

Cours G12 :

PRB : Probabilités de base

Espaces probabilisés, variables aléatoires	3
Indépendance	23
Convergence d'une suite de variables aléatoires	41
Loi des grands nombres	57
Théorème limite central	71

Philippe BRIAND

philippe.briand@univ-rennes1.fr

<http://perso.univ-rennes1.fr/philippe.briand/proba/proba.html>

Chapitre I. Espaces probabilisés, variables aléatoires

En 1933, N. Kolmogorov, s'appuyant sur la théorie de la mesure, a donné un cadre mathématique précis pour le calcul des probabilités permettant ainsi à d'autres mathématiciens, au premier rang desquels Paul Lévy, de développer l'analyse stochastique.

Ce chapitre d'introduction a pour but d'introduire les objets mathématiques de base utilisés pour le calcul des probabilités.

1. Vocabulaire probabiliste.

L'axiomatique de Kolmogorov repose sur la théorie de la mesure ; néanmoins, les origines du calcul des probabilités venant en partie de la Statistique, son vocabulaire est différent.

1.1. Espace probabilisé.

Il n'est pas toujours possible de prédire avec certitude le résultat d'une expérience, lancer d'un dé, d'une pièce de monnaie, sexe d'un enfant à naître. Voici le formalisme mathématique permettant de décrire une telle expérience aléatoire.

L'espace fondamental est un triplet généralement noté $(\Omega, \mathcal{F}, \mathbb{P})$. Ω est un ensemble non vide décrivant toutes les réalisations ou épreuves possibles. Par exemple, pour le lancer d'un dé on peut prendre l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}$. Un point $\omega \in \Omega$ est une réalisation particulière de cette expérience. Lors d'une réalisation de l'expérience, un « événement » peut se réaliser ou pas ; par exemple le résultat est pair ou impair. L'événement « le résultat est pair » se traduit au plan ensembliste par un sous ensemble A de Ω ici $\{2, 4, 6\}$ et A se réalise si $\omega \in A$. On aimerait associer à A un nombre $\mathbb{P}(A)$ qui rende compte des « chances » de réalisation de A . Pour cela, on doit en général se limiter à un sous-ensemble \mathcal{F} de parties de Ω : la tribu des événements et considérer une mesure de probabilité \mathbb{P} sur (Ω, \mathcal{F}) .

Commençons par quelques précisions.

Définition. Une tribu, \mathcal{F} , sur Ω – on parle également de σ -algèbre – est une classe de parties de Ω i.e. $\mathcal{F} \subset \mathcal{P}(\Omega)$ vérifiant :

- (i) $\Omega \in \mathcal{F}$;
- (ii) si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$;
- (iii) si $A_n \in \mathcal{F}$, pour tout $n \in \mathbf{N}$, alors $\cup_{\mathbf{N}} A_n \in \mathcal{F}$.

Dans ce cas, (Ω, \mathcal{F}) s'appelle un espace mesurable ou probabilisable et on parle d'événements pour désigner les éléments de \mathcal{F} . De plus, $\emptyset \in \mathcal{F}$ et \mathcal{F} est stable par union finie, par intersection finie et dénombrable, par différence ...

Exemple. $\mathcal{F}_1 = \{\emptyset, \Omega\}$ est une tribu sur Ω appelée tribu grossière ; $\mathcal{F}_2 = \mathcal{P}(\Omega)$, la collection des parties de Ω , est une tribu : la tribu discrète ; enfin, si $A \subset \Omega$, $\mathcal{F}_3 = \{\emptyset, \Omega, A, A^c\}$ est aussi une tribu.

Une intersection quelconque de tribus est encore une tribu. Si donc $\mathcal{C} \subset \mathcal{P}(\Omega)$, l'intersection de toutes les tribus contenant \mathcal{C} est une tribu qui par construction est la plus petite tribu pour l'inclusion contenant \mathcal{C} ; on l'appelle tribu engendrée par \mathcal{C} et on la note $\sigma(\mathcal{C})$.

Lorsque Ω est un espace topologique, la tribu engendrée par la classe des ouverts s'appelle la tribu borélienne de Ω ; on la note $\mathcal{B}(\Omega)$.

Exemple. La tribu $\mathcal{B}(\mathbf{R})$ est engendrée par les intervalles ouverts ou bien par les intervalles du type $]s, t]$ ou encore par ceux du type $] - \infty, t]$. La tribu $\mathcal{B}(\mathbf{R}^d)$ est, quant à elle, engendrée par les pavés du type $\prod_{i=1}^d]s^i, t^i]$.

Si, pour $i = 1, \dots, n$, $(\Omega_i, \mathcal{F}_i)$ est un espace mesurable, la tribu produit $\mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$ sur le produit cartésien $\Omega_1 \times \dots \times \Omega_n$ est la tribu engendrée par les pavés mesurables c'est à dire les ensembles de la forme $A_1 \times \dots \times A_n$, $A_i \in \mathcal{F}_i$. Par exemple, $\mathcal{B}(\mathbf{R}^d) = \otimes_{1 \leq i \leq d} \mathcal{B}(\mathbf{R})$.

Venons-en à présent à la mesure de probabilité \mathbb{P} .

Définition. Une mesure de probabilité, ou plus simplement probabilité, sur un espace mesurable (Ω, \mathcal{F}) est une application \mathbb{P} de \mathcal{F} dans $[0, 1]$ vérifiant

- (i) $\mathbb{P}(\emptyset) = 0$;
- (ii) σ -additivité : si $(A_n)_{\mathbf{N}} \subset \mathcal{F}$ et $A_n \cap A_m = \emptyset$ pour $n \neq m$, $\mathbb{P}(\cup_{\mathbf{N}} A_n) = \sum_{\mathbf{N}} \mathbb{P}(A_n)$.
- (iii) $\mathbb{P}(\Omega) = 1$.

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où (Ω, \mathcal{F}) est un espace mesurable et \mathbb{P} une probabilité sur cet espace s'appelle un espace probabilisé.

Exemple. Le premier exemple qui vient à l'esprit est celui de la mesure de Lebesgue sur $[0, 1]$. Comme vous l'avez vu, construire une probabilité sur $[0, 1]$ muni de la tribu borélienne vérifiant $\mathbb{P}(]s, t]) = t - s$ est loin d'être aisé. Cette mesure de probabilité a un rôle très important.

Sur $(\mathbf{N}, \mathcal{P}(\mathbf{N}))$, la donnée d'une mesure de probabilité est simplement celle d'une suite de réels positifs p_n tels que $\sum_{n \geq 0} p_n = 1$. On a alors $\mathbb{P}(A) = \sum_{n \geq 0} p_n \mathbf{1}_A(n)$ c'est à dire $\mathbb{P} = \sum_{n \geq 0} p_n \delta_n$ où δ_n désigne la masse de Dirac en n .

Une probabilité est additive, c'est à dire : si A et B sont deux éléments disjoints de \mathcal{F} i.e. $A \cap B = \emptyset$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. En particulier, \mathbb{P} est croissante sur \mathcal{F} i.e. si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$ et $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Voici quelques propriétés supplémentaires.

Proposition 1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Pour $(A_n)_{\mathbf{N}} \subset \mathcal{F}$, on a

1. $\mathbb{P}(\cup_n A_n) \leq \sum_{n \geq 0} \mathbb{P}(A_n)$;
2. si $A_n \subset A_{n+1}$ pour tout $n \in \mathbf{N}$, $\mathbb{P}(\cup_{\mathbf{N}} A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \sup_{n \geq 0} \mathbb{P}(A_n)$;
3. si $A_{n+1} \subset A_n$ pour tout $n \in \mathbf{N}$, $\mathbb{P}(\cap_{\mathbf{N}} A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \inf_{n \geq 0} \mathbb{P}(A_n)$.

Démonstration. Les deux premiers points résultent de la construction suivante : on pose $B_0 = A_0$ et $B_n = A_n \setminus \cup_{i \leq n-1} A_i$. Alors les B_n sont deux à deux disjoints et $\cup_{i \leq n} B_i = \cup_{i \leq n} A_i$, pour tout $n \in \mathbf{N}$, de même que $\cup_{n \geq 0} B_n = \cup_{n \geq 0} A_n$. Le dernier point s'obtient par passage au complémentaire. \square

Rappelons que si $(A_n)_{\mathbf{N}}$ est une suite de parties de \mathcal{F} , ses limites supérieures et inférieures sont les événements définis par

$$\limsup A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k, \quad \liminf A_n = \bigcup_{n \geq 0} \bigcap_{k \geq n} A_k;$$

cf. $\limsup x_n = \inf_{n \geq 0} \sup_{k \geq n} x_k$ et $\liminf x_n = \sup_{n \geq 0} \inf_{k \geq n} x_k$ pour une suite $(x_n)_{\mathbf{N}}$ de $\overline{\mathbf{R}}$.

