega \in \mathcal{C}$ est indispensable.

3.3 Quelques exemples

Le cas $X = \mathbf{1}_A$: la probabilité conditionnelle

Quand $X = \mathbf{1}_A$, on note $E(X|\mathcal{G}) = P(A|\mathcal{G})$. Les propriétés de cette fonction de A ressemblent à celles d'une probabilité. Ainsi, on démontre facilement que :
- $0 \leqslant P(A|\mathcal{G}) \leqslant 1$ p.s.
- si les A_n sont disjoints, $P(\cup_n A_n|\mathcal{G}) = \sum_n P(A_n|\mathcal{G})$ p.s. (Attention aux négligeables.)
- si A_n tend vers A en croissant, resp. en décroissant, alors $P(A_n|\mathcal{G})$ tend vers $P(A|\mathcal{G})$ p.s. en croissant, resp. en décroissant.
- si $P(A) = 0$ ou 1, alors $P(A|\mathcal{G}) = 0$ ou 1 p.s.

Remarque Tout fonctionne si l'on reste dans le dénombrable.

Le cas densitable

Supposons que la loi de (X, Y) est $f_{X,Y}(x, y)\, dx\, dy$ avec $f_{X,Y} > 0$ partout. Alors la loi de X est $f_X(x)\, dx$ et la loi de Y est $f_Y(y)\, dy$ avec

$$f_X(x) = \int f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int f_{X,Y}(x, y)\, dx.$$

[Dessin.]
Soit h une fonction borélienne telle que $h(X)$ est intégrable. Calculons $Z = E(h(X)|Y)$. D'après (c1), Z est de la forme $Z = g(Y)$. D'après (c2), pour tout A dans la tribu engendrée par Y c'est-à-dire pour tout $A = \{Y \in B\}$ avec $\mathcal{B}(\mathbb{R})$, on peut écrire :

$$\begin{aligned} E(h(X) : A) &= E(h(X)\, \mathbf{1}_B(Y)) = \int h(x)\, \mathbf{1}_B(y)\, f_{X,Y}(x, y)\, dx\, dy, \\ E(g(Y) : A) &= \int g(y)\, \mathbf{1}_B(y)\, f_Y(y)\, dy. \end{aligned}$$

On sait résoudre en g ! Il vient

$$g(y) = (f_Y(y))^{-1} \int h(x)\, f_{X,Y}(x, y)\, dx.$$

Loi conditionnelle

On se donne deux variables aléatoires X et Y définies sur le même espace. La loi conditionnelle de X sachant Y est une collection $(\nu_y)_y$ indexée par $\mathbb{R}$ (ou par l'espace où Y prend ses valeurs) et telle que, pour toute h borélienne bornée,

$$E(h(X)\,|\,Y) = g(Y) \quad \text{p.s.,} \quad \text{avec} \quad g(y) = \int h(x)\, d\nu_y(x).$$

Le cas densitable (suite)

Dans le cas de la Section 3.3, on définit donc *la loi conditionnelle de X sachant Y* par sa densité $f_{X|Y}$ comme suit :

$$f_{X|Y}(x|y) = f_{X,Y}(x, y)/f_Y(y).$$

Alors l'espérance conditionnelle de $h(X)$ sachant Y est bien donnée par :

$$E(h(X)|Y)(\omega) = \int h(x)\, f_{X|Y}(x|y)\, dx \quad \text{pour} \quad y = Y(\omega),$$

donc, avec les notations de la Section 3.3, $\nu_y(dx) = f_{X|Y}(x|y)\, dx$ convient.

Dessin : bande dans le plan et masses comparées des portions de barreaux horizontaux.

Une justification a posteriori dans le cas absolument continu :

$$\begin{aligned} P(X \in A|Y \in (y, y+\varepsilon)) &= P(X \in A, Y \in (y, y+\varepsilon))/P(Y \in (y, y+\varepsilon)) \\ &= \frac{\iint \mathbf{1}_A(x)\, \mathbf{1}(y \leqslant y' \leqslant y+\varepsilon)\, f_{X,Y}(x, y')\, dx\, dy'}{\int \mathbf{1}(y \leqslant y' \leqslant y+\varepsilon)\, f_Y(y')\, dy'} \\ &\sim \frac{\int \mathbf{1}_A(x)\, dx\, \varepsilon\, f_{X,Y}(x, y)}{\varepsilon\, f_Y(y)}, \end{aligned}$$

donc il est naturel de penser que la limite du membre de droite quand ε devient petit peut être choisie comme une valeur raisonnable de

$$P(X \in A|Y = y).$$

Une extension : soit h borélienne telle que $h(X, Y)$ est intégrable. Alors,

$$E(h(X, Y)|Y) = g(Y) \quad \text{avec} \quad g(y) = \int h(x, y)\, f_{X|Y}(x|y)\, dx.$$

On sait qu'alors, $E(h(X, Y)) = E(g(Y))$. Montrons que cette formule n'est rien d'autre qu'une formule de Fubini :

$$E(h(X, Y)) = \int h(x, y)\, f_{X,Y}(x, y)\, dx\, dy,$$
$$E(g(Y)) = \int g(y)\, f_Y(y)\, dy,$$

donc la mesure $f_{X|Y}(x|y)\, dx$ est simplement la bonne mesure sur le barreau $\{Y = y\}$ pour décomposer la loi de (X, Y) selon

$$P_{(X,Y)}(A \times B) = \int_B P_X(A|Y = y)\, P_Y(dy).$$

Conditionnement et indépendance

Soit X et Y deux vecteurs aléatoires indépendants et f une fonction borélienne bornée. Notons

$$\phi_f(x) = E(f(x, Y)).$$

Alors $\phi_f(X)$ est une version de $E(f(X, Y)|X)$.

Démonstration Soit A un borélien. On veut montrer

$$E(f(X, Y)\, \mathbf{1}_A(X)) = E(\phi_f(X)\, \mathbf{1}_A(X)).$$

Deux méthodes : classe monotone sur f en partant de f indicatrice d'un pavé de la forme $f(x, y) = \mathbf{1}_B(X)\, \mathbf{1}_C(Y)$; ou bien, Fubini. ■

Conditionnement et lois

Proposition *Soit X et Y (resp. X' et Y') deux vecteurs aléatoires définies sur Ω (resp. Ω') tels que (X,Y) suit la même loi que (X',Y'). On sait que $E(X|Y)=f(Y)$ p.s., pour une fonction f borélienne. Alors $f(Y')$ est une version de $E(X'|Y')$.*

Moralité : l'espérance conditionnelle ne dépend pas des lois.
Corollaire : les lois conditionnelles non plus.

Démonstration On teste tout en calculant des espérances absolues qui ne font intervenir que la loi $\mu = P_{(X,Y)} = P_{(X',Y')}$. ■

3.4 Les propriétés de l'espérance conditionnelle

On se donne des variables aléatoires X intégrables et des sous–tribus de $\mathcal{F}$ notées $\mathcal{G}$ et $\mathcal{H}$.

- Si Y est une version de $E(X|\mathcal{G})$, alors $E(Y)=E(X)$.
- si X est $\mathcal{G}$–mesurable, alors $X=E(X|\mathcal{G})$ p.s.
- linéarité
- positivité
  Déjà fait en lemme pour X bornée.
- cMON
  On utilise l'argument de la construction de $E(X|\mathcal{G})$ pour $X \geqslant 0$ intégrable. Ici, on pose $Y_n = E(X_n|\mathcal{G})$ et $Y = \limsup Y_n$. Alors Y vérifie (c1) et (c2).
- cFATOU
- cDOM
  cMON implique cFATOU implique cDOM tout comme MON implique FATOU implique DOM.
  Rappels : (MON) soit $f_n \geqslant 0$ mesurables avec f_n converge vers f en croissant. Alors $\mu(f_n)$ tend vers $\mu(f)$.
  L'outil : on pose $e_n(x) = 2^{-n}\cdot[2^n x]$ si $0 \leqslant x \leqslant n$ et $e_n(x)=n$ si $x>n$. Alors $f_n = e_n(f)$ est une suite croissante de fonctions simples tendant vers f.
  (FATOU) Si $f_n \geqslant 0$, on pose $g_n = \inf_{k\geqslant n} f_k$. Alors la limite inférieure des f_n est la limite croissante des g_n. On montre alors
  $$\mu(g_n) \leqslant \inf_{k\geqslant n} \mu(f_k),$$
  donc (MON) appliqué à la suite g_n donne
  $$\begin{aligned}\mu(\liminf f_n) &= \lim \mu(g_n)\\ &\leqslant \lim_n \inf_{k\geqslant n} \mu(f_k) = \liminf \mu(f_n),\end{aligned}$$
  soit (FATOU).
  (DOM) On sait que $0 \leqslant 2g - |f_n - f| \leqslant 2g$ et $\mu(2g)$ est fini donc (FATOU) sur la suite des $2g-|f_n-f|$ donne (DOM).
- cJENSEN. Corollaire : pour $p \geqslant 1$, $\|E(X|\mathcal{G})\|_p \leqslant \|X\|_p$.
  Preuve analogue au cas non conditionnel : une fonction convexe est un supremum de fonctions affines. (Attention tout de même aux négligeables.) Pour le corollaire, $c(t)=t^p$ est convexe, puis on prend l'espérance des deux côtés.
- cascade
- si Z est $\mathcal{G}$–mesurable et bornée, alors $E(ZX|\mathcal{G}) = ZE(X|\mathcal{G})$ p.s.
  Procédé standard et intégrabilités correctes.

– si $\mathcal{H}$ est indépendante de $\sigma(X, \mathcal{G})$ alors $E(X|\sigma(\mathcal{G}, \mathcal{H})) = E(X|\mathcal{G})$ p.s.
On suppose que $X \geqslant 0$ et on pose $Y = E(X|\mathcal{G})$. Soit $A \in \mathcal{G}$ et $B \in \mathcal{H}$. Alors,

$$E(X : A \cap B) = E((X\,\mathbf{1}_A)\mathbf{1}_B) = E(X\,\mathbf{1}_A)\,P(B).$$

Or, $Y\mathbf{1}_A$ est indépendant de B donc $E((Y\,\mathbf{1}_A)\,\mathbf{1}_B) = E(Y\,\mathbf{1}_A)\,P(B)$. on a montré que les deux mesures $C \mapsto E(X : C)$ et $C \mapsto E(Y : C)$ coïncident sur le π–système des $C = A \cap B$ avec $A \in \mathcal{G}$ et $B \in \mathcal{H}$. Elles coïncident donc sur la tribu $\sigma(\mathcal{G}, \mathcal{H})$. ∎

3.5 Le problème des versions régulières

Motivation et "résultat"

Pour $A \in \mathcal{F}$ fixé, on a construit une (classe d'équivalence de) variable(s) aléatoire(s), notée $P(A|\mathcal{G})$. Pour un $\omega \in \Omega$ fixé, quel est le comportement de l'application

$$\mathcal{F} \to \mathbb{R}, \quad A \mapsto P(A|\mathcal{G})(\omega)\ ?$$

Est-ce une probabilité ? Presque, comme vu plus haut. Cela dit, l'exemple de la section 3.2 pose problème :

$$P(A|\mathcal{G}) = c_A\,\mathbf{1}_{B_0} + \sum_i P(A|B_i)\,\mathbf{1}_{B_i},$$

donc, pour $\omega \in B_0$ et un mauvais choix de $A \mapsto c_A$, on n'obtient pas une probabilité. Ici, le problème est circonscrit à B_0 qui est négligeable. Y a-t-il des cas où les négligeables s'additionnent mal ?

Plus délicat : on veut avoir

$$P(\bigcup_n A_n|\mathcal{G}) = \sum_n P(A_n|\mathcal{G}) \text{ p.s.}$$

pour toute suite $(A_n)_n$ d'éléments disjoints de $\mathcal{F}$. Pour chaque suite $\mathcal{A} = (A_n)_n$, le "presque sûrement" élimine un négligeable $N(\mathcal{A})$ mais les suites $\mathcal{A}$ peuvent être en nombre non dénombrable.

Définition *On dit que $p : \mathcal{F} \times \Omega \to \mathbb{R}_+$ est une version régulière de la probabilité P sachant $\mathcal{G}$ si on a :*
(r1) pour tout $\omega \in \Omega$, l'application $A \mapsto p(A, \omega)$ est une probabilité sur $\mathcal{F}$;
(r2) pour tout $A \in \mathcal{F}$, l'application $\omega \mapsto p(A, \omega)$ est une version de $P(A|\mathcal{G})$, c'est-à-dire qu'elle vérifie (c1) et (c2) associées à $\mathbf{1}_A$:
(c1) $p(A, \cdot)$ est $\mathcal{G}$–mesurable ;
(c2) pour tout $B \in \mathcal{G}$, $P(A \cap B) = \int_B p(A, \omega)\,dP(\omega)$.

"Théorème" *Dans la plupart des cas, une version régulière de la probabilité P sachant $\mathcal{G}$ existe.*

Remarque Dans le cas densitable, on a déjà construit p, version régulière de la probabilité P_X sachant $\sigma(Y)$:

$$p(A, \omega) = \int_A f_{X|Y}(x|Y(\omega))\,dx.$$

Appendice : le cas $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et le théorème de Jirina

Tout repose sur le cas de $\mathbb{R}$ muni de ses boréliens. On dispose alors d'un π–système dénombrable qui engendre la tribu. Cette dénombrabilité est cruciale et permet de montrer le

Théorème *Si* $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *alors* p *existe.*

Démonstration On utilise évidemment le π–système $\pi(\mathbb{R})$. Pour tout rationnel r, soit $q(r, \cdot)$ une version de $P(]-\infty, r]|\mathcal{G})$. En éliminant un négligeable, on peut rendre $q(\cdot, \omega)$ croissante, càd, tendant vers 0 en $-\infty$ et vers 1 en $+\infty$. En effet, soit

$$A_{r,s} = \{q(r, \cdot) < q(s, \cdot)\}, \; B_r = \{\lim_n q(r + 1/n, \cdot) \neq q(r, \cdot)\},$$

ainsi que $C_{\pm} = \{\lim_{n\to\pm\infty} q(n, \cdot) \neq 1 \text{ ou } 0\}$, et posons

$$N = \bigcup_{r>s} A_{r,s} \cup \bigcup_r B_r \cup C_- \cup C_+$$

où r et s parcourent $\mathbb{Q}$. Alors N est négligeable et on définit m sur $\mathbb{R} \times \Omega$ par

$$m(x, \omega) = \lim_{r\to x+} q(r, \omega)$$

si $\omega \notin N$ et par $m(x, \omega) = g(x)$ si $\omega \in N$ où g est une fonction de répartition quelconque. Alors, pour tout $\omega \in \Omega$, la fonction $m(\cdot, \omega)$ est croissante, càd, tend vers 1 ou 0 en $\pm\infty$ donc elle correspond à une mesure de probabilité notée $p(\cdot, \omega)$. Ainsi, $p(A, \omega)$ est une version de $P(A|\mathcal{G})$ pour tout $A \in \pi(\mathbb{R})$. Par un théorème de classe monotone, $p(A, \cdot)$ est $\mathcal{G}$–mesurable pour tout $A \in \mathcal{B}(\mathbb{R})$ car $\mathcal{B}(\mathbb{R})$ est la tribu engendrée par $\pi(\mathbb{R})$. Soit $B \in \mathcal{G}$. Alors

$$P(A \cap B) = \int_B p(A, \cdot)\, dP$$

pour tout $A \in \pi(\mathbb{R})$ donc (Dynkin) pour tout $A \in \mathcal{B}(\mathbb{R})$. Finalement,

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \forall B \in \mathcal{G}, \quad P(A \cap B) = \int_B p(A, \cdot)\, dP.$$

■

Définition *Un espace* $(\Omega, \mathcal{F})$ *est standard s'il existe une injection bi–mesurable* f *de* $(\Omega, \mathcal{F})$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *telle que* $f(\Omega) \in \mathcal{B}(\mathbb{R})$.

Théorème 3.1 (Jirina) *Si* $(\Omega, \mathcal{F})$ *est standard, alors* p *existe.*

On peut donner une idée de l'extension du champ d'application de ce théorème en rappelant que si Ω est un espace polonais, $(\Omega, \mathcal{B}(\Omega))$ est standard : un espace topologique Ω est polonais si sa topologie est métrisable par une distance faisant de Ω un espace complet et séparable (c'est-à-dire possédant une suite dense).

Démonstration Soit q une version régulière de $f(P)$ sachant $f(\mathcal{G})$ (la tribu image réciproque de $\mathcal{G}$ par f^{-1}). Posons $p(A, \omega) = q(f(A), f(\omega))$. Alors, pour ω fixé, $p(\cdot, \omega)$ est une probabilité sur $f^{-1}(f(\Omega)) = \Omega$ et pour $A \in \mathcal{F}$ et $B \in \mathcal{G}$, on sait que $f(A) \in \mathcal{B}(\mathbb{R})$ et $f(B) \in f(\mathcal{G})$ donc

$$f(P)(f(A) \cap f(B)) = \int_{f(B)} q(f(A), x)\, f(P)(dx).$$

Le changement de variable $\omega = f^{-1}(x)$ montre que l'on a bien :

$$P(A \cap B) = \int_B p(A, \omega)\, P(d\omega).$$

■

Nous indiquons l'idée de la preuve du fait qu'un espace polonais muni de la tribu de ses boréliens est standard.