$\omega \in \limsup A_n$ signifie que ω appartient à une infinité de A_n et $\omega \in \liminf A_n$ signifie que ω appartient à tous les A_n sauf un nombre fini. On a d'autre part, $(\limsup A_n)^c = \liminf A_n^c$ et $(\liminf A_n)^c = \limsup A_n^c$.

Voici un résultat très utile en pratique connu sous le nom de lemme de Borel-Cantelli.

Lemme 2 (Borel–Cantelli). *Soit $(A_n)_{\mathbf{N}} \subset \mathcal{F}$. Si $\sum_{n \geq 0} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup A_n) = 0$.*

Démonstration. Pour tout $n \in \mathbf{N}$, $\limsup A_n \subset \bigcup_{k \geq n} A_k$. Comme $\mathbb{P}(\bigcup_{k \geq n} A_k) \leq \sum_{k \geq n} \mathbb{P}(A_k)$, on a

$$\forall n \in \mathbf{N}, \quad \mathbb{P}(\limsup A_n) \leq \sum_{k \geq n} \mathbb{P}(A_k),$$

qui est le reste d'une série convergente. □

D'autre part, la Proposition 1 implique que

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

Exercice. Démontrer les inégalités précédentes.

Remarque. Lorsqu'une propriété est vraie pour tout $\omega \in N^c$ avec $\mathbb{P}(N) = 0$ ou pour tout $\omega \in \Omega_1$ avec $\mathbb{P}(\Omega_1) = 1$ on dit que la propriété est vraie \mathbb{P} -presque sûrement ou avec probabilité un.

1.2. Variable aléatoire.

Imaginons une personne jouant une partie de dés en dix coups : il gagne deux euros s'il obtient un 5 ou un 6, un euro pour un quatre, rien pour un 3, perd un euro pour un 2 et trois euros s'il obtient 1. Si on cherche à modéliser cette expérience, on peut prendre $\Omega = \{1, \dots, 6\}^{10}$, $\mathcal{F} = \mathcal{P}(\Omega)$ et si le dé est équilibré $\mathbb{P}(\{\omega\}) = 6^{-10}$ pour tout $\omega \in \Omega$. Du point de vue du joueur, le plus important n'est pas le résultat de l'expérience en lui-même (la succession des valeurs du dé) mais il s'intéresse réellement au gain que va lui donner cette expérience. Bien sûr, son gain G dépend du résultat de l'expérience : il s'agit donc d'une application définie sur Ω qui, ici, est à valeurs dans $\{-30, \dots, 20\}$. On parle de variable aléatoire.

Commençons par quelques précisions. Soit X une application de Ω dans E . Si $B \subset E$, l'image réciproque de B par X , $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$, sera noté $\{X \in B\}$ et, si $\mathcal{C} \subset \mathcal{P}(E)$, on rappelle que $X^{-1}(\mathcal{C})$ désigne $\{\{X \in B\} : B \in \mathcal{C}\}$.

Définition. Soient (Ω, \mathcal{F}) et (E, \mathcal{E}) deux espaces mesurables. On appelle variable aléatoire – v.a. dans la suite – définie sur (Ω, \mathcal{F}) à valeurs dans (E, \mathcal{E}) toute application $X : \Omega \rightarrow E$ $(\mathcal{F}, \mathcal{E})$ -mesurable c'est à dire vérifiant $X^{-1}(\mathcal{E}) \subset \mathcal{F}$ soit

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

Exemple. Une v.a. X est étagée si elle ne prend qu'un nombre fini de valeurs x_1, \dots, x_n . On a dans ce cas, $X = \sum_{1 \leq i \leq n} x_i \mathbf{1}_{A_i}$ avec $A_i = \{X = x_i\} \in \mathcal{F}$. Une v.a. étagée est donc une combinaison linéaire d'indicatrices ; il n'y a pas unicité de la représentation.

Remarque. Dans la plupart des cas, l'espace (Ω, \mathcal{F}) est sous-entendu et lorsqu'il n'y a pas d'ambiguïté sur les tribus on dit simplement que X est une v.a. à valeurs dans E voire même que X est une v.a.

D'autre part, si E est un espace topologique, en l'absence d'indication contraire, la tribu considérée sur E sera la tribu borélienne.

On parle de v.a. positive si $E = \overline{\mathbf{R}}_+$, de v.a. numérique si $E = \overline{\mathbf{R}}$, de v.a. réelle ou v.a.r. si $E = \mathbf{R}$, de vecteur aléatoire ou v.a. vectorielle si $E = \mathbf{R}^d$ et enfin de v.a. entière si $E = \mathbf{N}$.

Proposition 3. Si $\mathcal{E} = \sigma(\mathcal{C})$, X est une v.a. si et seulement si $X^{-1}(\mathcal{C}) \subset \mathcal{F}$.

Démonstration. Supposons $X^{-1}(\mathcal{C}) \subset \mathcal{F}$. Puisque $\{X \in \cup_{\mathbf{N}} B_n\} = \cup_{\mathbf{N}} \{X \in B_n\}$ et $\{X \in B\}^c = \{X \in B^c\}$, $\{B \in \mathcal{E} : \{X \in B\} \in \mathcal{F}\}$ est une tribu sur E qui contient \mathcal{C} . Elle contient donc $\sigma(\mathcal{C}) = \mathcal{E}$. \square

Exemple. Si $\{X \leq t\} \in \mathcal{F}$ pour tout réel t alors X est une v.a. réelle puisque la tribu borélienne de \mathbf{R} est engendrée par les intervalles $] - \infty, t]$.

Comme la tribu borélienne de \mathbf{R}^d est engendrée par les pavés $\prod_{1 \leq i \leq d}] - \infty, t^i]$, $t^i \in \mathbf{R}$, une application X à valeurs dans \mathbf{R}^d est un vecteur aléatoire si et seulement si chacune de ses composantes sont des v.a. réelles.

Rappelons que si X est une v.a. dans E et $u : E \rightarrow E'$ une application mesurable, alors $u(X)$ est une v.a. dans E' . En particulier, la somme, le produit, le quotient (lorsqu'il est défini) de v.a. réelles est encore une v.a. Si X est une v.a.r. sa partie positive $X^+ = \max(X, 0)$, sa partie négative $X^- = \max(-X, 0)$, si X est une v.a. complexe, sa partie réelle, sa partie imaginaire, son module sont également des v.a. D'autre part si X_n est une suite de v.a. numériques, $\inf X_n$, $\sup X_n$, $\liminf X_n$, $\limsup X_n$ sont aussi des v.a.

Exercice. Soient $(X_n)_{\mathbf{N}}$ une suite de v.a. complexes et $(\varepsilon_n)_{\mathbf{N}}$ une suite de réels positifs tendant vers 0. Montrer que, si $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon_n) < +\infty$, la suite $(X_n)_{\mathbf{N}}$ converge presque sûrement vers 0.

1.3. Espérance.

Comme nous venons de le voir, une v.a. – disons positive pour l'instant – n'est rien d'autre qu'une application mesurable à valeurs dans $\overline{\mathbf{R}}_+$; on peut alors définir son intégrale. En langage probabiliste, on parle d'espérance. Rappelons qu'une v.a. numérique ou complexe est intégrable si

$$\int_{\Omega} |X(\omega)| \mathbb{P}(d\omega) < +\infty.$$

Définition. Soit X une v.a. positive ou intégrable. On appelle espérance de X , que l'on note $\mathbb{E}[X]$, l'intégrale de X par rapport à \mathbb{P} soit

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} X d\mathbb{P}.$$

Remarque. Il faut bien noter que comme une mesure de probabilité est de masse finie, les constantes sont des v.a. intégrables.

Avant de poursuivre, rappelons le mécanisme de l'intégration; je vous renvoie à [Rud87] ou au cours d'intégration de F. MALRIEU [Mal] disponible en ligne à l'adresse

Si X est une v.a. étagée positive i.e. $X = \sum_{1 \leq i \leq n} x_i \mathbf{1}_{A_i}$, $x_i \in \mathbf{R}_+$, $A_i \in \mathcal{F}$, on pose

$$\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{P}(A_i) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) ;$$

le résultat est indépendant de la représentation considérée.

Si X est une v.a. positive, l'espérance de X est l'élément de $\overline{\mathbf{R}}_+$ défini par

$$\mathbb{E}[X] = \sup\{\mathbb{E}[Y] : Y \text{ v.a. étagée positive, } Y \leq X\}.$$

En fait, toute v.a. positive est limite simple croissante d'une suite de v.a. étagées positives. En effet, si $x \geq 0$,

$$\beta_n(x) = \min(2^{-n} \lceil 2^n x \rceil, n) = \sum_{i=1}^{n2^n} (i-1) 2^{-n} \mathbf{1}_{\frac{i-1}{2^n} \leq x < \frac{i}{2^n}} + n \mathbf{1}_{x \geq n} \quad (1)$$

converge en croissant vers x . Les v.a. $X_n = \beta_n(X)$ sont étagées, positives et convergent simplement en croissant vers X . On montre alors que $\mathbb{E}[X] = \lim_{n \rightarrow +\infty} \mathbb{E}[X_n]$.

Enfin si X est une v.a. numérique telle que $\mathbb{E}[|X|] < +\infty$,

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-], \quad x^+ = x \vee 0 = \max(x, 0), \quad x^- = \max(-x, 0).$$

Finalement pour X v.a. complexe telle que $\mathbb{E}[|X|] < +\infty$, $\mathbb{E}[X] = \mathbb{E}[\operatorname{Re}(X)] + i \mathbb{E}[\operatorname{Im}(X)]$.

Tous les résultats de convergence se traduisent en terme d'espérance. En guise d'illustration, rappelons les trois principaux. Soit $(X_n)_{n \geq 0}$ une suite de v.a. numériques.

Convergence monotone. Si, pour tout $n \geq 0$, $0 \leq X_n \leq X_{n+1}$,

$$\mathbb{E} \left[\sup_{n \geq 0} X_n \right] = \sup_{n \geq 0} \mathbb{E}[X_n].$$

Lemme de Fatou. Si, pour tout $n \in \mathbf{N}$, $X_n \geq 0$,

$$\mathbb{E} [\liminf X_n] \leq \liminf \mathbb{E}[X_n].$$

Convergence dominée. On suppose que, X_n converge vers X presque sûrement. S'il existe une v.a. Y telle que, pour tout $n \in \mathbf{N}$, $|X_n| \leq Y$ p.s. et $\mathbb{E}[Y] < +\infty$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|X_n - X|] = 0, \quad \text{et donc} \quad \lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Exercice. Traduire les résultats sur les intégrales à paramètres en termes de v.a. et d'espérance.

Soit X une v.a réelle. Montrer que $\varphi_X(t) = \mathbb{E} \left[e^{itX} \right]$ est continue sur \mathbf{R} .

2. Loi d'une variable aléatoire.

2.1. Généralités.

Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probablisé et X une v.a. à valeurs dans (E, \mathcal{E}) . On pose, pour tout $B \in \mathcal{E}$,

$$\mathbb{P}_X(B) = \mathbb{P}(\{X \in B\}) = \mathbb{P}(X \in B).$$

Proposition 4. \mathbb{P}_X est une probabilité sur (E, \mathcal{E}) : c'est la probabilité image de \mathbb{P} par X .

Démonstration. On a $\mathbb{P}_X(\emptyset) = \mathbb{P}(X \in \emptyset) = \mathbb{P}(\emptyset) = 0$, $\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \mathbb{P}(\Omega) = 1$. De plus, si $(B_n)_{\mathbf{N}}$ est une suite de parties deux à deux disjointes de \mathcal{E} , les événements $\{X \in B_n\}$ sont deux à deux disjoints et par suite

$$\mathbb{P}_X(\cup_{\mathbf{N}} B_n) = \mathbb{P}(X \in \cup_{\mathbf{N}} B_n) = \mathbb{P}(\cup_{\mathbf{N}} \{X \in B_n\}) = \sum_{n \geq 0} \mathbb{P}(X \in B_n) = \sum_{n \geq 0} \mathbb{P}_X(B_n).$$

□

Définition. Soit X une variable aléatoire à valeurs dans (E, \mathcal{E}) . On appelle loi (ou distribution) de X , notée \mathbb{P}_X , la probabilité sur (E, \mathcal{E}) définie par

$$\forall B \in \mathcal{E}, \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Remarque. Nous étudierons principalement des variables aléatoires à valeurs dans \mathbf{R}^d et donc la loi d'une telle v.a. est une mesure de probabilité sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$.

Proposition 5. Soient X une v.a. dans E et $f : E \rightarrow \overline{\mathbf{R}}_+$ une application mesurable. Alors,

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

Si $f : E \rightarrow \overline{\mathbf{R}}$ ou \mathbf{C} est mesurable, la v.a. $f(X)$ est intégrable si et seulement si f est \mathbb{P}_X -intégrable sur E et dans ce cas la formule précédente est valable.

Démonstration. Si f est une fonction étagée positive, $f(x) = \sum_{1 \leq i \leq n} y_i \mathbf{1}_{B_i}(x)$, $B_i = \{x \in E : f(x) = y_i\}$, alors $f(X) = \sum_{1 \leq i \leq n} y_i \mathbf{1}_{\{X \in B_i\}}$ et

$$\mathbb{E}[f(X)] = \sum_{i=1}^n y_i \mathbb{P}(X \in B_i) = \sum_{i=1}^n y_i \mathbb{P}_X(B_i) = \int_E f(x) \mathbb{P}_X(dx).$$

Si f est mesurable positive, alors f est limite simple d'une suite croissante de fonctions étagées positives $(f_n)_{\mathbf{N}}$ - $f_n = \beta_n(f)$ par exemple cf. (1). D'après l'étape précédente, pour tout $n \in \mathbf{N}$, $\mathbb{E}[f_n(X)] = \int_E f_n(x) \mathbb{P}_X(dx)$ et par convergence monotone,

$$\mathbb{E}[f(X)] = \lim_{n \rightarrow +\infty} \mathbb{E}[f_n(X)] = \lim_{n \rightarrow +\infty} \int_E f_n(x) \mathbb{P}_X(dx) = \int_E f(x) \mathbb{P}_X(dx).$$

Pour la fin de la preuve, on utilise $f = f^+ - f^-$ puis $f = \operatorname{Re}(f) + i\operatorname{Im}(f)$. □

Dire que X a pour loi μ signifie donc que $\mathbb{P}(X \in B) = \mu(B)$ pour tout B de \mathcal{E} et dans ce cas, si $f : E \rightarrow \overline{\mathbf{R}}$ ou \mathbf{C} est mesurable, positive ou μ -intégrable,

$$\mathbb{E}[f(X)] = \int_E f(x) \mu(dx).$$

De plus, si la relation précédente est valable pour tout fonction f mesurable positive et bornée - $f \in \mathcal{M}_b^+$ - alors $\mathbb{P}_X = \mu$; il suffit de prendre $f(x) = \mathbf{1}_B(x)$, $B \in \mathcal{E}$.

2.2. Exemples.

Avant de donner quelques exemples, si f est une fonction mesurable positive ou intégrable par rapport à μ , l'intégrale de f par rapport à μ sera notée indifféremment

$$\int_E f d\mu \quad \int_E f(x) d\mu(x) \quad \int_E f(x) \mu(dx).$$

1. Variable discrète. Une mesure de probabilité μ est discrète s'il existe une partie au plus dénombrable D telle que $\mu(D^c) = 0$ soit $\mu(D) = 1$. Une v.a. X est discrète si sa loi l'est c'est à dire si $\mathbb{P}(X \in D) = 1$. On dit – un peu abusivement – que X est à valeurs dans D . μ est entièrement caractérisée par les valeurs $\mathbb{P}(X = d)$, $d \in D$ puisque $\mu(dx) = \sum_{d \in D} \mathbb{P}(X = d) \delta_d(dx)$, où δ_d désigne la masse de Dirac au point d ; dans ce cas,

$$\mathbb{E}[f(X)] = \int_E f(x) \mu(dx) = \sum_{d \in D} f(d) \mathbb{P}(X = d).$$

Exemple. Loi de Poisson de paramètre $\alpha > 0$. Pour tout $k \in \mathbf{N}$, posons $p_k = e^{-\alpha} \alpha^k / k!$. La probabilité sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, $\mu = \sum_{k \geq 0} p_k \delta_k$ est appelée loi de Poisson. On dit que X suit la loi de Poisson, $\mathcal{P}(\alpha)$, $\alpha > 0$, si $\mathbb{P}(X = k) = p_k$, pour tout $k \in \mathbf{N}$.