Démonstration 1) Soit (Ω, d) métrique complet séparable. Alors $d' = d/(1+d)$ induit la même topologie que d donc on peut supposer $d \leqslant 1$.
2) La topologie de la convergence simple sur $\Omega_0 = [0,1]^{\mathbb{N}}$ correspond à la métrique

$$d_0(x,y) = \sum_{n \geqslant 0} 2^{-n} \, |x_n - y_n|.$$

3) Soit $(\omega_n)_{n\geqslant 0}$ une suite de Ω dense dans Ω et $f : \Omega \to \Omega_0, \omega \mapsto f(\omega) = (d(\omega, \omega_n))_{n \geqslant 0}$. Alors f est injective et bi–continue.
4) L'ensemble $f(\Omega)$ est un borélien de Ω_0.
5) Soit $S = \{0,1\}$ et $\mathcal{S} = \mathcal{P}(S)$. Il existe $g : ([0,1[, \mathcal{B}([0,1[)) \to (S^\infty, \mathcal{S}^\infty)$ telle que g est bi–mesurable et $g([0,1[) \in \mathcal{S}^\infty$.
6) Par conséquent, il existe $h : ([0,1], \mathcal{B}([0,1])) \to (S^\infty, \mathcal{S}^\infty)$ bijective bi–mesurable.
7) Si on se donne une suite d'espaces $((\Omega_n, \mathcal{F}_n))_n$ telle que chaque $(\Omega_n, \mathcal{F}_n)$ est équivalent à $(S_n, \mathcal{S}_n)$, alors le produit $(\prod \Omega_n, \otimes \mathcal{F}_n)$ est équivalent au produit $(\prod S_n, \otimes \mathcal{S}_n)$.
8) Enfin, $(\mathcal{B}([0,1]))^\infty = \mathcal{B}([0,1]^\infty)$ car $[0,1]$ possède un π–système générateur dénombrable. Donc $(\Omega, \mathcal{F})$ est équivalent à un borélien de $([0,1], \mathcal{B}([0,1]))$. ∎

Appendice : un contrexemple

Soit $\Omega = [0,1]$, $\mathcal{G} = \mathcal{B}(\Omega)$ et m la mesure de Lebesgue. Fixons $E \subset \Omega$ qui n'est pas Lebesgue mesurable et tel que $m_*(E) = 0$ et $m^*(E) = 1$. Posons alors $\mathcal{F} = \sigma(\mathcal{G}, E)$ et prolongeons m en une probabilité P sur $\mathcal{F}$ en posant $P(E) = 1/2$. Ainsi, si $A \in \mathcal{F}$, il existe deux boréliens A_1 et A_2 tels que A est la réunion de $A_1 \cap E$ et de $A_2 \cap E^c$. On pose simplement

$$P(A) = (m(A_1) + m(A_2))/2.$$

Dans ces conditions, il n'existe pas de version régulière de P sachant $\mathcal{G}$.

Démonstration Soit p une version régulière et $D = \{p(E, \cdot) = 1\}$. On va montrer que D^c est négligeable. Ainsi, on aura $P(E \cap D) = P(E) = 1/2$ mais

$$\int_D p(E, \omega) \, dP(\omega) = 1.$$

Tout d'abord, D est mesurable pour la tribu engendrée par $P(E|\mathcal{G})$ qui est $\mathcal{G}$ mesurable donc $D \in \mathcal{G}$. La tribu $\mathcal{G}$ des boréliens est engendrée par un π–système dénombrable $\pi = \{B_n \,;\, n \geqslant 1\}$. Posons

$$C_n = \{\omega \in \Omega \,;\, P(B_n|\mathcal{G})(\omega) = \mathbf{1}_{B_n}(\omega)\}.$$

Alors, $P(C_n) = 1$ donc l'intersection C des C_n vérifie $C \in \mathcal{G}$ et $P(C) = 1$. Soit $\omega \in C$: pour tout n, $p(B_n, \omega) = \mathbf{1}_{B_n}(\omega)$ donc tout $B \in \pi$ vérifie $p(B, \omega) = \mathbf{1}_B(\omega)$, donc tout $B \in \mathcal{G}$ aussi. On en déduit que si $\omega \in C$, $p(\{\omega\}, \omega) = 1$. Si $\omega \in E \cap C$, on voit que :

$$p(E, \omega) \geqslant p(\{\omega\}, \omega) = 1.$$

Finalement, $E \cap C \subset D$ donc $E \subset C^c \cap D$ donc $m(D) = P(D) = P(C^c \cap D) = 1$. ∎

3.6 Conditionnement

On va généraliser la définition de la version régulière d'une probabilité par rapport à une tribu. Le premier pas est de remplacer cette tribu par une variable aléatoire. On veut donc décomposer P selon les valeurs de Y.

Définition *Soit P une probabilité sur $\mathcal{F}$ et Y une variable aléatoire. On appelle conditionnement de P par rapport à Y une application $q : \mathcal{F} \times \mathbb{R} \to [0,1]$ telle que*
(cp1) pour tout $y \in \mathbb{R}$, $q(\cdot, y)$ est une probabilité sur $\mathcal{F}$;
(cp2) pour tout $A \in \mathcal{F}$, $q(A, Y) = P(A|Y)$ p.s.

On a alors, pour tout $y \in \mathbb{R}$, $q(\{Y \neq y\}, y) = 0$, et

$$P = \int q(\cdot, y)\, P_Y(dy),$$

et ces deux propriétés suffisent à caractériser la famille $(q(\cdot, y))_y$.

Définition *Soit X et $Y : \Omega \to \mathbb{R}$ deux variables aléatoires. La loi de X sachant Y est une application $q_{X|Y} : \mathcal{B}(\mathbb{R}) \times \mathbb{R} \to [0,1]$ telle que*
(lc1) pour tout $y \in \mathbb{R}$, $q_{X|Y}(\cdot, y)$ est une probabilité sur $\mathcal{B}(\mathbb{R})$;
(lc2) pour tout $B \in \mathcal{B}(\mathbb{R})$, $q_{X|Y}(B, Y) = P(X \in B|Y)$ p.s.

On a alors

$$P_X = \int q(\cdot, y)\, P_Y(dy).$$

Si $q_{X|Y}$ est la loi de X sachant Y, alors $q_{X|Y} \circ (X \otimes Id_{\mathbb{R}})$ est un conditionnement de P par rapport à Y.

Exercice Montrer que X et Y sont indépendantes si et seulement si $q : (B, y) \mapsto P_X(B)$ est un conditionnement de X sachant Y.

Définition *Soit $\mathcal{G}$ une sous–tribu de $\mathcal{F}$ et $X : \Omega \to \mathbb{R}$ une variable aléatoire. La loi de X sachant $\mathcal{G}$ est une application $q_X : \mathcal{B}(\mathbb{R}) \times \Omega \to [0,1]$ telle que*
(tc1) pour tout ω, $q_X(\cdot, \omega)$ est une probabilité sur $\mathcal{B}(\mathbb{R})$;
(tc2) pour tout $B \in \mathcal{B}(\mathbb{R})$, $q_X(B, \cdot) = P(X \in B|\mathcal{G})$ p.s.

On a alors : $P_X = \int q_X(\cdot, \omega)\, P(d\omega)$.
Si $q_{X|Y}$ est la loi conditionnelle de X sachant Y, alors $q_{X|Y} \circ (Id_{\mathcal{B}(\mathbb{R})} \otimes Y)$ est la loi conditionnelle de X sachant $\mathcal{G} = \sigma(Y)$.

$$\begin{tikzcd}
\mathcal{B}(\mathbb{R}) \times \Omega \arrow[d, "Id \otimes Y"'] \arrow[dr, "P_X|\sigma(Y)"] & \\
\mathcal{B}(\mathbb{R}) \times \mathbb{R} \arrow[r, "P_X|Y"] & {[0,1]} \\
\mathcal{F} \times \mathbb{R} \arrow[u, "X \otimes Id"] \arrow[ur, "P|Y"'] &
\end{tikzcd}$$

Exercice Soit q un conditionnement de P sachant Y et $B \in \mathcal{B}(\mathbb{R})$ tel que $P(Y \in B) > 0$. Montrer que tout $A \in \mathcal{F}$ vérifie :

$$P(A|Y \in B) = (P(Y \in B))^{-1} \int_B q(A, y)\, P_Y(dy).$$

Montrer que les deux propriétés suivantes suffisent à assurer que q est un conditionnement de P sachant Y si l'on suppose que $y \mapsto q(A, y)$ est borélienne pour tout $A \in \mathcal{F}$:
(cp3) presque sûrement en y pour P_Y, $q(\cdot, y)$ est portée par $\{Y = y\}$;
(cp4) $P = \int q(\cdot, y)\, P_Y(dy)$.

Exercice On suppose que Ω est un produit, que P admet une densité f par rapport à une mesure produit $\mu \otimes \nu$ et que Y est la projection sur la deuxième coordonnée $Y(x, y) = y$. La loi de Y admet alors une densité g par rapport à ν. Soit

$$N = \{y\,;\, g(y) = 0 \text{ ou } +\infty\}.$$

Montrer que $P_Y(N) = 0$. Soit r définie sur $y \notin N$ par

$$r(dx, y) = (f(x, y)/g(y))\, \mu(dy),$$

et par $r(\cdot, y) = r_0(\cdot)$ pour $y \in N$ pour r_0 une probabilité quelconque. Montrer que $q(\cdot, y) = r(\cdot, y) \otimes \delta_y$ est un conditionnement de P sachant Y.

Exercice Soit $q_{X|X}$ une loi conditionnelle de X sachant X. Montrer que

$$q_{X|X}(\cdot, x) = \delta_x \quad P_X\text{–p.s.}$$

Soit $q_{X|Y}$ une loi conditionnelle de X sachant Y. Montrer que

$$q_2(\cdot, y) = q_{X|Y}(\cdot, y) \otimes \delta_y$$

est une loi conditionnelle de (X, Y) sachant Y.

Une façon de reformuler l'indépendance conditionnelle est la

Définition *Soit X et $Y : \Omega \to \mathbb{R}$ des variables aléatoires et $\mathcal{G}$ une sous–tribu. Supposons qu'il existe des lois conditionnelles de X et Y sachant $\mathcal{G}$, notées q_X et q_Y. On dit que X et Y sont indépendantes conditionnellement à $\mathcal{G}$ si*

$$q_2(B_1 \times B_2, \omega) = q_X(B_1, \omega)\, q_Y(B_2, \omega)$$

fournit une loi conditionnelle de (X, Y) sachant $\mathcal{G}$.

Exercice X et Y sont indépendantes conditionnellement à $\sigma(Z)$ si et seulement si

$$q_{X|(Y,Z)}(\cdot, (y, z)) = q_{X|Z}(\cdot, z)$$

P_Y–presque sûrement.

3.7 Le cas gaussien

Les variables aléatoires gaussiennes

Définition *Loi gaussienne réelle centrée réduite :*

$$\mathcal{N}(0,1)(dx) = \exp(-x^2/2)\,dx/\sqrt{2\pi}.$$

Loi gaussienne réelle : $\mathcal{N}(m,\sigma^2) = m + \sigma\,\mathcal{N}(0,1)$ *pour* $m \in \mathbb{R}$ *et* $\sigma \in \mathbb{R}$.
Famille gaussienne : $X = (X_i)_i$ *est gaussienne si et seulement si pour toute partie finie* J *de* I *et tous réels* $(a_j)_j$*, la variable aléatoire* $\sum_{j\in J} a_j\,X_j$ *est gaussienne (réelle).*

Exemple Soit X et Y deux variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. Les couples suivants sont gaussiens : (X,Y) ; $(0,0)$; $(X,0)$; (X,Y) ; $(X,2X+1)$; $(X+Y,X-Y)$.
La loi de $aX+bY$ est $\mathcal{N}(0,a^2+b^2)$.
[Faire le calcul.]
Soit $\varepsilon = \pm 1$ indépendant de X. Alors, X et $\varepsilon\,X$ sont deux variables aléatoires gaussiennes mais $(X,\varepsilon\,X)$ n'est pas gaussien. (preuve ?)
La loi de $\varepsilon\,X$ sachant X est $\frac{1}{2}(\delta_X + \delta_{-X})$.

Quelques propriétés :

Soit X de loi $\mathcal{N}(m,\sigma^2)$. Alors X est de carré intégrable et

$$E(X) = m, \quad \mathrm{var}(X) = \sigma^2, \quad E(e^{i\,t\,X}) = \exp(i\,t\,m - \sigma^2\,t^2/2).$$

Soit X un vecteur gaussien de moyenne M et de covariance C :

$$\begin{aligned} M = E(X), \qquad C = E((X-M)\,(X-M)^*),\\ E(\exp(i < t, X >)) = \exp(i < t, M > -t^*Ct/2). \end{aligned}$$

En particulier, quand $C > 0$, on connaît la loi de X. Elle admet une densité :

$$P_X(dx) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp(-\frac{1}{2}(x-M)^* C^{-1}(x-M))\,dx.$$

La loi de X est déterminée par M et C. On note cette loi $\mathcal{N}(M,C)$. Soit A une matrice de taille $d \times d'$ et B un vecteur constant de $\mathbb{R}^{d'}$. Alors $A \cdot X + B$ est un vecteur gaussien de $\mathbb{R}^{d'}$ de loi

$$\mathcal{N}(AM+B, ACA^*).$$

Réciproquement, soit $M \in \mathbb{R}^d$, C une matrice $d\times d$ symétrique positive, A une matrice $d\times d$ telle que $AA^* = C$ et X un vecteur gaussien de loi $\mathcal{N}(0,I_d)$. Alors, $Y = AX + M$ suit la loi $\mathcal{N}(M,C)$. Par conséquent, on peut réaliser toutes les lois gaussiennes (puisque $\mathcal{N}(0,I_d)$ est réalisée par un d-uplet de variables aléatoires réelles indépendantes de lois $\mathcal{N}(0,1)$) et la loi $\mathcal{N}(M,C)$ admet pour support $M + \mathrm{Im}(A) = M + \mathrm{Im}(C)$.

Démonstration Astuce du carré : $\int_{-\infty}^{+\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$.
Soit Y de loi $\mathcal{N}(O,I_n)$ et $Z = M + AY$ où A est une racine carrée de C c'est-à-dire que $C = AA^*$. Alors Z suit la loi de X dès que l'on connaît la valeur de $E(\exp(i < t, X >))$. Or, la variable aléatoire réelle $T =< t, X >$ est gaussienne donc il suffit d'identifier $E(T) =< t, E(X) >$ et la variance de T soit

$$\begin{aligned} E(< t, X - M >^2) &= E([t^*(X-M)]\,[t^*(X-M)]^*)\\ &= E(t^*(X-M)(X-M)^*t) = t^*\,C\,t. \end{aligned}$$

Par injectivité de la transformée de Fourier, on a terminé. ∎

Indépendance et orthogonalité

Proposition *Soit X un vecteur gaussien. Les X_k pour $k \leqslant n$ sont indépendants si et seulement si C est diagonale :*

$$(X_k)_k \text{ indépendantes} \quad \Leftrightarrow \quad \forall k \neq l,\ \operatorname{cov}(X_k, X_l) = 0.$$

Démonstration L'orthogonalité entraîne que C est diagonale, donc

$$E(\exp(i < t, X >)) = \prod_{k=1}^{n} E(\exp(i\, t_k\, X_k)).$$

Par injectivité de la transformée de Fourier sur $\mathbb{R}^n$, on en déduit que P_X est aussi $\otimes_k P_{X_k}$ donc l'indépendance. Indépendance implique orthogonalité (trivial). ■

Suite du contrexemple important : si X est de loi $\mathcal{N}(0,1)$ et $\varepsilon = \pm 1$ indépendant de X, alors $(X, \varepsilon\, X)$ est orthogonal mais pas indépendant.

Conditionnement et lois gaussiennes

Théorème *Soit (X, Y) un vecteur gaussien tel que Y ne soit pas dégénéré. Alors la loi de X sachant Y est gaussienne $\mathcal{N}(M(Y), \Sigma^2)$. On connaît ses paramètres :*

$$\begin{aligned} M(Y) &= E(X) + \operatorname{cov}(X,Y)\operatorname{cov}(Y)^{-1}(Y - E(Y)), \\ \Sigma^2 &= \operatorname{cov}(X) - \operatorname{cov}(X,Y)\operatorname{cov}(Y)^{-1}\operatorname{cov}(Y,X), \end{aligned}$$

où rappelons que l'on note $\operatorname{cov}(a,b)$ la matrice $E((a - E(a))\,(b - E(b))^*)$ et $\operatorname{cov}(a) = \operatorname{cov}(a,a)$. On peut remarquer que $\Sigma^2 < \operatorname{cov}(X)$ (au sens des matrices symétriques ; pourquoi est-ce normal ?) et que Σ^2 ne dépend pas de Y.

Démonstration On note Z le vecteur formé de X et Y, et on suppose tout d'abord que $E(Z) = 0$ et que Z est non dégénérée. Soit $C = \operatorname{cov}(Z)$, $A = \operatorname{cov}(X) = E(X\,X^*)$, $B = \operatorname{cov}(Y) = E(Y\,Y^*)$ et $D = \operatorname{cov}(X,Y) = E(X\,Y^*)$. On note E, F et $-G$ les composantes de C^{-1}. Donc :

$$C = \begin{pmatrix} A & D \\ D^* & B \end{pmatrix}, \qquad C^{-1} = \begin{pmatrix} E & -G \\ -G^* & F \end{pmatrix}.$$

Toutes les lois ont des densités donc il s'agit de calculer $f_{(X,Y)}(x,y)/f_Y(y)$. On trouve une expression de la forme $c\,\exp(-g(x|y)/2)$ avec :

$$\begin{aligned} g(x|y) &= x^*Ex + y^*Fy - x^*Gy - y^*G^*x - y^*B^{-1}y, \\ &= (x - E^{-1}Gy)^*E(x - E^{-1}Gy) + y^*Hy, \end{aligned}$$

avec $H = F - B^{-1} - G^*E^{-1}G$. En effet, B est inversible car C est inversible, de même pour E. (Par exemple, si $y^*By = 0$, on a $z^*Cz = 0$ pour $z^* = (0, y^*)$, donc $y = 0$ car C est inversible.)