Voici d'autres exemples de lois discrètes que nous rencontrerons. Si $p \in [0, 1]$, on note $q = 1 - p$.

Loi de Bernoulli, $\mathcal{B}(p)$, $0 \leq p \leq 1$: X ne prend que deux valeurs 0 et 1 et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = q$ i.e. $\mu = p \delta_1 + (1 - p) \delta_0$;

Loi binomiale, $\mathcal{B}(n, p)$, $n \geq 1$, $0 \leq p \leq 1$: pour $k = 0, \dots, n$, $\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$;

Loi géométrique, $\mathcal{G}(p)$, $0 < p < 1$: pour $k \in \mathbf{N}^*$, $\mathbb{P}(X = k) = p q^{k-1}$;

Loi binomiale négative, $\mathcal{B}_-(n, p)$, $n \geq 1$, $0 < p < 1$: pour $k \geq n$, $\mathbb{P}(X = k) = C_{k-1}^{n-1} p^n q^{k-n}$;

Loi de Poisson, $\mathcal{P}(\alpha)$, $\alpha > 0$: pour $k \in \mathbf{N}$, $\mathbb{P}(X = k) = e^{-\alpha} \alpha^k / k!$.

2. Variable à densité. Soient ν une mesure sur (E, \mathcal{E}) et p une fonction mesurable positive; on dit que p est une *densité de probabilité* par rapport à ν si $\int_E p(x) \nu(dx) = 1$. Si μ est une probabilité sur (E, \mathcal{E}) , on dit que μ a pour densité p par rapport à ν si

$$\forall B \in \mathcal{E}, \quad \mu(B) = \int_B p(x) \nu(dx).$$

Si μ possède une densité par rapport à ν celle ci est unique à un ensemble de ν -mesure nulle près.

On dit que X a pour densité p par rapport à ν s'il en est ainsi de \mathbb{P}_X c'est à dire si

$$\forall B \in \mathcal{E}, \quad \mathbb{P}(X \in B) = \int_B p(x) \nu(dx).$$

On a, alors,

$$\mathbb{E}[f(X)] = \int_E f(x) p(x) \nu(dx).$$

Si $E = \mathbf{R}^d$ et qu'aucune mesure de référence ν n'est donnée, il s'agit de la mesure de Lebesgue λ_d .

Si X a pour densité $\mathbf{1}_S(x)p(x)$, on vérifie sans peine que $\mathbb{P}(X \in S) = 1$ et par abus on considère que X est à valeurs dans S .

Exemple. Nous rencontrerons bon nombre de probabilités possédant une densité par rapport à la mesure de Lebesgue. Citons la loi normale centrée réduite qui est la probabilité μ sur \mathbf{R}^d de densité par rapport à λ_d

$$p(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}},$$

$|x|$ désignant la norme euclidienne de x .

Voici d'autres exemples. Toutes les lois sont considérées sur \mathbf{R} muni de la tribu borélienne.

Loi uniforme $\mathcal{U}(a, b)$: $p(x) = (b - a)^{-1} \mathbf{1}_{[a, b]}(x)$;

Loi de Cauchy, $\mathcal{C}(\alpha)$, $\alpha > 0$: $p(x) = \alpha / (\pi(\alpha^2 + x^2))$;

Loi de Laplace ou double exponentielle : $p(x) = \alpha e^{-\alpha|x|}/2$, $\alpha > 0$;

Loi exponentielle, $\mathcal{Exp}(\alpha)$, $\alpha > 0$: $p(x) = \alpha \exp(-\alpha x) \mathbf{1}_{\mathbf{R}_+}(x)$;

Loi gaussienne ou normale $\mathcal{N}(m, \sigma^2)$: $p(x) = (2\pi\sigma^2)^{-1/2} \exp(-(x - m)^2/(2\sigma^2))$.

Loi normale centrée réduite dans \mathbf{R}^d : $p(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$.

Comme nous le verrons plus loin, certaines v.a. ne sont pas discrètes et ne possèdent pas pour autant de densité.

2.3. Loi image.

Nous abordons ici un point que nous reprendrons ultérieurement celui du calcul de loi. Nous commençons par un point très simple.

1. Lois marginales. Soit $X = (X_1, X_2)$ une v.a. à valeurs dans $(E = E_1 \times E_2, \mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2)$. La loi de X est une probabilité sur (E, \mathcal{E}) . On appelle loi marginale de X – on devrait dire de \mathbb{P}_X – les lois de X_1 et X_2 . La loi du couple (X_1, X_2) permet de déterminer les lois marginales puisque, par exemple pour X_1 , si $B_1 \in \mathcal{E}_1$,

$$\mathbb{P}_{X_1}(B_1) = \mathbb{P}(X_1 \in B_1) = \mathbb{P}(X_1 \in B_1, X_2 \in E_2) = \mathbb{P}_X(B_1 \times E_2).$$

En particulier, si X à valeurs dans $\mathbf{R}^{d_1+d_2}$, a pour densité $p(x_1, x_2)$, X_1 et X_2 ont pour densités respectives

$$p_1(x_1) = \int_{\mathbf{R}^{d_2}} p(x_1, x_2) dx_2, \quad p_2(x_2) = \int_{\mathbf{R}^{d_1}} p(x_1, x_2) dx_1.$$

En effet, soit $f : \mathbf{R}^{d_1} \rightarrow \mathbf{R}_+$ borélienne bornée. On a, $f(X_1) = g(X)$ avec $g(x_1, x_2) = f(x_1)$ et

$$\mathbb{E}[f(X_1)] = \mathbb{E}[g(X)] = \int_{\mathbf{R}^{d_1+d_2}} g(x_1, x_2) p(x_1, x_2) dx_1 dx_2 = \int_{\mathbf{R}^{d_1}} f(x_1) \left(\int_{\mathbf{R}^{d_2}} p(x_1, x_2) dx_2 \right) dx_1.$$

La généralisation à $X = (X_1, \dots, X_n)$ est immédiate.

Exercice. Soit X une v.a. normale centrée réduite dans \mathbf{R}^d . Quelles sont les lois marginales de X ?

2. Calcul de lois. Le problème est le suivant. La donnée est une v.a. X à valeurs dans E dont la loi est connue, disons μ . On cherche à déterminer la loi de la v.a. $Y = u(X)$ où u est une fonction mesurable de E dans E' .

Pour déterminer la loi de Y , on peut calculer, pour $f \in \mathcal{M}_b^+$, $\mathbb{E}[f(Y)] = \mathbb{E}[f(u(X))]$ en partant de

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(u(X))] = \int_E f(u(x)) \mu(dx)$$

et avec pour objectif la formule

$$\mathbb{E}[f(Y)] = \int_{E'} f(y) \nu(dy)$$

où ν est une probabilité sur E' . Dans ce cas, $\mathbb{P}_Y = \nu$.

Exemple. Soit X une v.a. de loi de Cauchy, $\mathcal{C}(1)$, c'est à dire de densité $\pi^{-1}(1+x^2)^{-1}$ sur \mathbf{R} . Cherchons la loi de $Y = X^+$. Soit $f \in \mathcal{M}_b^+$.

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(X^+)] = \int_{\mathbf{R}} f(x^+) \frac{1}{\pi} \frac{1}{1+x^2} dx = \int_{]-\infty, 0[} f(0) \frac{1}{\pi} \frac{1}{1+x^2} dx + \int_{\mathbf{R}_+} f(x) \frac{1}{\pi} \frac{1}{1+x^2} dx,$$

et donc

$$\mathbb{E}[f(Y)] = \frac{1}{2} f(0) + \int_{\mathbf{R}} f(x) \frac{1}{\pi} \frac{1}{1+x^2} \mathbf{1}_{x \geq 0} dx.$$

Par conséquent, on a, pour $f \in \mathcal{M}_b^+$,

$$\mathbb{E}[f(Y)] = \int_{\mathbf{R}} f(y) \nu(dy), \quad \text{avec } \nu(dy) = \frac{1}{2} \delta_0(dy) + \frac{1}{\pi} \frac{1}{1+y^2} \mathbf{1}_{y \geq 0} dy.$$

D'où $\mathbb{P}_Y = \nu$. Ici on a considéré que Y était une v.a. réelle. De plus Y n'est pas discrète et ne possède pas de densité par rapport à la mesure de Lebesgue.