Il reste à montrer que H est nulle et à identifier $M(y) = My$ avec $M = E^{-1}G$ et $\Sigma^2 = E^{-1}$ en fonction des coefficients de C. Pour cela, on effectue le produit $C\,C^{-1} = I$ et on obtient :

$$AE = DG^* + I, \quad GB = ED, \quad AG = DF, \quad BF = D^*G + I.$$

Les deux premières relations donnent $G = EDB^{-1}$ et $(A - DB^{-1}D)E = I$ (rappelons que A, B, E et F sont symétriques), donc

$$\Sigma^2 = E^{-1} = A - DB^{-1}B^* \quad \text{et} \quad M = DB^{-1}.$$

On montre que $H = 0$ en utilisant les deux autres relations, comme suit :

$$BH = BF - I - BG^*E^{-1}G = (D^*E - BG^*)E^{-1}G = (ED - GB)^*E^{-1}G = 0.$$

On suppose maintenant que $E(Z) = 0$ et $C \geqslant 0$. On introduit un vecteur U de loi $\mathcal{N}(0, I)$, de la taille de X, et indépendant de (X, Y), et on pose $X_0 = X + U$ et $Z_0^* = (X_0^*, Y^*)$. Alors $C_0 = \mathrm{cov}(Z_0)$ s'écrit par blocs comme C avec $A_0 = A + I$, B et D étant inchangés. En appliquant ce qui précède à Z_0, on sait que, pour tout t,

$$E(\exp(it^*X_0) \,|\, Y) = \exp(it^*MY - t^*\Sigma_0^2 t/2),$$

avec $\Sigma_0^2 = A_0 - DB^{-1}B^* = \Sigma^2 + I$. Or, $\exp(it^*X_0) = \exp(it^*X)\exp(it^*U)$ et $\exp(it^*U)$ est indépendant de Y et de moyenne $\exp(-t^*t/2)$, donc

$$E(\exp(it^*X_0) \,|\, Y) = E(\exp(it^*X) \,|\, Y)\ \exp(-t^*t/2).$$

Il vient $E(\exp(it^*X) \,|\, Y) = \exp(it^*MY - t^*\Sigma^2 t/2)$, c'est-à-dire le résultat. Enfin, si Z n'est pas centrée, ajouter $E(X)$ à $X - E(X)$ ajoute une constante à la loi conditionnelle, et ajouter $E(Y)$ à $Y - E(Y)$ ne change rien. ■

Remarque : On a posé $f_Y(y) = c_Y\ \exp(-\frac{1}{2}\, y^*\, B^{-1}\, y)$, et

$$f_{X,Y}(x, y) = c_{X,Y}\ \exp\left(-\tfrac{1}{2}\, z^*\, C^{-1}\, z\right) \quad \text{avec} \quad z = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Appendice : l'inégalité de Gebelein

Dym–McKean, chapitre 3.

Omis.

Proposition *Soit G un sous–espace gaussien de $L^2 = L^2(\mathcal{F})$. Deux sous–espaces vectoriels de G sont indépendants si et seulement si ils sont orthogonaux pour le produit scalaire*

$$< X, Y >= \mathrm{cov}(X, Y).$$

Proposition *Soit G gaussien et A un sous–espace vectoriel de G. Alors $Y \in G$ est mesurable pour la tribu engendrée par les variables aléatoires de A si et seulement si $Y \in A$. Soit p_A la projection orthogonale sur A et soit $Y \in G$. On a :*

$$p_A(Y) = E(Y|\sigma(A)).$$

Soit A et B deux sous–espaces fermés de G. On pose

$$\cos(A, B) = \sup\{\mathrm{cov}(X, Y)\,;\, X \in A,\, Y \in B,\, \mathrm{var}(X) \leqslant 1,\, \mathrm{var}(Y) \leqslant 1\}.$$

Théorème 3.2 (Kolmogorov–Rozanov (1960))
Notons A_∞ l'ensemble des $X \in L^2$ $\sigma(A)$–mesurables tels que $E(X) = 0$. Alors,

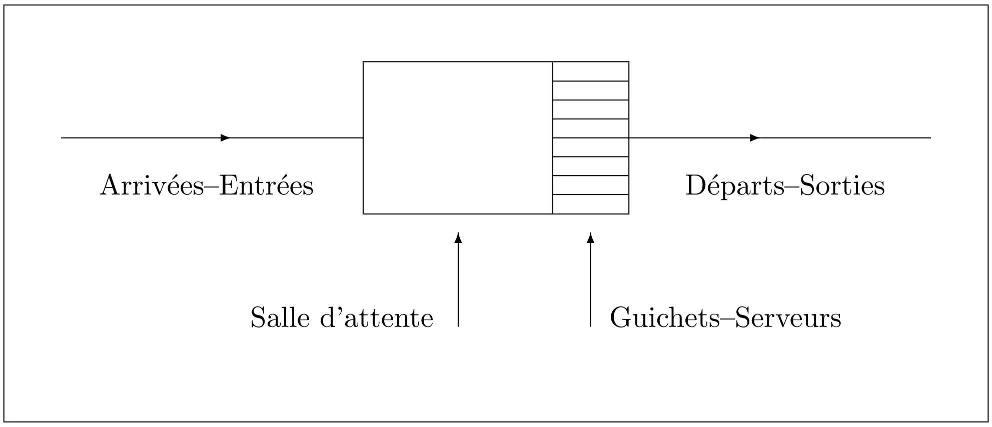
$$\cos(A_\infty, B_\infty) = \cos(A, B).$$

Lemme 3.3 (inégalité de Gebelein)
Soit (X, Y) un vecteur gaussien et f une fonction mesurable. Alors,

$$\mathrm{var}(X)\,\mathrm{var}(Y)\, E(E(f(Y)|X)^2) \leqslant \mathrm{cov}(X, Y)^2\, E(f(Y)^2).$$

3.8 Introduction aux files d'attente

On va étudier un modèle très général de problème de gestion : stocks, temps de service, travail partagé. . .Pour cela on considère un système de la forme suivante :



On veut évaluer des quantités comme le temps d'attente d'un client, la longueur de la queue qu'il a dû faire avant d'être servi, les surcharges éventuelles du système. . .

Le flot d'arrivée

Trois descriptions possibles et équivalentes d'un flot ϕ :
– les instants d'arrivée des clients : une suite $(T_n)_{n\geqslant 0}$ de variables aléatoires positives avec $T_0 = 0$ et $T_n \leqslant T_{n+1}$;
– les durées inter–arrivées : une suite $(D_n)_{n\geqslant 1}$ de variables aléatoires positives avec $D_n = T_n - T_{n-1}$;
– un processus de comptage $(N_t)_{t\geqslant 0}$ relié aux autres descriptions par l'égalité

$$\{N_t = n\} = \{T_n \leqslant t < T_{n+1}\}.$$

Définition *Un flot sans mémoire est un flot où les $(D_n)_n$ sont i.i.d. de loi exponentielle $\mathcal{E}(a)$.*

Proposition *Soit D une variable aléatoire de loi exponentielle et S une variable aléatoire $S \geqslant 0$ indépendante de D. Alors la loi de $D - S$ sachant $D \geqslant S$ est la loi de D.*

Démonstration Si S est une constante, c'est immédiat. Sinon, on intègre. ■

Soit ϕ un flot et $s \geqslant 0$ un réel. Le flot ϕ^* restreint après s est donné, sur l'événement $\{N_s = n\}$ par les durées inter–arrivées suivantes :

$$D_1^* = T_{n+1} - s, \qquad D_k^* = D_{n+k}.$$

Théorème *Soit $S \geqslant 0$ une variable aléatoire indépendante d'un flot sans mémoire. Alors le flot restreint après S suit la même loi que le flot original, indépendante de S.*

Démonstration Supposons que S est constante et calculons la loi de D_1^* :

$$P(D_1^* \geqslant t \,|\, N_S = n) = P(T_{n+1} - S \geqslant t | T_n \leqslant S < T_{n+1}).$$

Notons $T = S - T_n$. Alors, on sait que :

$$P(T_{n+1} - S \geqslant t | T_n \leqslant S < T_{n+1}) = P(D_{n+1} \geqslant t + T | 0 \leqslant T < D_{n+1}),$$

donc si T était toujours positif, on aurait fini. On va remplacer T par une variable aléatoire positive. Pour $c > 0$, soit $T^c = T$ si $T \geqslant 0$ et $T^c = c$ si $T < 0$. Alors, T^c dépend de S et de T_n uniquement donc T^c et D_{n+1} sont indépendants. On a :

$$\begin{aligned}
\{D_{n+1} \geqslant T^c\} &= \{D_{n+1} \geqslant T,\, T \geqslant 0\} \cup \{D_{n+1} \geqslant c,\, T < 0\},\\
\{D_{n+1} \geqslant T^c + t\} &= \{D_{n+1} \geqslant T + t,\, T \geqslant 0\} \cup \{D_{n+1} \geqslant c + t,\, T < 0\}.
\end{aligned}$$

Quand c tend vers l'infini,

$$P(D_{n+1} \geqslant T^c + t | D_{n+1} \geqslant T^c) \longrightarrow P(D_{n+1} \geqslant T + t | 0 \leqslant T < D_{n+1}).$$

La première quantité ne dépend pas de c et vaut e^{-at}. Finalement, D_1^* suit bien une loi $\mathcal{E}(a)$. En réalité, on sait même que la loi conditionnelle de D_1^* sachant N_S est $\mathcal{E}(a)$. Le même raisonnement fonctionne si S est une variable aléatoire.

Ensuite, T_n, S et D_{n+1} sont indépendants de la suite des D_k pour $k \geqslant n+2$ donc

$$\begin{aligned}
P(D_1^* \geqslant t_1, \ldots, D_k^* \geqslant t_k) &= \sum_{n=0}^{+\infty} P(D_{n+1} \geqslant t_1 + T, D_{n+2} \geqslant t_2, \ldots\\
&\qquad \ldots, D_{n+k} \geqslant t_k, 0 \leqslant T < D_{n+1})\\
&= \prod_{i=2}^{k} e^{-at_i} \cdot \sum_{n=0}^{+\infty} P(D_{n+1} \geqslant t_1 + T, 0 \leqslant T < D_{n+1})\\
&= \prod_{i=2}^{k} e^{-at_i} \cdot P(D_1^* \geqslant t_1),
\end{aligned}$$

d'où l'indépendance et les lois désirées. ■

Proposition *Soit $(N_t)_t$ le processus de comptage d'un flot sans mémoire. Alors $(N_t)_t$ est poissonnien :*
(P1) N_t suit une loi $\mathcal{P}(at)$;
(P2) $N_{t+s} - N_s$ suit la loi de N_t ;
(P3) $N_t - N_s$ est indépendant de N_s pour tout $s \leqslant t$.

Démonstration Pour $S = s$, on sait que $N_{t+s} - N_s = N_t^*$ donc (P2) et (P3). Pour (P1), on écrit :

$$P(N_t = n) = P(T_n \leqslant t < T_{n+1}),$$

puis on connaît la loi de $(D_n)_n$ donc celle de T_n et T_{n+1} donc on peut calculer l'intégrale, qui donne $e^{-at}\,(at)^n/n!$ ■

Nous admettrons la preuve d'une réciproque, qui s'énonce comme suit :

Proposition *Soit $(N_t)_t$ un processus vérifiant $N_0 = 0$, (P2), (P3) et $P(N_t \geqslant 2) = o(t)$ quand t tend vers* 0. *Alors, $(N_t)_t$ est un processus poissonnien.*

Exercice Dans ce cadre poissonnien, la loi de T_n est de densité $e^{-at}(at)^{n-1}/(n-1)!$ sur $\mathbb{R}_+$.

Processus de vie et de mort

Définition *Le processus $(N_t)_t$ est un processus de vie et de mort si N_t prend des valeurs entières et s'il existe deux suites $(a_n)_n$ et $(b_n)_n$ positives telles que :*
$P(N_{t+h} = n+1 | N_t = n) = a_n\, h + o(h)$,
$P(N_{t+h} = n-1 | N_t = n) = b_n\, h + o(h)$,
$P(|N_{t+h} - n| \geqslant 2 | N_t = n) = o(h)$, *avec $b_0 = 0$ et des $o(h)$ uniformes en n.*

Exemple Soit N_t la taille à l'instant t d'une population d'individus indépendants. Entre t et $t+dt$, chacun peut mourir avec une probabilité de $\theta\, dt$, ou se scinder en deux avec une probabilité de $\zeta\, dt$. Alors, $a_n = n\zeta$ et $b_n = n\theta$ conviennent.

Exemple Soit N_t le nombre de clients dans un service (donc dans la queue ou aux guichets) dont le flot des arrivées est sans mémoire (donc des variables aléatoires de loi $\mathcal{E}(a)$), dont le flot des services est sans mémoire (donc des variables aléatoires de loi $\mathcal{E}(b)$) et doté de k guichets (service M/M/k). Alors $a_n = a$ et $b_n = (n \wedge k)\, b$ conviennent.

Proposition *La probabilité $p_n(t) = P(N_t = n)$ vérifie l'équation de Kolmogorov :*

$$p'_n = a_{n-1}\, p_{n-1} + b_{n+1}\, p_{n+1} - (a_n + b_n)\, p_n.$$

On appelle régime permanent une solution constante de ces équations : $p_n(t) = p_n$. L'intérêt de cette notion est que l'on est souvent capable de montrer que la fonction $p_n(t)$ tend vers le régime permanent p_n quand t tend vers l'infini. Dans le cas où les a_n, b_n sont tous strictements positifs (cas sans barrière) on trouve

$$p_n = p_0 \prod_{k=1}^{n} \frac{a_{k-1}}{b_k}.$$

Régime permanent si et seulement si la somme ci-dessous est finie :

$$\sum_{n \geqslant 1} \prod_{k=1}^{n} \frac{a_{k-1}}{b_k}.$$

Exemple M/M/k. La condition est $a/b < k$.
M/M/∞. Pas de condition. Régime permanent : $p_n = e^{-a/b}\, (a/b)^n / n!$
Cas général : un régime permanent est possible sur chaque intervalle inter–barrières, c'est-à-dire chaque composante connexe de $a_n > 0$, $b_{n+1} > 0$.

Etude du système M/M/1

On note $r = a/b$ l'intensité du trafic et on suppose que $r < 1$. Le régime permanent est alors $p_n = r^n(1-r)$. Soit N^Q le nombre de clients dans la queue, N^G le nombre de clients au guichet et $N = N^Q + N^G$ le nombre total de clients dans le système. Alors

$$E(N) = \sum_{n \geqslant 0} n p_n = r/(1-r), \qquad E(N^G) = \sum_{n \geqslant 1} p_n = r.$$

Soit T le temps total d'attente d'un client dans le système : $T = T^Q + T^G$ et par hypothèse, T^G est $\mathcal{E}(b)$ donc $E(T^G) = b^{-1}$. Sur $\{N = n\}$, T^Q est somme de $n-1$ variables aléatoires indépendantes de loi $\mathcal{E}(b)$ et d'une variable aléatoire $\mathcal{E}(b) - t_0$ conditionnée par $\{\mathcal{E}(b) \geqslant t_0\}$ donc de n variables aléatoires indépendantes de loi $\mathcal{E}(b)$. Si $N = 0$, $T^Q = 0$ donc

$$E(T^Q) = 0\,(1-r) + \sum_{n \geqslant 1} n\, b^{-1}\, r^n\,(1-r) = r/(b(1-r)).$$

De même, la loi de T^Q est $(1-r)\,\delta_0 + r\,(b-a)\,\exp(-(b-a)t)\,\mathbf{1}(t>0)\,dt$, d'où la

Proposition 3.4 (loi de Little) *Le rapport $E(T)/E(N) = a^{-1}$ ne dépend pas de la rapidité b de service au guichet.*

Notation : On note un service $xx/yy/zz$ où xx désigne le flot des arrivées, yy celui des services et zz le nombre de guichets. Un flot markovien est noté M, un flot déterministe D et le cas général G.

Etude du système M/G/1

Rappelons que $r = a/b$. Les arrivées sont séparées par des D_n de loi $\mathcal{E}(a)$ et les temps de service sont des X_n avec $E(X) = b^{-1}$ et $\mathrm{var}(X) = \sigma^2$. Alors le temps T_n passé dans la queue par le n–ième client est donné par

$$T_{n+1} = (T_n + X_n - D_n)^+.$$

On s'intéresse au régime stationnaire. Si T_n et T_{n+1} ont la même loi, on a, pour $c \geqslant 0$:

$$E(e^{-cT}) = \int_0^{+\infty} P_T(dt)\,E(e^{-c(t+X-D)^+}).$$

On introduit donc la fonction $f(s) = E(e^{-c(s-D)^+})$, calculable aisément puisque l'on connaît la loi de D : $f(s) = (a\,e^{-cs} - c\,e^{-as})/(a-c)$. On obtient :

$$\begin{aligned}(a-c)\,E(e^{-cT}) &= (a-c)\,E(f(T+X)) \\ &= a\,E(e^{-cX})\,E(e^{-cT}) - c\,E(e^{-aX})\,E(e^{-aT}).\end{aligned}$$

Cherchons le terme en c^2 dans le développement en série entière près de 0 de cette égalité :

$$(a/2)\,E(T^2) + E(T) = a((1/2)\,E(T^2) + E(X)\,E(T) + (1/2)\,E(X^2)),$$

soit $E(T) = (a/2)\,E(X^2)/(1 - a\,E(X))$. Avec les notations ci-dessus, on trouve

$$E(T) = \frac{a^2\sigma^2 + r^2}{2a(1-r)}.$$

En particulier, si X est déterministe, $\sigma^2 = 0$ donc

Proposition $E_{M/D/1}(T^Q) = E_{M/M/1}(T^Q)/2$.

Interprétation Dans le cas d'un temps de service en $\mathcal{E}(b)$, quel est le temps gagné ou perdu à cause du n–ième client par rapport au système M/D/1 ? Si $T^G < b^{-1}$, on gagne $b^{-1} - T^G$. Si $T^G > b^{-1}$, on perd

$$T^G - b^{-1} + \sum T_n^Q + T_n^G,$$

où la somme porte sur tous les clients n arrivés pendant un laps de temps de $T^G - b^{-1}$. Au total, on y perd.

Récurrence/transience du système G/G/1

Soit $U_n = X_n - D_n$ et S_n la somme des n premiers U_k. Alors T_n vaut

$$T_n = S_n - \inf_{0 \leqslant k \leqslant n} S_k,$$

donc si $\inf_k S_k$ vaut presque sûrement $-\infty$, alors T_n est nul une infinité de fois. Sinon, T_n est nul un nombre presque sûrement fini (et même intégrable) de fois.
Ensuite, T_n suit la loi de $\sup_{0 \leqslant k \leqslant n} S_k$ donc si $\sup_k S_k$ est infini presque sûrement, on sait que $P(T_n \leqslant t)$ tend vers 0 pour tout t. Sinon, T_n converge en loi vers $\sup_k S_k$.

Démonstration Les U_n sont des variables aléatoires indépendantes de même loi absolument continue par rapport à la mesure de Lebesgue. Soit τ_k l'instant du k–ième passage de $(T_n)_n$ par 0. Alors, $P(\tau_k < +\infty) = P(\tau_1 < +\infty)^k$ et

$$E\left(\sum_{n \geqslant 1} \mathbf{1}(T_n = 0)\right) = (1 - P(\tau_1 < +\infty))^{-1}.$$

∎

Exercices

1 Il y a $d_H = 5\%$ de daltoniens chez les hommes et $d_F = 0,25\%$ chez les femmes. Il y a $p_H = 48\%$ d'hommes et $p_F = 52\%$ de femmes. Probabilité pour qu'un daltonien soit un homme ?

2 Soit X une variable aléatoire de Poisson $\mathcal{P}(a)$. Calculer $E(X)$ et $\mathrm{var}(X)$.
La loi de Y sachant $\{X = n\}$ est binômiale $\mathcal{B}(n,p)$. Loi de Y ? Calculer $E(Y)$ et $\mathrm{var}(Y)$.
Loi de $X - Y$ sachant $\{Y = m\}$? Conclusion ?

3 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi intégrable et S_n la somme des n premiers X_k. On note $\mathcal{G}_n$ la tribu engendrée par les S_k pour $k \geqslant n$. Montrer

$$\forall k \leqslant n, \quad E(X_k|\mathcal{G}_n) = E(X_k|S_n) = S_n/n.$$

4 Pour $i = 1, 2, 3$, on note S_i la somme d'un échantillon de taille n_i. Les trois échantillons sont indépendants et de même loi intégrable. Calculer

$$E(S_1 + S_3|S_2 + S_3).$$

5 Soit X et Y des variables aléatoires indépendantes de lois $\mathcal{B}(n,p)$ et $\mathcal{B}(m,p)$. Que vaut $E(X|X+Y)$? (Sans calcul.)

6 Soit X de loi uniforme sur $[0,1]$; soit Y de loi sachant $\{X = x\}$:

$$(y-x)\, e^{-(y-x)}\, \mathbf{1}_{y \geqslant x}\, dy,$$

et enfin, soit Z de loi sachant $\{X = x, Y = y\}$:

$$(y-x)\, e^{-z\,(y-x)}\, \mathbf{1}_{z \geqslant 0}\, dz.$$

Donner la loi de (X,Y,Z). Donner la loi de Z.
On pose $U = X$, $V = Y - X$ et $W = Z\,(Y-X)$. Donner la loi de (U,V,W). Quelle indépendance remarque-t-on ? Réaliser Y et Z à l'aide de X et de variables aléatoires indépendantes exponentielles. Réaliser (X,Y,Z) à l'aide de quatre variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

7 Le temps d'attente d'un taxi est exponentiel $\mathcal{E}(a)$. Au bout d'une heure, aucun taxi n'est passé. Loi du temps qu'il reste à attendre ?
Simulation : on jette les $\mathcal{E}_i(a)$ tant que $\mathcal{E}_i(a) \geqslant 1$. On garde le premier $\leqslant 1$. Comment retrouver la loi $\mathcal{E}(a)$? Quel est l'intérêt du procédé ?
(On ne simule que sur $[0,1]$, donc la précision est plus grande.)

8 1) Soit $\Omega = \{0 \leqslant y \leqslant x \leqslant 1\}$, P la mesure uniforme et X, resp. Y, la projection sur l'axe des abscisses, resp. des ordonnées. Calculer $E(Y|X)$ et $E(f(Y)|X)$.
2) Soit $\Omega = \{x^2 + y^2 = 1\}$ et (X,Y) de loi uniforme sur Ω. Calculer $E(X^2 + X - 1|Y)$.
3) Soit (X,Y) de loi uniforme sur $\Omega = \{x^2 + y^2 \leqslant 1\}$. Calculer $E(Y^2|X)$.

9 Soit X_k des variables aléatoires indépendantes de loi uniforme sur $[0,1]$ et S_n, resp. I_n, le maximum, resp. le minimum, des X_k pour $k \leqslant n$. Lois de S_n, de I_n, de (S_n, I_n), de S_n sachant que $\{I_n = x\}$. Calculer $E(S_n|I_n)$.

10 Soit X et Y de carré intégrable. Montrer l'inégalité de Cauchy–Schwarz conditionnelle suivante :

$$E(XY|\mathcal{G})^2 \leqslant E(X^2|\mathcal{G})\, E(Y^2|\mathcal{G}).$$

11 Pour $X \in L^2$, on définit la variance conditionnelle de X sachant $\mathcal{G}$ comme

$$\mathrm{var}(X|\mathcal{G}) = E(X^2|\mathcal{G}) - E(X|\mathcal{G})^2.$$

Montrer que $\mathrm{var}(X) = E(\mathrm{var}(X|\mathcal{G})) + \mathrm{var}(E(X|\mathcal{G}))$.

12 Soit $X \geqslant 0$ intégrable. Montrer que $\{E(X|\mathcal{G}) > 0\}$ est le plus petit ensemble $\mathcal{G}$–mesurable qui contient $\{X > 0\}$.

13 Soit X intégrable, Z presque sûrement constante et $B \in \mathcal{F}$ avec $0 < P(B) < 1$. Calculer $E(X|Z)$ et $E(X|\sigma(B))$.

14 Soit X et Y deux variables aléatoires de carré intégrable telles que $E(X|Y) = Y$ p.s. et $E(X^2|Y) = Y^2$ p.s. Montrer que $X = Y$ p.s.

15 Soit Γ un groupe fini de bijections bimesurables de $(\Omega, \mathcal{F})$ qui préservent P et posons

$$\mathcal{G} = \{A \in \mathcal{F}\,;\, \forall \gamma \in \Gamma,\ P(A \triangle \gamma(A)) = 0\}.$$

Montrer que $\mathcal{G}$ est une tribu. Calculer $E(X|\mathcal{G})$ pour X intégrable. On pourra commencer par étudier la tribu $\mathcal{G}' = \{A \in \mathcal{F}\,;\, \forall \gamma \in \Gamma,\ A = \gamma(A)\}$.

16 Soit $(X_k)_{k \leqslant n}$ des variables aléatoires de même loi conditionnellement à $\mathcal{G}$ et conditionnellement indépendantes par rapport à $\mathcal{G}$. Montrer que les X_k sont échangeables.

17 Soit $X \geqslant 0$ pas forcément intégrable. Montrer qu'il existe une variable aléatoire Y presque sûrement unique à valeurs dans $[0, +\infty]$ telle que Y est $\mathcal{G}$–mesurable et $E(X\mathbf{1}_A) = E(Y\mathbf{1}_A)$ pour tout $A \in \mathcal{G}$. Donner un exemple où X est fini p.s. et Y est infini p.s.

18 Soit $\mathcal{G}$ une sous–tribu de $\mathcal{F}$ et A, $B \in \mathcal{F}$. On a

$$E(E(\mathbf{1}_A|\mathcal{G})\, E(\mathbf{1}_B|\mathcal{G})) = E(\mathbf{1}_A\, E(\mathbf{1}_B|\mathcal{G})).$$

A-t-on : $E(E(\mathbf{1}_A|\mathcal{G})\, E(\mathbf{1}_B|\mathcal{G})) \leqslant E(\mathbf{1}_A\, \mathbf{1}_B)$?

19 Soit X et Y intégrables telles que $E(X|Y) = Y$ p.s. et $E(Y|X) = X$ p.s. Montrer que $X = Y$ p.s. (Difficile.)

20 Soit Z une variable aléatoire $\mathcal{G}$–mesurable et X une variable aléatoire indépendante de $\mathcal{G}$. Alors, si f est borélienne et $f(X, Z)$ intégrable, on a :

$$E(f(X, Z)|\mathcal{G}) = \int f(x, Z)\, P_X(dx).$$

21 Soit $\mathcal{H}$ indépendante de $\sigma(X, \mathcal{G})$. On sait que $E(X|\mathcal{G}) = E(X|\sigma(\mathcal{G}, \mathcal{H}))$. Donner un contrexemple avec $\mathcal{H}$ indépendante de X et de $\mathcal{G}$.

22 Si Z est $\mathcal{G}$–mesurable et $E(X|\mathcal{G}) = Z$, alors $E(X|Z) = Z$.
La réciproque est fausse.

23 Soit $(X_k)_{k\leqslant n}$ des variables aléatoires indépendantes de même loi intégrable et absolument continue $f(x)\,dx$ sur $\mathbb{R}$ et soit M leur maximum. Donner la loi de X_k sachant M.

24 (gaussiennes) Soit X et Y deux variables aléatoires indépendantes de même loi $\mathcal{N}(0,1)$. Les couples suivants sont-ils gaussiens ?

$$(X, X-Y), \qquad (X\,\mathrm{sgn}(Y), Y), \qquad (X\,\mathrm{sgn}(Y), Y\,\mathrm{sgn}(X)).$$

25 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi $\mathcal{E}(a)$. On pose

$$S_n = \max_{1\leqslant k\leqslant n} X_k, \qquad I_n = \min_{1\leqslant k\leqslant n} X_k.$$

1) Calculer $E(X_n)$, $\mathrm{var}(X_n)$, $P(I_n > x, S_n \leqslant y)$. En déduire la loi de S_n et de I_n.
2) Montrer que I_n tend vers 0 p.s. et S_n vers l'infini p.s.
3) Calculer $P(X_k \geqslant x | I_n \geqslant y)$ pour $1 \leqslant k \leqslant n$.
4) Soit $(P(\cdot|I_n = y))_{y\geqslant 0}$ une version régulière de P sachant I_n. Utiliser (3) pour calculer $P(X_k \geqslant x|I_n = y)$, puis la loi $P_{X_k}(\cdot|I_n = y)$.
5) Calculer $P(X_k = y|I_n = y)$ et retrouver ce résultat par un argument de symétrie. Que doit-on montrer sur

$$P(X_k = X_l = y|I_n = y)$$

avec $k \neq l$ pour que cet argument soit valide.
6) Calculer la loi $P_{(X_k,X_l)}(\cdot|I_n = y)$ et vérifier que le raisonnement de (5) est valable.

26 (martingales) Soit $(X_n)_n$ une suite de variables aléatoires intégrables et $\mathcal{F}_n$ la tribu engendrée par les X_k pour $k \leqslant n$. On dit que $(X_n, \mathcal{F}_n)_n$ est une martingale (mg) si et seulement si

$$E(X_{n+1}|\mathcal{F}_n) = X_n \quad \text{p.s.}$$

1) Pour $p \geqslant n \geqslant 0$, on a $E(X_p|\mathcal{F}_n) = X_n$ presque sûrement.
2) Soit $Y_n = X_{n+p}$. Trouver une "filtration" $\mathcal{G}_n$ pour laquelle Y est une mg.
3) On suppose que $\sup_n E(X_n^2)$ est fini. Calculer $E(X_{n+1} - X_n)^2$ et en déduire que X_n converge dans L^2 vers une variable aléatoire X et que $X_n = E(X|\mathcal{F}_n)$.
4) On veut montrer, toujours sous l'hypothèse $\sup E(X_n^2)$ fini, que la suite X_n converge presque sûrement.
a) Notons $A_n = \{\sup_{k\leqslant n} X_k > \lambda)$. Montrer l'inégalité de Doob :

$$\lambda\, P(A_n) \leqslant E(X_n\, \mathbf{1}_{A_n}).$$

b) Notons $B_n = \{\sup_{k\leqslant n} |X_k| > \lambda)$. Montrer que l'on a :

$$\lambda\, P(B_n) \leqslant E(|X_n|\, \mathbf{1}_{A_n}).$$

c) On rappelle la formule $E(Y^2) = 2\int y\, P(|Y| > y)\, dy$. Montrer que l'on a :

$$E(\sup_{k\leqslant n} X_k^2) \leqslant c\, E(X_n^2).$$

d) Appliquer ce qui précède à la mg $X'_k = X_{n+k} - X_n$ et en déduire que X_n converge presque sûrement.

27 (W 15.8 observation bruitée de gaussiennes)
Soit X de loi $N(0,\sigma^2)$ (le signal), $(B_n)_{n\geqslant 1}$ des v.a.i.i.d. indépendantes de X et de lois $N(0,1)$ (le bruit, inconnu), et $(c_n)_{n\geqslant 1}$ une suite de réels (les amplitudes du bruit, connues). La suite des observations est

$$Y_n = X + c_n\, B_n.$$

Quand X est-elle $\mathcal{G}$ mesurable, où $\mathcal{G}$ est la tribu engendrée par $(Y_n)_{n\geqslant 1}$? On cherche donc une CNS pour que X soit reconstructible à partir des Y_n, puisque si X est $\mathcal{G}$ mesurable, il existera une fonction mesurable g telle que $X = g(Y_1, Y_2, \cdots)$.
Réponse : la CNS est que la série des $1/c_n^2$ diverge.

Solutions

12 Soit $Y = E(X|\mathcal{G})$, $A = \{X > 0\}$ et $B = \{E(X|\mathcal{G}) > 0\}$.
Alors, $E(Y : B) = E(X : B)$ par définition de Y, et $E(Y : B) = E(Y) = E(X)$ par définition de B. Donc $E(X : B^c) = 0$, soit $B^c \subset A^c$.
Si $A \subset C \in \mathcal{G}$, $E(Y) = E(X) = E(X : C) = E(Y : C)$ donc $E(Y : C^c) = 0$, donc $C^c \subset B^c$.

17 Sur $]0,1]^2$, mesure uniforme, $X(x,y) = 1/x$ et $\mathcal{G}$ engendrée par y.

18 Sur $[0,1]^2$, mesure uniforme, $A = \{x \leqslant y\}$, $B = \{x \geqslant y\}$ et $\mathcal{G}$ engendrée par x.
Mieux : sur $\{1,2\}$, $\mathcal{G} = \sigma(\emptyset)$, $A = \{1\}$ et $B = \{2\}$.

19 Soit c un réel ; alors $\{Y \leqslant c\}$ est Y–mesurable donc $E(X - Y; Y \leqslant c) = 0$. Donc,

$$0 = E(X - Y; Y \leqslant c < X) + E(X - Y; Y \leqslant c, X \leqslant c).$$

Le premier terme est positif donc le second est négatif :

$$E(X; Y \leqslant c, X \leqslant c) \leqslant E(Y; Y \leqslant c, X \leqslant c).$$

Par symétrie, ces deux termes sont égaux donc $E(X - Y; Y \leqslant c < X)$ est en fait nul. Finalement, $P(Y \leqslant c < X) = 0$ pour tout c rationnel donc $Y \geqslant X$ presque sûrement. Par symétrie, on a fini.

21 $X = \varepsilon_1\varepsilon_2$, $\mathcal{G} = \sigma(\varepsilon_1)$ et $\mathcal{H} = \sigma(\varepsilon_2)$.

24 oui, oui, non (ne charge que deux quadrants sur quatre).

26 a) On découpe selon le premier indice où $X_k \geqslant \lambda$.
c) $c = 4$. d) $E(\sup_{k\geqslant n}(X_k - X_n)^2) \leqslant 4(E(X^2) - E(X_n^2))$.

Chapitre 4

Convergence en loi

On peut (on doit ?) traiter la section 4.2 avant la section 4.1. Dans la partie 4.1, on peut passer sur : l'exercice 4.4, la réciproque dans la proposition 4.5 (le sens utile, et facile, est le sens qui assure que $\phi_X \in C^p$ dès que $X \in L^p$), l'exercice 4.6, la proposition 4.8. Dans la partie 4.2, on peut éviter la convergence faible et les caractérisations en dimension supérieure. Le lemme 4.21 peut être redescendu en dimension 1. Laplace facultatif. Dans la partie 4.4, les démonstrations d'unicité des lois limites sont facultatives (donc l'exercice 4.27). On peut omettre la partie 4.4 (TCL multidimensionnel) et la partie 4.4 est un exercice.

Le problème étudié dans ce chapitre peut se résumer simplement : supposons qu'une suite de réels $(p_n)_n$ est définie par

$$p_n = P(X_n \leqslant x),$$

pour des variables aléatoires X_n fixées. Que peut-on dire de son comportement asymptotique quand n tend vers l'infini ? Considérons par exemple une suite $(\xi_n)_n$ de variables aléatoires indépendantes de même loi centrée et intégrable et posons

$$X_n = (\xi_1 + \cdots + \xi_n)/n.$$

La loi des grands nombres indique que X_n tend presque sûrement vers 0, donc aussi en probabilité, et donc que p_n tend vers 0 si $x < 0$ et vers 1 si $x > 0$. Ainsi, $X_n(P)(]-\infty, x])$ tend vers $\delta_0(]-\infty, x])$ pour tout $x \neq 0$. Pour $x = 0$, on ne connaît pas la limite éventuelle. En fait, il est facile de construire des X_n pour lesquels X_n converge presque sûrement vers 0 et $P(X_n \leqslant 0)$ converge vers une limite quelconque entre 0 et 1 ou bien ne converge pas : on peut même remarquer que $(p_n)_n$ définie par $p_n = P(X_n \leqslant 0)$ peut être n'importe quelle suite à valeurs dans $[0, 1]$ (exercice 4.1). En conclusion, $X_n(P)(]-\infty, 0])$ ne converge pas forcément vers $\delta_0(]-\infty, 0]) = 1$.