Cette approche est particulièrement efficace dans le cas suivant : X est à valeurs dans \mathbf{R}^d et possède une densité de la forme $\mathbf{1}_U(x) p(x)$ où U est un ouvert et enfin u est un \mathcal{C}^1 -difféomorphisme sur U . On peut alors calculer $\mathbb{E}[f(u(X))]$ à l'aide de la formule du changement de variable ; cf. [Mal] pour les détails. Donnons un exemple dans \mathbf{R} pour ne pas compliquer les calculs.

Exemple. Soit X une v.a.r. de densité $\mathbf{1}_{]0,1[}(x)$ (X uniforme sur $]0,1[$). Cherchons la loi de la v.a. $Y = -\ln X$. Pour $f \in \mathcal{M}_b^+$, le changement de variable $y = -\ln x$ donne,

$$\mathbb{E}[f(-\ln X)] = \int_{]0,1[} f(-\ln x) dx = \int_{\mathbf{R}_+} f(y) e^{-y} dy = \int_{\mathbf{R}} f(y) e^{-y} \mathbf{1}_{y \geq 0} dy.$$

Y a pour densité $y \mapsto e^{-y} \mathbf{1}_{y \geq 0}$; c'est une v.a. de loi exponentielle de paramètre 1.

Lorsque Y est discrète, schématiquement Y prend un nombre au plus dénombrable de valeurs $(y_n)_{\mathbf{N}}$, il suffit de calculer $\mathbb{P}(Y = y_i)$ pour déterminer sa loi.

Exercice. Soit X une v.a. exponentielle de paramètre $\alpha > 0$. Calculer la loi de la v.a. $Y = 1 + [X]$.

2.4. Égalité en loi.

Soient X et Y deux v.a. à valeurs dans E . On dit que X et Y sont égales en loi (ou en distribution) ou encore que X et Y sont identiquement distribuées si $\mathbb{P}_X = \mathbb{P}_Y$ soit

$$\forall B \in \mathcal{E}, \quad \mathbb{P}(X \in B) = \mathbb{P}(Y \in B), \quad \text{on note } X \stackrel{(d)}{=} Y.$$

Deux variables aléatoires définies sur le même espace peuvent avoir la même loi et « être très différentes ». Par exemple, sur $[0, 1]$ muni de la tribu borélienne et de la mesure de Lebesgue, si $X(\omega) = \omega$ et $Y(\omega) = 2 \min(\omega, 1 - \omega)$, alors X et Y sont de loi uniforme sur $[0, 1]$. En effet, c'est évident pour X et pour Y , si $f \in \mathcal{M}_b^+$,

$$\mathbb{E}[f(Y)] = \int_0^1 f(2 \min(\omega, 1 - \omega)) d\omega = \int_0^{\frac{1}{2}} f(2\omega) d\omega + \int_{\frac{1}{2}}^1 f(2(1 - \omega)) d\omega = \int_0^1 f(\omega) d\omega.$$

Notons que deux variables X et Y peuvent avoir la même loi sans être définies sur le même espace probabilisé. Voici un exemple très simple. On considère tout d'abord le triplet $\Omega = \{0, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P} = \frac{1}{2}(\delta_0 + \delta_1)$ et on considère la v.a. $X(\omega) = \omega$. On a alors $\mathbb{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$. Maintenant, on prend $\Omega' =]0, 1[$, $\mathcal{F}' = \mathcal{B}(]0, 1[)$, \mathbb{P}' la mesure de Lebesgue sur Ω' et $Y(\omega) = [2\omega]$. On a $\mathbb{P}'(Y = 0) = \mathbb{P}'(]0, 1/2[) = 1/2$, $\mathbb{P}'(Y = 1) = \mathbb{P}'([1/2, 1[) = 1/2$. D'où, $\mathbb{P}'_Y = \mathbb{P}_X$.

3. Identification de deux probabilités.

Dans bien des applications, il est important d'identifier la loi d'une v.a. D'un point de vue général, le problème est le suivant : étant données deux probabilités μ et ν sur (E, \mathcal{E}) à quelles conditions sont-elles égales ? Du point de vue des v.a., on cherche des conditions si possibles simples permettant de répondre aux questions :

- (a) X suit-elle la loi μ ?
- (b) X et Y ont-elles même loi ?

3.1. Égalité de deux probabilités.

Nous allons dans ce paragraphe, démontrer un résultat général sur l'égalité de deux mesures de probabilité. Il repose sur un argument dit de classe monotone. Soit (E, \mathcal{E}) un espace mesurable.

Définition. Soit $\mathcal{D} \subset \mathcal{P}(E)$. On dit que \mathcal{D} est un λ -système si :

- (i) $E \in \mathcal{D}$;
- (ii) si $(A, B) \in \mathcal{D}^2$ et $A \subset B$ alors $B \setminus A \in \mathcal{D}$;
- (ii) si $(A_n)_{\mathbb{N}} \subset \mathcal{D}$ est croissante alors $\cup_{\mathbb{N}} A_n \in \mathcal{D}$.

Un π -système est simplement une classe de parties non vide stable par intersection finie. Un π -système contenant E est dit complet.

Théorème 6 (Dynkin). *Soit \mathcal{C} un π -système. Tout λ -système contenant \mathcal{C} contient $\sigma(\mathcal{C})$.*

Démonstration. Comme dans le cas des tribus, on montre qu'une intersection de λ -systèmes est encore un λ -système : ceci permet de définir le plus petit λ -système contenant \mathcal{C} que nous noterons \mathcal{D} . Il suffit donc de montrer que \mathcal{D} contient $\sigma(\mathcal{C})$.

Pour cela, nous allons montrer que \mathcal{D} est une tribu. Un λ -système stable par intersection finie est une tribu (preuve laissée en exercice). Il suffit de montrer que \mathcal{D} est un π -système. Considérons

$$\mathcal{D}_1 = \{B \in \mathcal{D}, \quad \forall A \in \mathcal{C}, B \cap A \in \mathcal{D}\}.$$

Comme \mathcal{D} contient \mathcal{C} qui est stable par intersection finie, on a $\mathcal{C} \subset \mathcal{D}_1$. Montrons que \mathcal{D}_1 est un λ -système. $E \in \mathcal{D}_1$. Soient B et C dans \mathcal{D}_1 , $B \subset C$. On a $A \cap (C \setminus B) = (A \cap C) \setminus (A \cap B)$. Or, si $A \in \mathcal{C}$, $A \cap B$ et $A \cap C$ sont éléments de \mathcal{D} et $A \cap B \subset A \cap C$; \mathcal{D} étant un λ -système, $A \cap (C \setminus B)$ appartient à \mathcal{D} . Donc $C \setminus B \in \mathcal{D}_1$.

Soient $(B_n)_{\mathbf{N}} \subset \mathcal{D}_1$ une suite croissante et $A \in \mathcal{C}$. On a $A \cap \cup_{\mathbf{N}} B_n = \cup_{\mathbf{N}} A \cap B_n$. $(A \cap B_n)_{\mathbf{N}}$ est une suite croissante du λ -système \mathcal{D} . On en déduit que $\cup_{\mathbf{N}} A \cap B_n \in \mathcal{D}$ et donc que $\cup_{\mathbf{N}} B_n \in \mathcal{D}_1$.