Malgré cet accident, on a bien l'impression que $X_n(P)$ ressemble de plus en plus à δ_0. Il se trouve que la bonne définition de la convergence en loi consiste précisément à éliminer les valeurs de x pour lesquelles la fonction de répartition de la loi limite (ici, δ_0) est discontinue.

Une autre façon de poser le problème est de chercher à trouver la limite de $E(f(X_n))$ pour le plus de fonctions f possible, tout en gardant une définition raisonnable. On veut donc avoir :

$$E(f(X_n)) \to E(f(X_\infty)),$$

pour une variable aléatoire X_∞ donnée. On sait déjà qu'on ne peut pas demander la convergence pour toutes les fonctions indicatrices. Il se trouve que la bonne définition (les fonctions continues bornées) coïncide avec la définition raisonnable obtenue à partir des fonctions de répartition.

L'outil principal pour étudier la convergence en loi est la transformée de Fourier de la loi d'une variable aléatoire ou, plus généralement, d'une mesure, que l'on appelle sa fonction caractéristique (attention à ne pas confondre avec la fonction *indicatrice* $\mathbf{1}_A$ d'un ensemble A). On commence par définir les fonctions caractéristiques et par étudier leurs propriétés ; puis on relie convergence en loi et fonctions caractéristiques grâce au théorème de Lévy ; enfin, on applique ces résultats à diverses situations concrètes et on démontre en particulier un théorème central limite.

Exercice 4.1 Quelques exemples. Soit X une variable aléatoire et $X_n = X/n$. Alors X_n converge presque sûrement vers 0 et $p_n = P(X_n \leqslant x)$ tend vers 0 si $x < 0$; vers 1 si $x > 0$; vers $P(X_1 \leqslant 0)$ si $x = 0$. On peut choisir X pour obtenir une limite quelconque entre 0 et 1.
Dans la même idée, soit $(Y_n)_n$ une suite de v.a. bornées par 1. Alors $X_n = Y_n/n$ tend p.s. vers 0 et $p_n = P(X_n \leqslant 0) = P(Y_n \leqslant 0)$ est quelconque : par exemple, $Y_n(P) = p_n\delta_{-1} + (1 - p_n)\delta_1)$. ■

4.1 Fonctions caractéristiques

Définition

Définition 4.2 *La fonction caractéristique $\widehat{m}$ d'une probabilité m sur $\mathbb{R}^d$ est la transformée de Fourier de m, soit*

$$\widehat{m}(t) = \int e^{i(t\cdot x)}\, dm(x).$$

Si $m = X(P)$ est la loi d'un vecteur aléatoire X, on appelle $\widehat{m}$ la fonction caractéristique de X et on la note ϕ_X donc : $\phi_X(t) = E\left[e^{i(t\cdot X)}\right]$.

Exemple Si m est absolument continue par rapport à la mesure de Lebesgue et admet pour densité f, alors $\widehat{m} = \widehat{f}$.

Propriétés

Exercice 4.3 Calculer les fonctions caractéristiques des lois usuelles : uniforme, Bernoulli, discrète, Dirac, Cauchy, gaussienne.
Les lois fondamentales :
Loi normale : $\phi(t) = \exp(imt - \sigma^2 t^2/2)$ (on fait le calcul pour la loi centrée réduite). Loi uniforme sur $[0, 1]$: $\phi(t) = (e^{it} - 1)/(it)$. Loi uniforme sur $[-1, 1]$: $\phi(t) = \sin(t)/t$.
Plus anecdotiques :
Double exponentielle : densité $e^{-|x|}/2$, $\phi(t) = 1/(1 + t^2)$. Loi de Cauchy : densité $1/\pi(1 + x^2)$, $\phi(t) = e^{-|t|}$.
Loi triangulaire : densité $(1 - |x|)^+$, $\phi(t) = 2(1 - \cos(t))/t^2$. Loi Anon : densité $\pi^{-1}(1 - \cos(x))/x^2$, $\phi(t) = (1 - |t|)^+$.

Dualité évidente puisque $\widehat{\widehat{f}} = (2\pi)^d f$ dès que f et $\widehat{f} \in L^1$. ■

Quelques propriétés élémentaires : $\phi(0) = 1$; $|\phi(t)| \leqslant 1$; ϕ est continue ; de plus,

$$\phi_{(-X)}(t) = \phi(-t) = \overline{\phi(t)}, \qquad \phi_{(aX+b)}(t) = e^{ibt}\phi(at).$$

Exercice 4.4 ϕ est uniformément continue.

Démonstration (Facultatif) $|\widehat{m}(t) - \widehat{m}(s)| \leqslant E|1 - e^{i(t-s)X}|$ donc la continuité de la fonction $t \mapsto E|1 - e^{itX}|$ en 0 implique l'u.c. Ensuite, e^{itX} tend p.s. vers 1 donc, par convergence dominée de Lebesgue, la convergence est aussi L^1.

Remarque : on aurait pu utiliser $|e^{ia} - e^{ib}| \leqslant |a - b| \wedge 2$. ■

On sait relativement bien caractériser la dérivabilité d'une fonction caractéristique.

Proposition 4.5 *Soit m une probabilité.*
Si $m(\|x\|^n)$ est fini, $\widehat{m}$ est de classe C^n et pour tout multi-indice α de longueur $|\alpha| \leqslant n$,

$$\partial_\alpha \widehat{m}(t) = i^{|\alpha|} \int x^\alpha \, e^{i(t \cdot x)} \, dm(x).$$

En particulier, $\int x^\alpha \, dm(x) = i^{-|\alpha|} \, \partial_\alpha \widehat{m}(0)$.
Réciproquement, si $\widehat{m}$ est de classe C^{2n}, alors $m(\|x\|^{2n})$ est fini.

Pour une variable aléatoire X réelle, on obtient : si $X \in L^n$, $\phi \in C^n$ et $\phi^{(k)}(t) = E(i^k \, X^k \, e^{i(t \cdot X)})$ pour $k \leqslant n$. Réciproquement, si $\phi \in C^{2n}$, alors $X \in L^{2n}$.

Démonstration (Facultatif) La partie directe est évidente à partir du théorème de dérivation sous le signe somme. Rappelons qu'une condition suffisante pour pouvoir dériver est que la fonction dérivée soit uniformément dominée par une fonction intégrable. Ici, on utilise simplement

$$|i^{|\alpha|} \, x^\alpha \, e^{i(t \cdot x)}| \leqslant |x|^\alpha = \prod_k |x_k|^{\alpha_k} \leqslant \|x\|^n \wedge 1.$$

Pour la réciproque, on commence à se ramener au cas $d = n = 1$.
La norme $\|x\|^{2n}$ est contrôlée par la somme des x_k^{2n} donc il suffit de montrer que chaque $m(x_k^{2n})$ est finie. Or, cette intégrale est simplement $m_k(f)$ pour la fonction f définie sur $\mathbb{R}$ par $f(\xi) = \xi^{2n}$ et pour la mesure m_k projection de m sur la k-ième coordonnée. Si $\widehat{m}$ est de classe C^{2n}, $\widehat{m_k}$ aussi donc le résultat en dimension $d = 1$ implique le résultat en toute dimension.
Etablissons une récurrence sur $n \geqslant 1$. Supposons le résultat démontré pour n et considérons une mesure m telle que $\widehat{m}$ soit de classe C^{2n+2}. Posons $m_0(dx) = c \, x^{2n} \, m(dx)$ où c est choisie pour que m_0 soit une probabilité. Alors $\widehat{m_0}$ est C^2 et il suffit de montrer que $m_0(x^2)$ est fini.
Il reste à traiter le cas $d = n = 1$. La fonction $\widehat{m}$ est C^2 donc, quand $t \to 0$,

$$\int 2(1 - \cos(tx)) t^{-2} \, dm(x) = t^{-2} \left(2\widehat{m}(0) - \widehat{m}(t) - \widehat{m}(-t) \right) \longrightarrow -\widehat{m}''(0).$$

Quand $t \to 0$, $2(1 - \cos(tx))/t^2 \to x^2$ tout en étant positif donc le lemme de Fatou implique

$$\begin{aligned} \int x^2 \, dm(x) &= \int \liminf_{t \to 0} \left[2(1 - \cos(tx)) t^{-2} \right] \, dm(x) \\ &\leqslant \liminf_{t \to 0} \int 2(1 - \cos(tx)) t^{-2} \, dm(x) = -\widehat{m}''(0), \end{aligned}$$

c'est-à-dire que $m(x^2)$ est finie. ■

Exercice 4.6 On peut avoir $\widehat{m}$ de classe C^1 et $m(|x|)$ infini. Exemple (exercice 1) : la mesure de densité $f(x) = c \, \mathbf{1}_{|x| \geqslant e} / (|x|^2 \log |x|)$. ■

Le résultat d'injectivité et quelques conséquences.

Proposition 4.7 *Deux probabilités sur $\mathbb{R}^d$ dont les fonctions caractéristiques coïncident sont égales.*

Démonstration (Facultatif) Notons $\mathcal{S}$ l'espace de Schwartz des fonctions de classe C^∞ qui sont à décroissance rapide ainsi que toutes leurs dérivées. On sait que la transformation de Fourier $\mathcal{F}$ est un isomorphisme de $\mathcal{S}$ sur lui-même. Pour $f \in \mathcal{S}$, il existe $g \in \mathcal{S}$ telle que $f = \mathcal{F}(g)$. On connaît même g, qui vaut $g = (2\pi)^{-d}\,\overline{\mathcal{F}}(f)$. Supposons que m et n sont deux probabilités sur $\mathbb{R}^d$ dont les fonctions caractéristiques coïncident. On a alors

$$m(f) = \int dm(x) \int dt\, g(t)\, e^{i(x.t)} = \int \widehat{m}(t)\, g(t)\, dt$$

par application du théorème de Fubini car $|e^{i(x.t)}\, g(t)| \leqslant |g(t)|$ et que $|g| \otimes \mathbf{1}$ est $dt \otimes dm(x)$ intégrable. On a montré que $m(f) = n(f)$ pour toute fonction f de $\mathcal{S}$. A présent, si U est un ouvert, sa fonction indicatrice $\mathbf{1}_U$ peut s'écrire comme la limite croissante d'une suite de fonctions positives de $\mathcal{S}$. On en déduit que $m(U) = n(U)$ par convergence monotone. ■

Proposition 4.8 *Si $\widehat{m}$ est intégrable, m est absolument continue par rapport à la mesure de Lebesgue et sa dérivée de Radon-Nykodym est $(2\pi)^{-d}\,\overline{\mathcal{F}}(\widehat{m})$.*

Des conséquences importantes.

Proposition 4.9 • *Les vecteurs aléatoires X et Y ont même loi si et seulement si, pour tout t, $(t \cdot X)$ et $(t \cdot Y)$ ont même loi.*
• *Les vecteurs aléatoires X_1 et X_2 sont indépendants si et seulement si $\phi_{(X_1,X_2)} = \phi_{X_1} \otimes \phi_{X_2}$.*
En particulier, si X_1 et X_2 sont indépendants, alors $\phi_{X_1+X_2} = \phi_{X_1} \cdot \phi_{X_2}$.

Les deux derniers points sont à lire pour toutes tailles de vecteurs.

Démonstration • Le sens direct est évident. Pour la réciproque, on remarque que, pour t fixé, $\phi_X(t)$ ne fait intervenir que $(t \cdot X)$, donc que la loi de $(t \cdot X)$.

• Dans le sens direct, l'indépendance de X_1 et X_2 permet de factoriser l'intégrale $\phi_X(t_1, t_2)$. Réciproquement, on veut montrer que la loi m de $X = (X_1, X_2)$ est le produit des lois m_1 de X_1 et m_2 de X_2. Pour cela, on montre l'égalité des transformées de Fourier. Par hypothèse, on sait que $\widehat{m} = \phi_X$ vaut

$$\phi_{X_1} \otimes \phi_{X_2} = \widehat{m_1} \otimes \widehat{m_2} = \widehat{m_1 \otimes m_2},$$

la dernière égalité entre transformées de mesures étant toujours vraie d'après Fubini, et ceci permet de conclure. ■

Remarque L'égalité en loi des vecteurs X et Y ne peut résulter de l'égalité en loi de $(t \cdot X)$ et $(t \cdot Y)$ pour t décrivant une base de $\mathbb{R}^n$ car on sait bien que les lois marginales ne suffisent pas à caractériser une loi multidimensionnelle. Un exemple : $X = (\varepsilon, \varepsilon)$ et $Y = (\varepsilon, -\varepsilon)$.

4.2 Convergence en loi

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires. On veut savoir s'il existe une variable aléatoire X telle que

$$P(X_n \in A) \to P(X \in A), \tag{4.1}$$

pour autant de $A \in \mathcal{B}$ que possible. Si c'est le cas et si la loi de X est assez simple, on pourra ainsi estimer $P(X_n \in A)$ pour n grand.

Sous cette forme, la question est mal posée. Soit $X_n = 1/n$ pour $n \geqslant 1$. Alors X_n converge presque sûrement vers $X = 0$ mais $P(X_n \leqslant 0) = 0$ ne converge pas vers $P(X \leqslant 0) = 1$. Le problème vient bien sûr de l'existence d'un atome de la loi de X en 0.

Remarque (4.1) ne fait intervenir X_n et X qu'à travers leurs lois.

Dans toute cette partie, on donnera les preuves pour des mesures sur $\mathbb{R}$ tout en énonçant aussi souvent que possible les résultats sur $\mathbb{R}^d$. Notons que $\mathbb{R}$ peut être remplacé par un espace polonais (métrisable, complet, séparable).

Notations
• C_K, resp. C_0, resp. C_b, est l'ensemble des fonctions continues à support compact, resp. continues tendant vers 0 à l'infini, resp. continues bornées ; bien sûr, $C_K \subset C_0 \subset C_b$.
• $C(F)$ est l'ensemble des points de continuité d'une fonction F entre espaces topologiques.

Définition

Définition 4.10 *Soit $(m_n)_{n\geqslant 1}$ et m des mesures de Radon sur $\mathbb{R}$ ou $\mathbb{R}^d$. On dit que m_n converge étroitement vers m et on note $m_n \xrightarrow{\text{ÉTR}} m$ si $m_n(f)$ tend vers $m(f)$ pour toute fonction $f \in C_b$.*
Soit $(X_n)_{n\geqslant 1}$ et X des variables aléatoires à valeurs dans le même espace. On dit que X_n converge en loi vers X et on note $X_n \xrightarrow{\mathcal{L}} X$ si $X_n(P) \xrightarrow{\text{ÉTR}} X(P)$.
Soit $(F_n)_{n\geqslant 1}$ et F des fonctions de répartition associées aux mesures de probabilités $(m_n)_{n\geqslant 1}$ et m. On dit que F_n converge étroitement vers F et on note $F_n \xrightarrow{\text{ÉTR}} F$ si $m_n \xrightarrow{\text{ÉTR}} m$.

Définition 4.11 *Soit $(m_n)_{n\geqslant 1}$ et m des mesures bornées. On dit que m_n converge faiblement vers m et on note $m_n \xrightarrow{\text{FAI}} m$ si $m_n(f)$ converge vers $m(f)$ pour toute $f \in C_0$.*

Si m_n et m sont des probabilités, les deux convergences sont équivalentes par un argument analogue à celui qui permet de montrer $C_b \subset \mathcal{E}$ à partir de $C_K \subset \mathcal{E}$ dans la preuve du lemme de la section ci-dessous.

La distinction entre les deux types de convergence est tout de même utile, même en probabilité, à cause des phénomènes de fuite de masse à l'infini. Par exemple, si $m_n = \delta_n$ ou si m_n est la loi uniforme sur $[0, n]$, $m_n \xrightarrow{\text{FAI}} 0$ mais $m_n(1) = 1$, donc m_n ne peut pas converger étroitement. La convergence étroite de mesures de probabilités exige donc que la mesure limite soit une probabilité.

Exemples : si $x_n \to x$, alors $\delta_{x_n} \xrightarrow{\text{ÉTR}} \delta_x$. Si $\delta_{x_n} \xrightarrow{\text{ÉTR}} \mu$, alors μ est une Dirac δ_x et $x_n \to x$.

La définition "pratique"

Théorème 4.12 *Soit $(F_n)_{n\geqslant 1}$ et F des fonctions de répartition. Alors $F_n \xrightarrow{\text{ÉTR}} F$ si et seulement si $F_n(x)$ tend vers $F(x)$ pour tout $x \in C(F)$.*

Démonstration Si $F_n \xrightarrow{\text{ÉTR}} F$, on va montrer que

$$F(x-) \leqslant \liminf F_n(x-) \leqslant \limsup F_n(x) \leqslant F(x).$$

A cet effet, posons $h(y) = 1$ si $y \leqslant x$, $h(y) = 1 - (y - x)/a$ si $x < y < x + a$ et $h(y) = 0$ si $y \geqslant x + a$. Alors $m_n(h)$ tend vers $m(h)$ et on voit que $F_n(x) \leqslant m_n(h)$ et $m(h) \leqslant F(x + a)$. On en déduit :

$$\limsup F_n(x) \leqslant F(x + a).$$

A présent, F est continue à droite donc on fait tendre a vers $0+$. Pour l'autre inégalité, on travaille avec $y \mapsto h(y + a)$.

Réciproquement, notons $\mathcal{E}$ l'espace vectoriel des fonctions $f \in C_b$ telles que $m_n(f)$ tend vers $m(f)$ où les m_n et m sont les mesures associées aux F_n et à F. Alors, $\mathcal{E}$ est fermé pour la norme uniforme et contient par hypothèse les $\mathbf{1}_{]x,y]}$ pour $x, y \in C(F)$. Les points de discontinuité de F sont au plus en nombre dénombrable donc $C(F)$ est dense dans $\mathbb{R}$.

Montrons que $\mathcal{E}$ contient C_K. Soit $f \in C_K$. Alors, f est nulle en dehors d'un compact et uniformément continue sur ce compact. Pour tout $\varepsilon > 0$, il existe une fonction en escalier de Darboux g telle que $g \leqslant f \leqslant g+\varepsilon$ et dont la subdivision associée ne contient que des points de $C(F)$. Alors, $m_n(g)$ tend vers $m(g)$ donc :

$$\limsup m_n(f) \leqslant \varepsilon + \limsup m_n(g) = \varepsilon + m(g) \leqslant \varepsilon + m(f).$$

De même pour la limite inférieure, donc $m_n(f)$ converge vers $m(f)$ et $\mathcal{E}$ contient C_K.

Soit $f \in C_b$ avec $0 \leqslant f \leqslant 1$ et K un compact tel que $m(K^c) \leqslant \varepsilon$. Il existe une fonction $g \in C_K$ telle que $\mathbf{1}_K \leqslant g \leqslant 1$. Alors, le produit fg appartient à C_K et $f \geqslant fg$ donc :

$$\liminf m_n(f) \geqslant \liminf m_n(fg) = m(fg) \geqslant m(f\mathbf{1}_K) \geqslant m(f) - \varepsilon.$$

On a démontré $m(f) \leqslant \liminf m_n(f)$ et on applique ce résultat à $(1-f)$, donc $m_n(f)$ converge vers $m(f)$ et $\mathcal{E} = C_b$. ■

Remarque On voit que le raisonnement permettant de remonter de $f \in C_K$ à $f \in C_b$ fonctionne dès que $m_n(\mathbb{R})$ converge vers $m(\mathbb{R})$.

Proposition 4.13 *Soient m_n et m des mesures finies sur $\mathbb{R}^d$. Alors, $m_n \xrightarrow{\text{ÉTR}} m$ si et seulement si $m_n \xrightarrow{\text{FAI}} m$ et $m_n(\mathbb{R}^d) \to m(\mathbb{R}^d)$.*

Proposition 4.14 *Soit m_n et m des mesures bornées sur $\mathbb{R}^d$ telles que $m_n \xrightarrow{\text{FAI}} m$ (en particulier si m_n et m sont des probabilités avec $m_n \xrightarrow{\text{ÉTR}} m$). Alors, pour tout compact K, pour tout ouvert U et pour tout borélien borné B dont la frontière est m-négligeable, on a*

$$\begin{aligned} \limsup m_n(K) &\leqslant m(K), \\ \liminf m_n(U) &\geqslant m(U), \\ \lim m_n(B) &= m(B). \end{aligned}$$

Démonstration Soit K un compact et $f \in C_0$ telle que $f \geqslant \mathbf{1}_K$. Alors la limite supérieure des $m_n(K)$ est majorée par celle des $m_n(f)$ c'est-à-dire par $m(f)$. Le résultat s'en déduit en passant à l'infimum sur les fonctions f. Pour un ouvert U, la technique est la même en approchant $\mathbf{1}_U$ par en dessous et pour un borélien B, on applique ce qui précède à l'adhérence et à l'intérieur de B. ■

Remarque Si m_n et m sont des probabilités, chacune de ces trois propriétés (K compact, U ouvert, B borélien (borné ou non) de frontière négligeable) est donc équivalente à la convergence étroite de m_n vers m.

Relions la convergence en loi aux autres types de convergence.

Proposition 4.15 *Si $X_n \xrightarrow{P} X$, alors $X_n \xrightarrow{\mathcal{L}} X$. Si $X_n \xrightarrow{\mathcal{L}} c$ où c est une constante, alors $X_n \xrightarrow{P} c$.*

Démonstration Soit $X_n \xrightarrow{P} X$ et $f \in C_b$. Alors il est facile de montrer que $f(X_n) \xrightarrow{P} f(X)$ (c'est un exercice du chapitre 1, on commence par supposer que f est uniformément continue). Ensuite, f est bornée donc, par convergence dominée, $(f(X_n))_{n\geqslant 1}$ converge dans L^1 également. Pour la réciproque, soit $X_n \xrightarrow{\mathcal{L}} c$. La fonction $f_\varepsilon(x) = 1 \wedge (d(x,c)/\varepsilon)$ est continue bornée donc :

$$P(\|X_n - c\| \geqslant \varepsilon) \leqslant E(f_\varepsilon(X_n)) \to E(f_\varepsilon(c)) = 0.$$

Exercice : c'est la seule réciproque possible. Montrer que si X n'est pas presque sûrement constante, il existe une suite de variables aléatoires $(X_n)_n$ qui converge en loi vers X sans converger en probabilité. ■

Exercice 4.16 Soit X_n et X des variables aléatoires dont les lois sont portées par le même ensemble $D \subset \mathbb{R}$. Si D est fini, une CNS pour avoir $X_n \xrightarrow{\mathcal{L}} X$ est

$$\forall t \in D, \quad P(X_n = t) \to P(X = t).$$

Montrer que seule la réciproque subsiste dans le cas D dénombrable. ($X_n = 1/n$ sur D formé des $1/n$ pour $n \geqslant 1$ et de 0.) ■

[V.a. entières : page 45 de Bertoin.]

Helly–Bray

On étudie la convergence des mesures de probabilité ; il faut donc faire appel à une propriété de compacité pour pouvoir disposer de valeurs d'adhérence. Malheureusement $C_b(\mathbb{R})$ n'est pas séparable. Par contre, $\overline{\mathbb{R}}$ est compact donc $C(\overline{\mathbb{R}})$ est séparable. C'est le théorème d'Helly–Bray : l'ensemble des sous–probabilités de $\overline{\mathbb{R}}$ est un compact métrisable (donc séquentiellement compact) pour la topologie de la convergence étroite car c'est la boule unité du dual de $C(\overline{\mathbb{R}})$. On va démontrer à la main ce résultat en utilisant les fonctions de répartition et la séparabilité de $\mathbb{R}$.

Lemme 4.17 *Soit $(F_n)_{n\geqslant 1}$ une suite de fonctions de répartition. Il existe une fonction G càdlàg, croissante avec $0 \leqslant G \leqslant 1$ et une sous suite $(n_k)_k$ telles que $F_{n_k}(x)$ tend vers $G(x)$ pour tout $x \in C(G)$.*

Démonstration Soit $Q = \{q_n\,;\, n \geqslant 1\}$ un sous–ensemble dense de $\mathbb{R}$. La suite $(F_n(q_1))_n$ est bornée donc une sous–suite $(F_{\phi_1(n)}(q_1))_n$ converge vers $H(q_1)$. La suite $(F_{\phi_1(n)}(q_2))_n$ est bornée donc une sous–suite $(F_{\phi_2(n)}(q_2))_n$, où $(\phi_2(n))_n$ est extraite de $(\phi_1(n))_n$, converge vers $H(q_2)$. Et ainsi de suite. Soit $n_k = \phi_k(k)$. Alors $(F_{n_k}(q))_k$ converge vers $H(q)$ pour tout $q \in Q$. On voit que $0 \leqslant H \leqslant 1$ et que H est croissante sur Q. Il reste à construire une "version" de H qui soit continue à droite. Pour $x \in \mathbb{R}$, on pose $G(x) = \inf\{H(q)\,;\, q \in Q,\, q > x\}$. Enfin, si $x \in C(G)$, on peut vérifier que $F_{n_k}(x) \to G(x)$ en encadrant x par des éléments de Q, d'où le résultat. ■

On voit que G n'est pas forcément une fonction de répartition.

Tension

Dans les situations concrètes, on connaît les lois $X_n(P)$ mais pas l'éventuelle loi limite. En particulier, la loi de X_n pourrait converger faiblement vers une mesure de masse différente de 1 (exemple : $X_n = n$). Le bon critère pour vérifier que la mesure limite est bien une probabilité est le critère de tension ci-dessous, qui nous permet de compléter les préliminaires du théorème 4.20 de Lévy. On remarquera l'analogie entre la tension et l'uniforme intégrabilité.

Définition 4.18 *Une famille $\mathcal{M}$ de mesures est dite tendue si pour tout $\varepsilon > 0$, il existe un compact $K \subset \Omega$ tel que $m(\Omega \backslash K) \leqslant \varepsilon$ pour toute mesure $m \in \mathcal{M}$.*

En particulier, une famille $\mathcal{R}$ de fonctions de répartition est tendue si pour tout $\varepsilon > 0$, il existe $a > 0$ tel que $F(a) - F(-a) \geqslant 1 - \varepsilon$ pour toute $F \in \mathcal{R}$. La tension est donc la propriété qui empêche de "pousser (une partie non nulle de) la masse vers l'infini". Le lemme suivant est laissé en exercice.

Lemme 4.19 *Soit $(F_n)_n$ une suite de fonctions de répartition. Si $F_n \stackrel{\text{ÉTR}}{\longrightarrow} F$ pour une fonction de répartition F, alors $(F_n)_n$ est tendue. Réciproquement, si $(F_n)_n$ est tendue, il existe une sous–suite $(n_k)_k$ et une fonction de répartition F telles que $F_{n_k} \stackrel{\text{ÉTR}}{\longrightarrow} F$.*

Le résultat de Lévy

Théorème 4.20 (Paul Lévy) *Soit $(m_n)_{n \geqslant 1}$ une suite de probabilités sur $\mathbb{R}^d$. Si $m_n \stackrel{\text{ÉTR}}{\longrightarrow} m$, alors m est une probabilité et, pour tout $t \in \mathbb{R}^d$, $\widehat{m_n}(t)$ converge vers $\widehat{m}(t)$.*
Réciproquement, si $\widehat{m_n}$ converge simplement vers une fonction ϕ continue à l'origine, alors m_n converge étroitement vers une mesure de probabilité m et m vérifie $\phi = \widehat{m}$.

Contrexemples : $m_n = \delta_n$ ou m_n uniforme sur $[-n, n]$.

Démonstration La partie directe est évidente. Pour la réciproque, supposons que la suite $(m_n)_n$ est tendue. Alors, il existe une sous–suite m_{n_k} qui converge étroitement vers une probabilité m (Helly–Bray). On en déduit que $\widehat{m_{n_k}}$ converge simplement vers $\widehat{m}$ donc $\phi = \widehat{m}$.
A présent, $(m_n)_n$ converge étroitement vers m. En effet, dans le cas contraire, une deuxième sous–suite convergerait vers $m' \neq m$, ce qui est absurde car on aurait alors $\phi = \widehat{m} = \widehat{m'}$ donc $m' = m$ par le théorème d'injectivité.

Il reste à démontrer la tension. On sait que $\widehat{m_n}(at) \to \phi(at)$ pour tout t quand $n \to \infty$ et que $\phi(at)$ tend vers $\phi(0) = 1$ pour tout t quand $a \to 0$, avec $|\widehat{m_n}| \leqslant 1$ et $|\phi| \leqslant 1$. Par convergence dominée, on en déduit que
$$\int_{[-1,1]^d} (1 - \widehat{m_n}(a\,\cdot)) \longrightarrow K(a) = \int_{[-1,1]^d} (1 - \phi(a\,\cdot))$$
quand $n \to \infty$ et que $K(a) \to 0$ quand $a \to 0$. Fixons $\varepsilon > 0$ et soit a tel que $K(a) \leqslant \varepsilon$. En admettant pour l'instant le lemme 4.21, il vient, pour $n \geqslant n_\varepsilon$,
$$m_n(\mathbb{R}^d \backslash [-2/a, 2/a]^d) \leqslant 2c_d\varepsilon.$$
Pour chaque $n \leqslant n_\varepsilon$, on choisit a_n tel que $m_n(\mathbb{R}^d \backslash [-2/a_n, 2/a_n]^d)$ est également majoré par $2c_d\varepsilon$. Alors, si b est le maximum des $2a_n$ pour $n \leqslant n_\varepsilon$ et de $2a$, on obtient $m_n(\mathbb{R}^d \backslash [-1/b, 1/b]^d) \leqslant 2c_d\varepsilon$ pour tout $n \geqslant 1$, soit la tension voulue. ■

Lemme 4.21 *Soit m une probabilité sur $\mathbb{R}^d$ et $a > 0$. Alors*
$$m(\mathbb{R}^d \backslash [-2/a, 2/a]^d) \leqslant c_d \int_{[-1,1]^d} (1 - \widehat{m}(a\,t))\, dt.$$

Démonstration Par Fubini, on a :
$$\begin{aligned} \int_{[-1,1]^d} (1 - \widehat{m}(a\,t))\, dt &= \int dm(x) \left(2^d - \int_{[-1,1]^d} e^{ia\,(tx)}\, dt \right) \\ &= 2^d \int dm(x) \left(1 - \prod_{k=1}^{d} \sin a x_k / a x_k \right). \end{aligned}$$

Remarquons que $\sin\xi/\xi \leqslant 1$ pour tout ξ et $\sin\xi/\xi \leqslant 1/2$ pour $|\xi| \geqslant 2$. Si $x \notin [-2/a, 2/a]^d$, l'un des x_k vérifie $\sin ax_k/ax_k \leqslant 1/2$ donc la parenthèse dans la dernière intégrale est minorée par 1/2. On en déduit le résultat avec $c_d = 2^{1-d}$. ■

Si on enlève la condition de continuité à l'origine, le résultat de Lévy est évidemment faux : si m_n est la loi de probabilité uniforme sur $[-n, +n]$, la suite $\widehat{m_n}$ converge simplement vers $\mathbf{1}_{\{0\}}$.

Exemple L'application pratique la plus fréquente du théorème 4.20 consiste à affirmer qu'une suite X_n converge en loi vers une variable aléatoire X si et seulement si les fonctions ϕ_{X_n} convergent ponctuellement vers ϕ_X (ce qui est en général plus facile à vérifier).

Exercice 4.22 [Transformée de Laplace] Si m est une mesure sur $\mathbb{R}_+$, sa transformée de Laplace Lm est définie par

$$Lm(t) = \int_0^{+\infty} e^{-(tx)}\, dm(x)$$

pour $t \geqslant 0$. On remarquera que $\{0\}$ contribue pour $m(\{0\})$ dans $Lm(t)$.

• Montrer qu'il existe une constante c telle que, pour toute probabilité m sur $\mathbb{R}_+$ et tout $a > 0$,

$$m([1/a, +\infty[) \leqslant c \int_0^1 (1 - Lm(at))\, dt.$$

• Soit X et Y deux variables aléatoires à valeurs dans $[0, 1]$ telles que $E(X^n) = E(Y^n)$ pour tout $n \geqslant 1$ entier. Montrer successivement que $E(p(X)) = E(p(Y))$ pour tout polynôme, que $E(f(X)) = E(f(Y))$ pour toute fonction continue (Stone–Weierstrass !) puis que $P(X \leqslant x) = P(Y \leqslant x)$ pour tout $x \in [0, 1]$.
• Montrer que si $Lm_1 = Lm_2$ pour deux probabilités m_1 et m_2, alors $m_1 = m_2$ (théorème d'injectivité).
• En déduire que si $(X_n)_n$ est une suite de variables aléatoires positives telles que $L_n(t) = E(e^{-tX_n})$ converge pour tout $t \geqslant 0$ vers une fonction L_0 continue à l'origine, alors $(X_n)_n$ converge en loi vers une probabilité dont L_0 est la transformée de Laplace. ■

4.3 Convergence en loi : résumé actualisé

(A) Deux définitions équivalentes de la convergence en loi :

(i) une suite $(m_n)_n$ de mesures de probabilités converge étroitement vers une mesure m si et seulement si $m_n(f) \to m(f)$ pour toute fonction f continue bornée. Alors, m est une probabilité.

(ii) une suite $(F_n)_n$ de fonctions de répartition converge étroitement vers une fonction de répartition F si et seulement si $F_n(x) \to F(x)$ pour tout x point de continuité de F.

Si F_n est la fonction de répartition de m_n, $F_n \xrightarrow{\text{ÉTR}} F$ si et seulement si $m_n \xrightarrow{\text{ÉTR}} m$, où F est la fonction de répartition de m. Dans ce cas, si $X_n(\mathbb{P}) = m_n$ et $X(\mathbb{P}) = m$, on note $X_n \xrightarrow{\mathcal{L}} X$.

(B) Un peu de compacité :

(iii) Helly : pour toute suite $(F_n)_n$ de fonctions de répartition, une sous-suite converge vers la fonction de répartition G d'une sous-distribution (i.e. G est comme une fonction de répartition sauf que $G(+\infty) - G(-\infty) \leqslant 1$).

(iv) Prohorov : si $F_n \xrightarrow{\text{ÉTR}} F$, alors $\{F_n\}_n$ est tendue ; si $\{F_n\}_n$ est tendue, une sous-suite converge vers une “vraie” fonction de répartition.

(C) Lévy : si $\phi_{X_n}(t)$ converge pour tout t vers une limite $g(t)$ et si la fonction g est continue en 0, alors il existe X telle que $g = \phi_X$ et $X_n \xrightarrow{\mathcal{L}} X$.

On savait déjà que si $X_n \xrightarrow{\mathcal{L}} X$, alors $\phi_{X_n}(t)$ converge vers $\phi_X(t)$ pour tout t (puisque les exponentielles sont continues bornées). Lévy donne une réciproque :

Corollaire : $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\phi_{X_n}(t) \to \phi_X(t)$ pour tout t.

La preuve du résultat de Lévy :

1. Si $\{F_n\}_n$ est tendue, Prohorov assure qu'une sous-suite converge vers une fonction de répartition F_X. Donc une sous-suite de $(X_n)_n$ converge en loi vers X. Par la caractérisation (i), une sous-suite de $(\phi_n)_n$ converge vers ϕ_X. Donc $g = \phi_X$.