\mathcal{D}_1 est un λ -système qui contient \mathcal{C} . D'où $\mathcal{D} \subset \mathcal{D}_1$ puis $\mathcal{D}_1 = \mathcal{D}$. Ceci signifie que

$$\forall B \in \mathcal{D}, \forall A \in \mathcal{C}, \quad B \cap A \in \mathcal{D}.$$

Considérons à présent

$$\mathcal{D}_2 = \{A \in \mathcal{D}, \quad \forall B \in \mathcal{D}, B \cap A \in \mathcal{D}\}.$$

On montre de la même façon que \mathcal{D}_2 est un λ -système contenu dans \mathcal{D} . D'autre part, d'après l'étape précédente, $\mathcal{C} \subset \mathcal{D}_2$ et par suite $\mathcal{D} \subset \mathcal{D}_2$ et $\mathcal{D} = \mathcal{D}_2$. On a donc

$$\forall A \in \mathcal{D}, \forall B \in \mathcal{D}, \quad A \cap B \in \mathcal{D},$$

ce qui termine la preuve de cette proposition. □

Une des applications de ce résultat est le théorème d'égalité de deux probabilités.

Théorème 7. *Soient \mathcal{C} un π -système, μ et ν deux probabilités sur $\sigma(\mathcal{C})$ telles que $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{C}$. Alors μ et ν sont égales sur $\sigma(\mathcal{C})$.*

Démonstration. Considérons

$$\mathcal{D} = \{A \in \sigma(\mathcal{C}), \quad \mu(A) = \nu(A)\}.$$

$E \in \mathcal{D}$ par hypothèse. Si $A \subset B$ sont deux éléments de \mathcal{D} , on a, les mesures μ et ν étant finies,

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A),$$

ce qui montre que $B \setminus A \in \mathcal{D}$. Si $(A_n)_{\mathbf{N}} \subset \mathcal{D}$ est une suite croissante, on a

$$\mu(\cup_{\mathbf{N}} A_n) = \lim_{n \rightarrow +\infty} \mu(A_n) = \lim_{n \rightarrow +\infty} \nu(A_n) = \nu(\cup_{\mathbf{N}} A_n);$$

\mathcal{D} est donc stable par union dénombrable croissante et par suite un λ -système. Or par hypothèse, $\mathcal{C} \subset \mathcal{D}$ et le Théorème 6 implique que $\sigma(\mathcal{C}) \subset \mathcal{D}$ ce qui donne le résultat. □

Remarque. Lorsque E est un espace topologique, on utilise souvent le π -système des fermés ou celui des ouverts qui engendre la tribu borélienne. Dans \mathbf{R}^d , le π -système des compacts engendre la tribu borélienne.

Pour les variables aléatoires, ce théorème s'applique de la manière suivante : soient (E, \mathcal{E}) un espace mesurable, on suppose que $\mathcal{E} = \sigma(\mathcal{C})$ où \mathcal{C} est un π -système et μ une probabilité sur (E, \mathcal{E}) , X et Y deux v.a. à valeurs dans E . Alors,

1. X a pour loi μ si et seulement si $\mathbb{P}(X \in B) = \mu(B)$ pour tout $B \in \mathcal{C}$;
2. $X \stackrel{(d)}{=} Y$ si et seulement si $\mathbb{P}(X \in B) = \mathbb{P}(Y \in B)$ pour tout $B \in \mathcal{C}$.

Un des exemples les plus utilisés est celui de $\mathcal{B}(\mathbf{R})$ qui est engendrée par les intervalles du type $] - \infty, t]$, $t \in \mathbf{R}$. Il conduit à la définition suivante.

Définition. Soit μ une probabilité sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. On appelle fonction de répartition de μ la fonction réelle définie par

$$\forall t \in \mathbf{R}, \quad F_\mu(t) = \mu(] - \infty, t]).$$

Si X est une v.a. réelle, la fonction de répartition de X , F_X , est celle de sa loi \mathbb{P}_X soit

$$\forall t \in \mathbf{R}, \quad F_X(t) = \mathbb{P}_X(] - \infty, t]) = \mathbb{P}(X \leq t).$$

On montre sans peine que F_μ est une fonction croissante, continue à droite à valeurs dans $[0, 1]$ vérifiant $\lim_{t \rightarrow -\infty} F_\mu(t) = 0$ et $\lim_{t \rightarrow +\infty} F_\mu(t) = 1$. On a $\mu(]s, t]) = F_\mu(t) - F_\mu(s)$ et $\mu(\{t\}) = F_\mu(t) - F_\mu(t^-)$.

De plus, l'ensemble des intervalles du type $] - \infty, t]$ est un π -système qui engendre la tribu borélienne, ce qui conduit au résultat suivant :

Proposition 8. *Deux probabilités (resp. v.a.) sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ sont égales (resp. égales en loi) si et seulement si elles ont même fonction de répartition.*

Une v.a.r. X est de loi μ si et seulement si $F_X = F_\mu$.

Pour simplifier quelque peu la situation, l'étude d'une probabilité sur \mathbf{R} se ramène à l'étude d'une fonction réelle.

En fait toute fonction F de \mathbf{R} dans $[0, 1]$, croissante, continue à droite et telle que $F(+\infty) = 1$, $F(-\infty) = 0$ est la fonction de répartition d'une unique mesure de probabilité μ sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ soit $F(t) = \mu(] - \infty, t])$. En d'autres termes, il existe une v.a. réelle X telle que $F_X = F$. L'exercice suivant démontre ce résultat à l'aide de la mesure de Lebesgue. Pour le résultat général qui comprend la construction de la mesure de Lebesgue, je vous renvoie à l'annexe.

Exercice. On se place sur $]0, 1[$ muni de la tribu borélienne et de la mesure de Lebesgue. On note, pour $\omega \in]0, 1[$, $X(\omega) = \inf\{t \in \mathbf{R} : F(t) \geq \omega\}$. Montrer que X est croissante, continue à gauche et que sa fonction de répartition est F .

Pour déterminer la loi d'une v.a. réelle, il suffit donc de déterminer sa fonction de répartition.

3.2. Approche fonctionnelle.

Au paragraphe précédent, l'égalité de μ et ν s'obtenait en montrant que $\mu(B) = \nu(B)$ pour une certaine classe de parties de \mathcal{E} . Ici, nous cherchons à obtenir l'égalité de deux mesures en calculant des intégrales.

1. Rappels. Soit E un espace métrique muni de sa tribu borélienne. On note $\mathcal{C}_b(E)$ l'ensemble des fonctions réelles continues et bornées sur E , $\mathcal{C}_0(E)$ celui des fonctions continues tendant vers 0 à l'infini – pour tout $\varepsilon > 0$, il existe un compact K tel que $|f(x)| \leq \varepsilon$ si $x \in K^c$ – et $\mathcal{C}_c(E)$ celui des fonctions continues à support compact. On note pour $f \in \mathcal{C}_b(E)$, $\|f\|_\infty = \sup_{x \in E} |f(x)|$.

Proposition 9. Soient μ et ν deux probabilités sur E . $\mu = \nu$ si et seulement si, pour toute $f \in \mathcal{C}_b(E)$, Lipschitzienne et positive,

$$\int_E f(x) \mu(dx) = \int_E f(x) \nu(dx).$$

Démonstration. Si $A \subset E$, la fonction $x \mapsto d(x, A) = \inf\{|x - u|, u \in A\}$ est 1-Lipschitzienne. Soit F un fermé de E . Considérons la fonction $f(x) = (1 + d(x, F))^{-1}$. Alors, $0 \leq f(x) \leq 1$ et $f(x) = 1$ si et seulement si $x \in F$ puisque F est fermé. Par conséquent, f^n , qui est n -Lipschitzienne et positive, décroît simplement vers $\mathbf{1}_F$. Par convergence dominée,

$$\mu(F) = \lim_{n \rightarrow +\infty} \int_E f^n(x) \mu(dx) = \lim_{n \rightarrow +\infty} \int_E f^n(x) \nu(dx) = \nu(F).$$

Or le π -système des fermés engendre la tribu borélienne. Donc, d'après le Théorème 7, $\mu = \nu$. L'implication inverse est immédiate. \square

Plaçons nous à présent dans \mathbf{R}^d .

Proposition 10. Soit $\mathcal{H} \subset \mathcal{C}_b(\mathbf{R}^d)$ tel que $\mathcal{C}_c(\mathbf{R}^d) \subset \overline{\mathcal{H}}$ (l'adhérence pour $\|\cdot\|_\infty$). $\mu = \nu$ si et seulement si

$$\forall f \in \mathcal{H}, \quad \int_{\mathbf{R}^d} f(x) \mu(dx) = \int_{\mathbf{R}^d} f(x) \nu(dx).$$

Démonstration. Montrons tout d'abord que si

$$\int_{\mathbf{R}^d} f(x) \mu(dx) = \int_{\mathbf{R}^d} f(x) \nu(dx) \tag{2}$$

pour toute $f \in \mathcal{C}_c$ et positive alors $\mu = \nu$. Nous donnons ici une démonstration valable dans un contexte plus général. Soit K un compact de \mathbf{R}^d et G un ouvert tel que $K \subset G$ et \bar{G} est compact. Posons $f(x) = d(x, G^c) / (d(x, K) + d(x, G^c))$. f est continue sur \mathbf{R}^d et de plus $0 \leq f(x) \leq 1$. f est à support compact puisque $f(x) = 0$ si et seulement si $x \in G^c$ et de plus $f(x) = 1$ si et seulement si $x \in K$. Par suite, f^n décroît vers $\mathbf{1}_K$ et donc, par convergence dominée

$$\mu(K) = \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f^n(x) \mu(dx) = \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f^n(x) \nu(dx) = \nu(K).$$

Le résultat se déduit directement en considérant le π -système des compacts qui engendre la tribu borélienne de \mathbf{R}^d .

Supposons à présent l'égalité (2) valable seulement pour $h \in \mathcal{H}$. Soit $f \in \mathcal{C}_c$. Par hypothèse, si $\varepsilon > 0$, il existe $h \in \mathcal{H}$ telle que $\|f - h\|_\infty \leq \varepsilon$. Il vient alors

$$\left| \int_{\mathbf{R}^d} f(x) \mu(dx) - \int_{\mathbf{R}^d} h(x) \mu(dx) \right| \leq \int_{\mathbf{R}^d} |f(x) - h(x)| \mu(dx) \leq \varepsilon,$$

et par suite, l'inégalité précédente étant aussi valable pour ν ,

$$\left| \int_{\mathbf{R}^d} f(x) \mu(dx) - \int_{\mathbf{R}^d} f(x) \nu(dx) \right| \leq 2\varepsilon + \left| \int_{\mathbf{R}^d} h(x) \mu(dx) - \int_{\mathbf{R}^d} h(x) \nu(dx) \right| = 2\varepsilon ;$$

ce qui donne le résultat.

La réciproque est immédiate. \square

Remarque. La proposition précédente est valable si E est un espace localement compact à base dénombrable.

Rappelons que toute fonction continue à support compact est limite uniforme d'une suite de fonctions \mathcal{C}^∞ à support compact ; prenant $\mathcal{H} = \mathcal{C}_c^\infty$, on obtient le

Corollaire 11. *Soient μ et ν deux probabilités sur \mathbf{R}^d .*

$$\mu = \nu \iff \forall f \in \mathcal{C}_c^\infty, \quad \int_{\mathbf{R}^d} f(x) \mu(dx) = \int_{\mathbf{R}^d} f(x) \nu(dx).$$

2. Fonctions caractéristiques. Passons à un résultat utilisant la transformation de Fourier.

Définition. Soit μ une probabilité (ou une mesure finie) sur \mathbf{R}^d . On appelle transformée de Fourier de μ la fonction complexe définie par

$$\forall t \in \mathbf{R}^d, \quad \widehat{\mu}(t) = \int_{\mathbf{R}^d} e^{it \cdot x} \mu(dx),$$

où $t \cdot x = \langle t, x \rangle$ désigne le produit scalaire de t et x .

Si X est une variable aléatoire dans \mathbf{R}^d , on appelle fonction caractéristique de X , notée φ_X , la transformée de Fourier de \mathbb{P}_X soit

$$\varphi_X(t) = \widehat{\mathbb{P}_X}(t) = \mathbb{E} \left[e^{it \cdot X} \right].$$

L'inégalité élémentaire $\sup_{|t-s| \leq \eta} |e^{it \cdot x} - e^{is \cdot x}| \leq \min(2, \eta|x|)$ permet d'utiliser le théorème de convergence dominée pour montrer que $\widehat{\mu}$ est une fonction uniformément continue sur \mathbf{R}^d ; elle est aussi bornée par $\mu(\mathbf{R}^d)$.

Nous allons voir que la fonction caractéristique, comme son nom l'indique, caractérise la loi d'un vecteur aléatoire. Notons h la fonction définie par

$$\forall t \in \mathbf{R}^*, \quad h(t) = \frac{1}{\pi} \frac{1 - \cos t}{t^2}, \quad h(0) = \frac{1}{2\pi}.$$

Proposition 12. *Soient μ une probabilité sur \mathbf{R}^d et $f \in \mathcal{C}_b(\mathbf{R}^d)$. Alors,*

$$\left| \int_{\mathbf{R}^d} f(x) \mu(dx) - \int_{\mathbf{R}^d} f(x) \int_{\mathbf{R}^d} e^{-it \cdot x} H(\varepsilon t) \widehat{\mu}(t) dt dx \right| \leq \omega_f(\varepsilon \sqrt{d}),$$

où $H(t) = h(t_1) \dots h(t_d)$ et $\omega_f(\eta) = \sup_{|x-y| \leq \eta} |f(x) - f(y)|$.

Démonstration. Notons $K(x) = k(x_1)k(x_2) \dots k(x_d)$ avec, pour $u \in \mathbf{R}$, $k(u) = (1 - |u|)\mathbf{1}_{|u| \leq 1}$ et, pour $\varepsilon > 0$, $K_\varepsilon(x) = \varepsilon^{-d} K(x/\varepsilon)$. K est positive, intégrable et d'intégrale 1. On a, pour f continue bornée,

$$\left| \int_{\mathbf{R}^d} (f \star K_\varepsilon)(x) \mu(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq \|(f \star K_\varepsilon) - f\|_\infty.$$

D'autre part, pour $x \in \mathbf{R}^d$,

$$|(f \star K_\varepsilon)(x) - f(x)| \leq \int_{\mathbf{R}^d} |f(x - \varepsilon z) - f(x)| K(z) dz \leq \omega_f(\varepsilon \sqrt{d}).$$

Admettons pour l'instant que

$$\forall x \in \mathbf{R}, \quad \int_{\mathbf{R}} e^{itx} h(t) dt = k(x). \quad (3)$$

On obtient, via le théorème de Fubini,

$$\int_{\mathbf{R}^d} (f \star K_\varepsilon)(x) \mu(dx) = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} f(y) K_\varepsilon(x-y) dy \mu(dx) = \int_{\mathbf{R}^d} f(y) \int_{\mathbf{R}^d} K_\varepsilon(x-y) \mu(dx) dy.$$

Or, la formule (3) entraîne que

$$\forall x \in \mathbf{R}^d, \quad K_\varepsilon(x) = \frac{1}{\varepsilon^d} \int_{\mathbf{R}^d} e^{it \cdot \frac{x}{\varepsilon}} H(t) dt = \int_{\mathbf{R}^d} e^{it \cdot x} H(\varepsilon t) dt.$$

Le théorème de Fubini donne alors l'égalité

$$\int_{\mathbf{R}^d} K_\varepsilon(x-y) \mu(dx) = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} e^{it \cdot (x-y)} H(\varepsilon t) dt \mu(dx) = \int_{\mathbf{R}^d} e^{-it \cdot y} H(\varepsilon t) \widehat{\mu}(t) dt, \quad (4)$$

et par suite

$$\int_{\mathbf{R}^d} (f \star K_\varepsilon)(x) \mu(dx) = \int_{\mathbf{R}^d} f(y) \int_{\mathbf{R}^d} e^{-it \cdot y} H(\varepsilon t) \widehat{\mu}(t) dt dy.$$

Il reste à démontrer la formule (3). Soit $x \in \mathbf{R}$. Considérons la fonction ψ définie par

$$\forall u \geq 0, \quad \psi(u) = \int_{\mathbf{R}} h(t) e^{itx} e^{-u|t|} dt.$$

La fonction h étant intégrable sur \mathbf{R} , on montre facilement en appliquant les résultats sur les intégrales à paramètres que ψ est continue sur \mathbf{R}_+ , \mathcal{C}^∞ sur $]0, +\infty[$ et vérifie $\lim_{u \rightarrow +\infty} \psi(u) = 0$. On a d'autre part

$$\forall u > 0, \quad \psi''(u) = \frac{1}{\pi} \int_{\mathbf{R}} (1 - \cos t) e^{itx} e^{-u|t|} dt = \frac{1}{2\pi} \int_{\mathbf{R}} (2e^{itx} - e^{i(1+x)t} - e^{i(1-x)t}) e^{-u|t|} dt,$$

soit encore

$$\forall u > 0, \quad \psi''(u) = \frac{1}{\pi} \left(\frac{2u}{u^2 + x^2} - \frac{u}{u^2 + (1-x)^2} - \frac{u}{u^2 + (1+x)^2} \right);$$

par suite, on obtient

$$\forall u > 0, \quad \psi'(u) = \frac{1}{2\pi} \left\{ 2 \ln(u^2 + x^2) - \ln(u^2 + (1-x)^2) - \ln(u^2 + (1+x)^2) \right\} + c.$$