Si $(X_n)_n$ ne converge pas en loi vers X, $|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \geqslant \epsilon$ pour f continue bornée fixée et pour tout X_n d'une sous-suite. On applique Prohorov à cette sous-suite. Toute valeur d'adhérence Y doit vérifier $\phi_Y = g$ et $|\mathbb{E}(f(Y)) - \mathbb{E}(f(X))| \geqslant \epsilon$. C'est absurde.

2. On montre que $\{F_n\}_n$ est tendue grâce au

Lemme 4.23

$$\mathbb{P}(|Z| \geqslant 2/s) \leqslant s^{-1} \int_{-s}^{+s} (1 - \phi_Z(t))\,\mathrm{d}t.$$

Pour appliquer ce lemme, fixons $\epsilon > 0$. Comme $g(0) = \lim \phi_n(0) = 1$, la continuité de g en 0 donne $|1 - g(t)| \leqslant \epsilon$ pour tout $|t| \leqslant s$. Par convergence dominée de ϕ_n vers g sur $[-s, +s]$, il vient

$$s^{-1} \int_{-s}^{+s} |1 - \phi_n(t)|\,\mathrm{d}t \leqslant \epsilon + s^{-1} \int_{-s}^{+s} |1 - g(t)|\,\mathrm{d}t \leqslant 3\epsilon,$$

pour $n \geqslant n_\epsilon$. Donc $\mathbb{P}(|X_n| \geqslant 2/s) \leqslant 3\epsilon$ pour $n \geqslant n_\epsilon$. Pour chaque $n \leqslant n_\epsilon$, il existe t_n tel que $\mathbb{P}(|X_n| \geqslant t_n) \leqslant 3\epsilon$. Donc $\mathbb{P}(|X_n| \geqslant t(\epsilon)) \leqslant 3\epsilon$ pour tout $n \geqslant 1$, si $t(\epsilon) = \max\{2/s\} \cup \{t_n \,;\, n \leqslant n_\epsilon\}$, ce qui prouve la tension.

3. La preuve du lemme : par Fubini,

$$s^{-1} \int_{-s}^{+s} (1 - \phi_Z(t))\,\mathrm{d}t = \mathbb{E}\left[s^{-1} \int_{-s}^{+s} (1 - \mathrm{e}^{itZ})\,\mathrm{d}t\right] = 2\,\mathbb{E}[1 - \sin(sZ)/(sZ)].$$

On minore $1 - \sin(sZ)/(sZ)$ par 0 sur $\{|Z| < 2/s\}$ et par $1 - 1/|sZ| \geqslant \frac{1}{2}$ sur $\{|Z| \geqslant 2/s\}$. Il reste $2\,\mathbb{E}[\frac{1}{2} \,;\, |Z| \geqslant 2/s] = \mathbb{P}[|Z| \geqslant 2/s]$, soit le lemme.

4.4 Théorèmes limites

LGN

Considérons pour l'instant des variables aléatoires réelles. On peut retrouver sans effort la loi faible des grands nombres. Rappelons que, si $(X_n)_{n\geqslant 1}$ est une suite de v.a., $S_n = X_1 + \cdots + X_n$.

Proposition *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi intégrable. Alors, S_n/n converge en probabilité vers $c = E(X_1)$.*

Démonstration Si $m_1 = X_1(P)$ et $m_n = (S_n/n)(P)$, alors $\widehat{m_n}(t) = \widehat{m_1}(t/n)^n$. Or, m_1 est intégrable donc $\widehat{m_1}(t/n) = 1 + it\,c/n + o(1/n)$. Ainsi, $\widehat{m_n}(t)$ tend vers $\exp(it\,c) = \widehat{\delta_c}(t)$ donc la proposition découle du théorème de Lévy car dans ce cas, la convergence en loi implique la convergence en probabilité. ■

La loi de Dirac obtenue comme limite est une des seules possibles. En effet :

Théorème *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi et soit m_n la loi de S_n/n. Si $m_n \stackrel{\text{ÉTR}}{\longrightarrow} m$ où m est une probabilité, on doit avoir*

$$\widehat{m}(t) = e^{ibt}\, e^{-a|t|}$$

pour certains $b \in \mathbb{R}$, $a \geqslant 0$. Ainsi, si $a = 0$, m est une masse de Dirac en b et si $a > 0$, m est une loi de Cauchy $\mathcal{C}(b, a)$.

Démonstration On remarque que $\widehat{m}(t) = \lim \widehat{m_n}(t) = \lim \widehat{m_1}(t/n)^n$. On en déduit, en remplaçant n par un multiple de n, que $\widehat{m}$ vérifie l'équation fonctionnelle :

$$\forall n \geqslant 1, \qquad \widehat{m}(t) = \widehat{m}(t/n)^n.$$

Ensuite, l'approximation habituelle des réels par des rationnels associée à la continuité de $\widehat{m}$ fournit $\widehat{m}(t) = \exp(-at + ibt)$ sur $t \geqslant 0$ avec $a \geqslant 0$ car $|\widehat{m}| \leqslant 1$, puis la relation $\widehat{m}(-t) = \overline{\widehat{m}(t)}$ fournit la formule sur $t \leqslant 0$. ∎

Exemple 4.24 Regarder le cas où les X_n suivent une loi de Cauchy. ∎

Théorème central limite

Théorème 4.25 *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et de même loi de carré intégrable avec $E(X_1) = \mu$ et $\mathrm{var}(X_1) = \sigma^2$. Alors $T_n = (S_n - n \cdot \mu)/\sqrt{n \cdot \sigma^2}$ converge en loi vers la loi normale centrée réduite $\mathcal{N}(0, 1)$.*

Les variables aléatoires T_n sont réduites c'est-à-dire que $E(T_n) = 0$ et $\mathrm{var}(T_n) = 1$. La loi du logarithme itéré précise les fluctuations de T_n comme suit : l'ensemble des valeurs d'adhérence de la suite $T_n/\sqrt{2 \log \log n}$ est presque sûrement égal à l'intervalle $[-1, +1]$. En particulier, T_n n'est pas bornée mais le théorème 4.25 assure que la loi de T_n converge.

Remarque : la LLI regarde la suite $(T_n(\omega))_n$ pour un ω fixé. Au contraire, le TCL assure la convergence de (la répartition de) l'ensemble des valeurs de $T_n(\omega)$, $\omega \in \Omega$, pour une valeur de n fixée.

Démonstration On note $\widehat{m_n}$ la fonction caractéristique de T_n. Alors, la fonction $\widehat{m_n}$ est donnée par :

$$\widehat{m_n}(t) = \widehat{m_1}(t/\sqrt{n})^n.$$

De plus, $\widehat{m_1}$ est de classe C^2 car la loi de X_1 est de carré intégrable et $\widehat{m_1}'(0) = 0$ et $\widehat{m_1}''(0) = 1$ donc

$$\widehat{m_1}(t/\sqrt{n}) = 1 - t^2/2n + o(1/n),$$

et $\widehat{m_n}(t)$ converge vers $\exp(-t^2/2)$ qui est la fonction caractéristique de $\mathcal{N}(0, 1)$. Le théorème de Lévy permet de conclure. ∎

La loi normale est la seule loi limite possible. En effet :

Théorème 4.26 *Soit X_n des variables aléatoires indépendantes et de même loi. Si $S_n/\sqrt{n}$ converge en loi vers une loi m, alors $\widehat{m}(t) = e^{-\sigma^2 t^2/2}$ avec $\sigma^2 \geqslant 0$.*
Si $\sigma^2 = 0$, $m = \delta_0$, si $\sigma^2 > 0$, $m = \mathcal{N}(0, \sigma^2)$.

Démonstration Comme dans le théorème d'unicité de la LGN, on trouve que $\widehat{m}$ vérifie $\widehat{m}(t) = \widehat{m}(t/\sqrt{n})^n$ pour tout $n \geqslant 1$. À nouveau par approximation des réels par des rationnels, puis continuité, puis conjugaison de t à $-t$, on obtient

$$\widehat{m}(t) = \exp(-\sigma^2\, t^2/2 + ia \operatorname{sgn}(t)\, t^2),$$

pour $\sigma^2 \geqslant 0$ et a réel. Il reste à montrer que $a = 0$. On se reportera à l'exercice 4.27, tout en remarquant que si l'on ajoute l'hypothèse "la loi limite m est de carré intégrable", alors $\widehat{m}$ doit être de classe C^2 en 0, ce qui implique directement $a = 0$. ■

Exercice 4.27 Montrer que $\phi(t) = \exp(-\sigma^2\, t^2/2 + ia \operatorname{sgn}(t)\, t^2)$ n'est pas une fonction caractéristique si $a \neq 0$. Pour cela on remarque que les formes quadratiques associées sont positives (écrire une somme sur (n,m) de $z_n \overline{z_m} \phi(t_n - t_m)$ comme l'espérance d'une v.a. positive), donc

$$k(t) = \begin{vmatrix} \phi(0) & \phi(-t) & \phi(-2t) \\ \phi(t) & \phi(0) & \phi(-t) \\ \phi(2t) & \phi(t) & \phi(0) \end{vmatrix} \geqslant 0$$

pour tout t. On calcule facilement

$$\begin{aligned} k(t) &= 1 - 2\,|\phi(t)|^2 - |\phi(2t)|^2 + 2\operatorname{Re}(\phi(t)^2\overline{\phi(2t)}) \\ &= 1 - 2e^{-\sigma^2 t^2} - e^{-4\sigma^2 t^2} + 2e^{-3\sigma^2 t^2}\cos(2a\, t^2). \end{aligned}$$

Un équivalent en 0 est $k(t) = -4a^2t^4 + o(t^4)$, ce qui est absurde si $a \neq 0$. ■

Utilisation statistique du théorème central limite

Enonçons tout d'abord un corollaire historiquement et statistiquement important. Nous noterons

$$\gamma(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-y^2/2}\, dy$$

la fonction de répartition d'une loi $\mathcal{N}(0,1)$ gaussienne centrée réduite.

Corollaire (de Moivre–Laplace) *Soit $(A_k)_{k\geqslant 1}$ une suite d'événements indépendants de probabilité $0 < p < 1$ et f_n la fréquence de réalisation des A_k au temps n :*

$$f_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}(A_k).$$

Alors, pour tout couple de réels $a \leqslant b$,

$$P(a \leqslant \frac{f_n - p}{\sqrt{p\,(1-p)/n}} \leqslant b) \longrightarrow \gamma(b) - \gamma(a) \quad \textit{quand} \quad n \to +\infty.$$

On peut déduire du théorème de de Moivre–Laplace un intervalle de confiance pour la valeur de p connaissant f_n : on sait que f_n tend presque sûrement vers p mais, un seuil $\varepsilon > 0$ étant fixé, on cherche δ petit avec

$$P(p \in [f_n - \delta, f_n + \delta]) \geqslant 1 - \varepsilon.$$

Soit β tel que $\gamma(\beta) = 1 - (\varepsilon/2)$. Si on suppose n assez grand pour assimiler les deux membres de l'énoncé du théorème de de Moivre–Laplace, on en déduit

$$P(\frac{f_n - p}{\sqrt{p\,(1-p)}} \in [-\beta/\sqrt{n}, \beta/\sqrt{n}]) \approx 1 - \varepsilon,$$

d'où un intervalle de confiance non symétrique pour p. On peut encore simplifier les calculs en remarquant que $\sqrt{f_n\,(1-f_n)}$ converge presque sûrement vers $\sqrt{p\,(1-p)}$. Ainsi,

$$\frac{f_n - p}{\sqrt{f_n\,(1-f_n)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

d'où l'intervalle symétrique $[f_n - \beta\,\sqrt{f_n\,(1-f_n)/n}\,;\, f_n + \beta\,\sqrt{f_n\,(1-f_n)/n}]$.

Exercice 4.28 Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} a$, alors $Y_n X_n \xrightarrow{\mathcal{L}} aX$. ■

Exemple $n = 1024$, $f_n = 0,51$, $\varepsilon = 0,05$ donne $p \in [0,49\,;\,0,53]$.

Théorème central limite multi–dimensionnel

Rappelons qu'un vecteur X de $\mathbb{R}^d$ est dit gaussien si pour tout $t \in \mathbb{R}^d$, $(t \cdot X)$ est une variable aléatoire réelle gaussienne. On note $\mathcal{N}(M,C)$ la loi gaussienne de moyenne M et de matrice de covariance C. On dispose d'un théorème central limite multi–dimensionnel.

Théorème *Soit* $(X_n)_{n\geqslant 1}$ *une suite de vecteurs aléatoires indépendants et de même loi. On suppose que* $E\|X_1\|^2$ *est fini, on note* $M = E(X_1)$ *et* C *la matrice de covariance de* X_1*. Alors,* $(S_n - n\cdot M)/\sqrt{n}$ *converge en loi vers* $\mathcal{N}(O,C)$*.*

Démonstration Soit $x \in \mathbb{R}^d$ et $Y_n = (x \cdot X_n)$. Les Y_n sont des variables aléatoires indépendantes de même loi de carré intégrable et

$$E(Y_n) = (M \cdot x), \qquad \mathrm{var}(Y_n) = x^*\,C\,x,$$

donc grâce au TCL uni–dimensionnel, on trouve que

$$\left(x \cdot \frac{S_n - nM}{\sqrt{n}}\right) = \sum_{k=1}^{n} (Y_k - E(Y_k))/\sqrt{n} \longrightarrow \mathcal{N}(0, x^*\,C\,x).$$

On en déduit que la fonction caractéristique de $(S_n - nM)/\sqrt{n}$ évaluée en x converge vers

$$\exp(-x^*\,C\,x/2),$$

qui est la fonction caractéristique de la loi $\mathcal{N}(O,C)$ évaluée en x, donc le théorème de Lévy donne le résultat.
■

Répartition des étoiles dans l'univers

D'après Chandrasekhar (1950).

Univers $B(O,r)$, n étoiles, donc $n/\mathrm{vol}(B(O,r)) = \rho$ densité stellaire observée. Avec pour seule donnée ρ, on veut donner une valeur de D, distance entre notre soleil et l'étoile la plus proche.

Chaque étoile est en $X_n(i)$, $i \leqslant n$, avec les hypothèses suivantes : indépendance ; lois uniformes. Soit $s > 0$ un rayon et posons

$$N_n(s) = \sum_{i=1}^{n} \mathbf{1}(X_n(i) \in B(O,s)).$$

Alors $N_n(s)$ est le nombre d'étoiles présentes dans la boule de rayon s. En particulier, $P(N_n(s) = 0)$ est la probabilité pour que l'étoile la plus proche de O soit à une distance D_n supérieure à s. On postule que la situation réelle est bien représentée par le comportement du modèle quand r tend vers l'infini. Ainsi :

$$P(D_n \geqslant s) = (1 - P(X_n(1) \in B(O, s)))^n = (1 - v(s)/v(r))^n,$$

en notant $v(s)$ le volume de la boule de rayon s. Or, par hypothèse $v(r) = n/\rho$ donc $P(D_n \geqslant s)$ tend vers $\exp(-\rho v(s))$. On peut donc approcher $P(D \geqslant s)$ par $\exp(-\rho v(s))$ et en déduire la loi de D.

On peut obtenir mieux, à savoir la répartition limite des étoiles. Pour cela, soit ϕ_n la transformée de Fourier de $N_n(s)$. Par indépendance et équidistribution, on trouve en notant $v(s)$ le volume de la boule de rayon s :

$$\phi_n(t) = E(\exp(it\mathbf{1}[X_n(1) \in B(O, s)]))^n = [1 - (1 - e^{it})v(s)/v(r)]^n.$$

A nouveau, $v(r) = n/\rho$ donc $\phi_n(t)$ tend vers $\exp(-\rho v(s)(1-e^{it}))$, expression dans laquelle on peut reconnaître avec un peu d'habitude la transformée de Fourier de la mesure de Poisson sur les entiers d'intensité $\rho v(s)$ définie par

$$m = \sum_{n \geqslant 0} \delta_n \exp(-\rho v(s))(\rho v(s))^n/n!$$

En appliquant le théorème de Lévy, on a démontré la convergence en loi de la suite $N_n(s)$, $n \geqslant 1$. Remarquons à présent que, les variables $N_n(s)$ ne prenant que des valeurs entières, on a

$$P(N_n(s) \leqslant p) = P(N_n(s) \leqslant p + (1/2)).$$

De plus, la mesure m ne comporte pas d'atome en $p + (1/2)$ donc la probabilité $P(N_n(s) \leqslant p)$ tend vers

$$m(] - \infty, p + 1/2]) = \sum_{n=0}^{p} \exp(-\rho v(s))(\rho v(s))^n/n!$$

Ce modèle suggère donc de représenter le nombre d'étoiles situées à une distance inférieure à s du système solaire O par une variable aléatoire de Poisson $N(s)$ de moyenne $\rho v(s)$. Revenons au problème de l'estimation de la distance D qui nous sépare de l'étoile la plus proche. Alors,

$$P(D > s) = P(N(s) = 0) = e^{-\rho v(s)} = e^{-4\pi\rho s^3/3}.$$

On peut en déduire la densité de la loi de D, et surtout la moyenne

$$E(D) = \int_0^{+\infty} P(D > t)\, dt = \int_0^{+\infty} e^{-4\pi\rho t^3/3}\, dt = (3/4\pi\rho)^{1/3}\Gamma(4/3).$$

En remplaçant ρ par sa valeur observée, on trouve $E(D) = 9,8$ années–lumières. Très remarquablement, la distance réelle est du même ordre de grandeur !!! puisque l'étoile Alpha du Centaure est à $4,3$ années–lumières de la Terre. On peut aussi remplacer D par la valeur D^* où sa densité est maximale. On trouve $D^* = (2\pi\rho)^{-1/3} = 9,6$ années-lumières.

Test du χ_2

D'après Pearson (1920). pp. 18–23 du poly AG.

à faire.

Exercices

Si X est une variable aléatoire, on note $\phi_X(t) = E(e^{itX})$.

1 Soit m la mesure sur $\mathbb{R}$ de densité $f(x) = c\,\mathbf{1}_{|x|\geqslant e}/(|x|^2 \log|x|)$. Montrer que $\widehat{m} \in C^1$ bien que $m(|x|)$ soit infini.

2 (Duplicité des fonctions caractéristiques)
1) Montrer que $\phi(t) = (1-|t|)^+$ est la fonction caractéristique d'une variable aléatoire X de loi absolument continue par rapport à la mesure de Lebesgue.
2) Développer ϕ en série de Fourier pour montrer qu'il existe une variable aléatoire discrète Y telle que $\phi_Y = \phi$ sur $[-1,+1]$.
3) En déduire qu'il existe des v.a. indépendantes X, Y et Z telles que Y et Z n'ont pas la même loi mais $X+Y$ et $X+Z$ ont la même loi.

3 (Opérations élémentaires et convergence en loi)
1) Exhiber des variables aléatoires telles que $X_n \xrightarrow{\mathcal{L}} X$ mais $P(X_n \to X) = 0$, et même

$$P(\exists\,(n_k),\ X_{n_k} \to X) = 0.$$

2) Exhiber des variables aléatoires telles que $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} Y$ et $X_n + Y_n$ ne converge pas en loi vers $X+Y$.
3) On suppose que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} a$ où a est une constante. Montrer que le couple (X_n, Y_n) converge en loi vers (X, a).
4) Si $X_n \xrightarrow{\mathcal{L}} X$ et $f \in C^0$ (pas forcément bornée, alors $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.
5) Si $X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{\mathcal{L}} b$ et $Z_n \xrightarrow{\mathcal{L}} a$, alors $Z_nX_n + Y_n \xrightarrow{\mathcal{L}} aX + b$.
6) Si X et Y ont la même loi et si f est mesurable, $f(X)$ et $f(Y)$ ont la même loi. Si X, Y et Z sont définies sur le même espace et si X et Y ont la même loi, est-ce que XZ et YZ ont la même loi ?

4 Utiliser la loi de Poisson de paramètre $n \geqslant 1$ pour montrer

$$\sum_{k=0}^{n} n^k/k! \sim e^n/2.$$

Quelles relations analogues pouvait-on montrer avec la LGN seule ?

5 (Supports et convergence en loi)
1) Soit X une variable aléatoire de loi uniforme sur $[0,1]$ et $X_n = \lfloor nX \rfloor / n$ pour $n \geqslant 1$. Montrer que $X_n \xrightarrow{\mathcal{L}} X$ et trouver un borélien B de $[0,1]$ tel que $P(X \in B) = 0$ et $P(\forall n, X_n \in B) = 1$.
2) Montrer que $P(X_n \in B) = 1$ pour tout n et $X_n \xrightarrow{\mathcal{L}} X$ entraîne $P(X \in B) = 1$ si l'on ajoute l'hypothèse B fermé.
3) Le support de la loi d'une variable aléatoire X est le plus petit fermé F tel que $P(X \in F) = 1$. Montrer qu'un tel ensemble minimal existe.
4) Un point de croissance de la loi de X est un point x tel que $P(X \in V) > 0$ pour tout voisinage V de x. Montrer que l'ensemble des points de croissance est exactement le support de la loi de X.

6 Soit $X_n \xrightarrow{\mathcal{L}} X$ des variables aléatoires réelles et Y une variable aléatoire réelle positive telle que $|X_n|$ est dominé stochastiquement par Y au sens suivant : $P(|X_n| \leqslant x) \geqslant P(Y \leqslant x)$ pour tout $x \geqslant 0$. Montrer que $E(X_n)$ converge vers $E(X)$.

7 Les chances de gain à la roulette sont de $18/37$ car il y a 18 numéros rouges, 18 noirs et le numéro 0 sur lequel on ne peut pas miser. Combien de fois doit-on jouer en misant un franc pour avoir au moins une chance (?) sur deux d'avoir perdu au moins mille francs ?

8 1) Soit X une variable aléatoire. Montrer que ϕ_X est continue.
2) Soit X et Y deux variables aléatoires indépendantes de même loi telles que $X+Y$ suit la loi de X. Utiliser ϕ_X pour trouver la loi de X.
3) Même question qu'au (2) avec $(X+Y)/\sqrt{2}$ à la place de $X+Y$.

9 Soit Z une variable aléatoire de loi uniforme sur $[-1,+1]$. Montrer que sa fonction caractéristique est $\phi_Z(t) = \sin t/t$, et en déduire qu'il n'existe pas de variables aléatoires indépendantes de même loi X et Y telles que $X-Y$ suit une loi uniforme sur $[-1,+1]$.

10 1) Sur $[0,1]$ muni de ses boréliens et de la mesure de Lebesgue, on définit $Z(\omega) = 2\omega - 1$. Donner la loi de Z.
2) On note $(R_n(\omega))_{n\geqslant 1}$ le développement dyadique de ω :

$$\omega = \sum_{n\geqslant 1} 2^{-n}\, R_n(\omega),$$

et on pose

$$Q_n = 2R_n - 1, \qquad U = \sum_{n\geqslant 0} 2^{-(2n+1)}\, Q_{2n+1}.$$

Trouver une variable aléatoire V indépendante de U, de même loi que U et telle que la variable aléatoire $U + (V/2)$ suit la loi uniforme sur $[-1,+1]$.
3) On étudie la réciproque. Soit X et Y deux variables aléatoires indépendantes de même loi telles que $X + (Y/2)$ suit la loi uniforme sur $[-1,+1]$. Calculer $\phi_X(t)/\phi_X(t/4)$. Montrer que X doit suivre la loi de U de la question (2). En déduire qu'il existe un borélien A de $[-1,+1]$ de mesure de Lebesgue nulle tel que $P(X \in A) = 1$.

11 Théorème de Bochner : Soit f une fonction continue sur $\mathbb{R}$ telle que $f(0) = 1$. Alors f est une fonction caractéristique si et seulement si f est définie positive au sens où, pour tous $z_k \in \mathbb{C}$ et tous $x_k \in \mathbb{R}$, on a :

$$\sum_{k,l=1}^{n} z_k \overline{z_l}\, f(x_k - x_l) \geqslant 0.$$

12 (Loi de distribution des vitesses de Maxwell)
Soit X_r de loi uniforme sur la boule $B(O,r)$ de $\mathbb{R}^n$ (pour la norme euclidienne).
1) Calculer la loi de $\|X_r\|$ et celle de $(X_{1,r}, \|X_r\|)$.
2) En déduire la loi conditionnelle de $X_{1,r}$ sachant que $\|X_r\|^2 = u$, où $u < r^2$. Cette loi, notée μ_u, est la loi de la vitesse à énergie u fixée.
3) Montrer que μ_{nu} converge étroitement vers $\mathcal{N}(0,u)$ quand $n \to \infty$.
4) Étendre ce résultat en montrant que la loi de $(X_{1,r}, \ldots, X_{p,r})$ sachant que $\|X_r\|^2 = nu$ converge étroitement vers $\mathcal{N}(0, uI_p)$ lorsque $r \to \infty$ et $n \to \infty$ avec $r^2 > nu$.

13 (TCL aléatoire)
Soit X_n des variables aléatoires indépendantes réduites, c'est-à-dire telles que, pour tout n, $E(X_n) = 0$ et $\mathrm{var}(X_n) = 1$. Soit $T(n)$ des variables aléatoires à valeurs entières telles que $T(n)/n$ converge en probabilité vers une constante $c \in]0,+\infty[$. Montrer, sans hypothèse d'indépendance portant sur les $T(n)$, que la suite

$S_{T(n)}/\sqrt{T(n)}$ converge en loi vers $\mathcal{N}(0,1)$.
On pourra commencer par montrer que $(S_{T(n)} - S_{\lfloor cn \rfloor})/\sqrt{n}$ converge en probabilité vers 0, en utilisant l'inégalité de Kolmogorov utilisée pour démontrer la LGN.

14 (Variantes du TCL)
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi telles que $E(X_n) = 0$ et $E(X_n^2) = 1$. On note

$$S_n = \sum_{k=1}^{n} X_k, \qquad C_n = \sum_{k=1}^{n} X_k^2, \qquad Z_n = \exp(-S_n)\,\mathbf{1}(S_n \geqslant 0).$$

1) Étudier la convergence en loi des suites de terme général $S_n/\sqrt{C_n}$ et $\sqrt{n}\,S_n/C_n$.
2) Déterminer la limite de $P(S_n \geqslant 0)$ quand n tend vers l'infini puis montrer que $(EZ_n)^{1/n}$ tend vers 1 quand n tend vers l'infini. (On pourra même montrer que $(EZ_n)^{1/\sqrt{n}}$ tend vers 1.)

15 Soit $(X_n)_n$ une suite u.i. telle que $X_n \xrightarrow{\mathcal{L}} X$. Montrer que $E(X_n)$ tend vers $E(X)$.

16 (TCL et loi de Poisson : une preuve "élémentaire" de la formule de Stirling)
Pour tout $a > 0$, on se donne une variable aléatoire X_a de loi $\mathcal{P}(a)$ et on note $Y_a = (X_a - a)/\sqrt{a}$.
1) Étudier la convergence en loi de Y_a quand a tend vers l'infini.
2) Pour $t \geqslant 0$, on note $G_a(t) = \sum_{k=0}^{[at]} e^{-a}a^k/k!$ Déterminer la limite de $G_a(t)$ quand a tend vers l'infini. En déduire la limite de

$$\sum_{k=0}^{[bat]} e^{-bx} b^k x^k/k!$$

pour $x > 0$ et $t \geqslant 0$ quand b tend vers l'infini.
3) Calculer $E((Y_n)^-)$. En déduire la formule de Stirling :

$$n! \sim \sqrt{2\pi n}(n/e)^n.$$

17 Loi gaussienne si et seulement si moyenne et variance empiriques sont indépendantes.

18 (Loi des petits nombres)
Pour chaque $n \geqslant 1$, on considère $S_n = \sum_{k\geqslant 0} X_{n,k}$, où $(X_{n,k})_{k\geqslant 0}$ est une suite de v.a. indépendantes de loi de Bernoulli $p_{n,k}\,\delta_1 + (1 - p_{n,k})\,\delta_0$. On suppose que $\sum_{k\geqslant 0} p_{n,k}$ tend vers $\lambda \neq 0$ et que $\max_{k\geqslant 0} p_{n,k}$ tend vers 0 quand n tend vers l'infini.
Montrer que $S_n \xrightarrow{\mathcal{L}} S$, où S suit une loi de Poisson $\mathcal{P}(\lambda)$.

Solutions

1 Soit $g(t) = \int_e^{+\infty} \cos(tx)k(x)dx$, $k(x) = 1/(x^2 \log x)$. Il faut montrer que $g \in C^1$.
1. Une intégration par parties donne $tg(t) = -e^{-2}\sin(et) + h(t)$ avec :

$$h(t) = \int_e^{+\infty} \sin(tx)\,m(x)\,dx, \qquad m(x) = -k'(x) = 2/(x^3 \log x) + 1/(x^3 \log^2 x).$$

En particulier, $0 \leqslant m(x) \leqslant 3/(x^3 \log x)$. Donc la borne $|(\partial/\partial t)\sin tx| \leqslant x$ et l'intégrabilité de $x \mapsto xm(x)$ donnent la dérivabilité de h par convergence dominée. De plus, $t \mapsto \cos(tx)xm(x)$ est continue donc $h \in C^1(\mathbb{R})$.

2. Montrons que g est dérivable en $t=0$. Comme g est paire, il faut montrer que $g(t)-g(0)$ est négligeable devant t. Alors, la dérivée en 0 existera et sera nulle. Or, $0 \leqslant 1-\cos(tx) \leqslant 2 \wedge (t^2x^2/2)$ donc

$$0 \leqslant g(0)-g(t) \leqslant (t^2/2)\int_e^{1/t} x^2k(x)dx + 2\int_{1/t}^{+\infty} k(x)dx.$$

Comme $x^2k(x)$ tend vers 0, la première intégrale est négligeable devant $1/t$ et la deuxième devant celle de $1/x^2$ sur $(1/t,+\infty)$, soit t, donc tout marche.
3. Continuité de g' en $t=0$: il faut montrer que $(h(t)/t)' = o(1)$ quand $t \to 0$. On a :

$$(h(t)/t)' = \int_e^{+\infty} \left[\frac{x}{t}(\cos(tx)-1) - \frac{1}{t^2}(\sin(tx)-tx)\right] m(x)\,dx.$$

Grâce à $0 \leqslant 1-\cos(tx) \leqslant 2\wedge(t^2x^2/2)$ et $0 \leqslant tx-\sin(tx) \leqslant tx \wedge (t^3x^3/6)$, il suffit de montrer que

$$t\int_{x\leqslant 1/t} x^3 m(x)dx \qquad \text{et} \qquad t^{-1}\int_{x\geqslant 1/t} xm(x)dx$$

sont $o(1)$. Ce sont les mêmes intégrales que celles qui impliquaient k.

Rappel : Soit $F(t) = \int f(t,x)dx$ et I un intervalle en t.
1. Si $|f(t,x)| \leqslant g(x)$ et $g \in L^1$, alors F est continue sur I.
2. Si $|\partial f(t,x)/\partial t| \leqslant g(x)$ et $g \in L^1$, alors F est dérivable sur I et sa dérivée vaut ce que l'on pense.
(Deux applications directes du théorème de convergence dominée.)

2 Loi Anon et loi $\frac{1}{2}\delta_0 + \sum_{n\geqslant 1}(2/n^2\pi^2)(\delta_{n\pi}+\delta_{-n\pi})$ où la somme ne porte que sur les n impairs.

3 1. $X_n = \varepsilon$ et $X = -\varepsilon$.
3. $(X_n, a) \xrightarrow{\mathcal{L}} (X, a)$ donc il suffit de majorer $E(e^{itX_n+isY_n}) - E(e^{itX_n+isa})$ (l'autre décomposition ne marche pas).
5. impliqué par 3. en décomposant l'addition et la multiplication.

6 Comme $X_n^{\pm}$ est aussi dominé par Y, on suppose $X_n \geqslant 0$. Alors $E(X_n)$ est l'intégrale de $P(X_n \geqslant x)$ et on a convergence presque partout (sur les points de continuité de F_Y) et dominée par $P(Y \geqslant x)$. Donc, si $Y \in L^1$, $E(X_n) \to E(Y)$. Ensuite, $X_n \geqslant 0$ dominé par Y implique $X_n \wedge t$ dominé par $Y \wedge t$.

7 Moyenne $-1/37$, variance inutile, donc 37.000 coups.

9 $|\phi_X(t)|^2 = \sin t/t$ qui prend des valeurs négatives !

10 3. $\phi_X(t)$ doit valoir le produit des $\cos(t/2^k)$ pour $k \geqslant 1$ impair, soit $\phi_U(t)$.

12 $\mu_u(dx) = c_{n-2}\,(1-x^2/u)^{(n-3)/2}\,\mathbf{1}_{|x|\leqslant\sqrt{u}}\,dx/\sqrt{u}$ avec $1/c_n = \int_{-\pi/2}^{\pi/2}(\cos t)^n\,dt$, intégrale de Wallis. Pour $n\to\infty$, la méthode de Laplace donne $1/c_n \sim \int_{\mathbb{R}} \exp(-nt^2/2)\,dt = \sqrt{2\pi/n}$.

13 Soit $S_t = S_{\lfloor t \rfloor}$. On majore $\{|S_{T(n)} - S_{cn}| \geqslant \varepsilon\sqrt{n}\}$ par $\{T(n) \notin ((c-\alpha)n, (c+\alpha)n)\}$, plus

$$P(\exists k \leqslant \alpha n, |S_{cn+k} - S_{cn}| \geqslant \varepsilon\sqrt{n})$$

plus un terme analogue pour $((c-\alpha)n, cn)$. Par l'inégalité de martingales de Kolmogorov et en notant $\widetilde{S}$ les sommes décalées, $P(\cdots) \leqslant E(\widetilde{S}^2_{\alpha n})/(n\varepsilon^2) \longrightarrow \alpha/\varepsilon^2$. Donc $S_{T(n)} - S_{cn}/\sqrt{n} \xrightarrow{P} 0$.
Ensuite, l'exercice 3 donne $(S_{T(n)} - S_{cn})/\sqrt{T(n)} \xrightarrow{P} 0$ donc il suffit de regarder $S_{cn}/\sqrt{T(n)}$ qui se comporte comme $S_{cn}/\sqrt{cn}$, à nouveau par l'exercice 3.

14 Pour le (2), $Z_n \leqslant 1$. Dans l'autre sens :

$$E(Z_n) \geqslant e^{-c\sqrt{n}}\,P(0 \leqslant S_n \leqslant c\sqrt{n}).$$

On élève tout à la puissance $1/\sqrt{n}$ et on estime la probabilité grâce au TCL. Donc, la limite inférieure de $E(Z_n)^{1/\sqrt{n}}$ est au moins e^{-c}, pour tout $c > 0$.

15 Tout d'abord, $X \in L^1$. Par exemple, $x \mapsto x^+ \wedge t$ est continue bornée donc $E(X_n^+ \wedge t)$ tend vers $E(X^+ \wedge t)$. Donc $E(X^+)$ vaut au plus $\limsup E(X_n^+) < +\infty$. Donc $\{X_n\,;\, n \geqslant 1\} \cup \{X\}$ est u.i. Soit t tel que $E(|Y|\,;\,|Y| \geqslant t) \leqslant \varepsilon$ pour tout Y dans cette famille. Alors, $f_t : x \mapsto (x \wedge t) \vee (-t)$ est continue bornée donc $E(f_t(X_n))$ tend vers $E(f_t(X))$. Finalement, la limite supérieure de $|E(X_n - X)|$ vaut au plus 4ε.

16 $E(Y_a^2) = \mathrm{var}(\mathcal{P}(1))$ donc $(Y_a)_a$ est u.i.
De plus, $E(Y_n^-) = \sqrt{n}\, e^{-n}\, n^n/n!$ donc tout marche.