Une primitive de $\ln(u^2 + a^2)$ est $u \ln(u^2 + a^2) - 2u + 2|a| \arctan(u/|a|)$: il suffit de faire une intégration par parties. On a alors – pour $x = 0$ et $|x| = 1$ le terme en \arctan disparaît – pour tout $u \geq 0$,

$$\begin{aligned} \psi(u) &= \frac{u}{2\pi} \left\{ 2 \ln(u^2 + x^2) - \ln(u^2 + (1-x)^2) - \ln(u^2 + (1+x)^2) \right\} + cu + d \\ &\quad + \frac{1}{\pi} \left\{ 2|x| \arctan(u/|x|) - |1-x| \arctan(u/|1-x|) - |1+x| \arctan(u/|1+x|) \right\}. \end{aligned}$$

On a

$$\lim_{u \rightarrow +\infty} u \left\{ 2 \ln(u^2 + x^2) - \ln(u^2 + (1-x)^2) - \ln(u^2 + (1+x)^2) \right\} = 0$$

et donc, comme $\lim_{u \rightarrow +\infty} \psi(u) = 0$, $c = 0$ et

$$\psi(0) = d = \frac{1}{2} (-2|x| + |1-x| + |1+x|) = (1-|x|) \mathbf{1}_{|x| \leq 1}.$$

□

Corollaire 13. Deux probabilités sur \mathbf{R}^d , μ et ν , sont égales si et seulement si $\widehat{\mu}(t) = \widehat{\nu}(t)$ pour tout $t \in \mathbf{R}^d$.

Démonstration. Si $f \in C_0(\mathbf{R}^d)$, f est uniformément continue et donc,

$$\int_{\mathbf{R}^d} f(x) \mu(dx) = \lim_{\varepsilon \rightarrow 0^+} \int_{\mathbf{R}^d} f(x) \int_{\mathbf{R}^d} e^{-it \cdot x} H(\varepsilon t) \widehat{\mu}(t) dt dx,$$

et de même pour ν . Si donc $\widehat{\mu} = \widehat{\nu}$,

$$\forall f \in C_0(\mathbf{R}^d), \quad \int_{\mathbf{R}^d} f(x) \mu(dx) = \int_{\mathbf{R}^d} f(x) \nu(dx).$$

Il reste à appliquer la Proposition 10. □

Corollaire 14. Si $\widehat{\mu}$ (resp. φ_X) est intégrable alors μ (resp. X) possède une densité par rapport à λ_d donnée par

$$p(x) = (2\pi)^{-d} \int_{\mathbf{R}^d} e^{-it \cdot x} \widehat{\mu}(t) dt.$$

Démonstration. Si $\widehat{\mu}$ est intégrable, le théorème de convergence dominée donne

$$\lim_{\varepsilon \rightarrow 0^+} \int_{\mathbf{R}^d} e^{-it \cdot x} H(\varepsilon t) \widehat{\mu}(t) dt = p(x).$$

Remarquons que, d'après la formule (4), p est positive sur \mathbf{R}^d .

Si $f \in C_c$, appliquant encore le théorème de convergence dominée,

$$\int_{\mathbf{R}^d} f(x) \mu(dx) = \lim_{\varepsilon \rightarrow 0^+} \int_{\mathbf{R}^d} f(x) \int_{\mathbf{R}^d} e^{-it \cdot x} H(\varepsilon t) \widehat{\mu}(t) dt dx = \int_{\mathbf{R}^d} f(x) p(x) dx.$$

□

3.3. Conséquences pour les variables aléatoires.

Soient X une v.a. à valeurs dans \mathbf{R}^d et μ une probabilité sur \mathbf{R}^d .

Alors X a pour loi μ dès que l'une des conditions suivantes est remplie :

- (i) $\mathbb{P}(X \in C) = \mu(C)$ pour tout C appartenant à un π -système engendrant la tribu borélienne ;
- (ii) $\mathbb{E}[f(X)] = \int_{\mathbf{R}^d} f(x) \mu(dx)$, pour toute fonction $f \in C_c^\infty(\mathbf{R}^d)$ positive ;
- (iii) $\varphi_X(t) = \widehat{\mu}(t)$, pour tout $t \in \mathbf{R}^d$.

De la même façon, si X et Y sont deux vecteurs aléatoires, ils sont égaux en loi si

- (i) $\mathbb{P}(X \in C) = \mathbb{P}(Y \in C)$ pour tout C appartenant à un π -système engendrant la tribu borélienne ;
- (ii) $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$, pour toute fonction $f \in C_c^\infty(\mathbf{R}^d)$ positive ;
- (iii) $\varphi_X(t) = \varphi_Y(t)$, pour tout $t \in \mathbf{R}^d$.

3.4. Cas particuliers.

Soit X une v.a. de loi μ .

1. Variable entière. Lorsque X est une variable entière ou plus généralement à valeurs dans $\overline{\mathbf{N}}$, on utilise la *série génératrice* pour caractériser μ . Celle-ci est définie, pour $s \in [0, 1]$, par

$$G_\mu(s) = \int_{[0, +\infty[} s^x \mu(dx) = \sum_{n \geq 0} s^n \mu(\{n\}).$$

La série génératrice de X est celle de sa loi soit

$$G_X(s) = \mathbb{E} \left[s^X \mathbf{1}_{X < +\infty} \right] = \sum_{n \geq 0} s^n \mathbb{P}(X = n).$$

G_X caractérise la loi de X puisque

$$\forall n \in \mathbf{N}, \quad G_X^{(n)}(0) = n! \mathbb{P}(X = n), \quad \mathbb{P}(X = +\infty) = 1 - G_X(1).$$

La seule chose à remarquer est que G_X est normalement convergente sur $[-1, 1]$ et d'appliquer les résultats sur les séries entières.

2. Variable positive. Pour une v.a. positive c'est à dire à valeurs dans $\overline{\mathbf{R}}_+$, on utilise assez fréquemment la transformée de Laplace. La transformée de Laplace d'une probabilité μ sur $\overline{\mathbf{R}}_+$ est la fonction définie par

$$\forall s \geq 0, \quad \Lambda_\mu(s) = \int_{[0, +\infty[} e^{-sx} \mu(dx);$$

celle de X est bien évidemment

$$\Lambda_X(s) = \mathbb{E} \left[e^{-sX} \mathbf{1}_{X < +\infty} \right].$$

Le fait que la transformée de Laplace caractérise la loi d'une v.a. résulte du théorème de Stone-Weierstrass. Remarquons d'autre part que Λ_X est continue sur \mathbf{R}_+ et vérifie

$$\Lambda_X(0) = \mathbb{P}(X < +\infty), \quad \lim_{s \rightarrow +\infty} \Lambda_X(s) = \mathbb{P}(X = 0).$$

4. Moments d'une v.a..

Dans ce paragraphe, nous donnons quelques paramètres de la loi d'une v.a. vectorielle qui précise la taille de la v.a. sans être caractéristique de sa loi.

4.1. Définitions.

Soit X est v.a. positive. On a, pour $p > 0$,

$$\mathbb{E}[X^p] = \int_0^{+\infty} pt^{p-1} \mathbb{P}(X > t) dt.$$

En effet, $X = \int_0^{+\infty} \mathbf{1}_{t < X} dt$ et donc, en prenant l'espérance, le théorème de Fubini donne

$$\mathbb{E}[X] = \mathbb{E} \left[\int_0^{+\infty} \mathbf{1}_{t < X} dt \right] = \int_0^{+\infty} \mathbb{E}[\mathbf{1}_{t < X}] dt = \int_0^{+\infty} \mathbb{P}(X > t) dt.$$

Par suite, le changement de variable $t = s^p$, conduit à

$$\mathbb{E}[X^p] = \int_0^{+\infty} \mathbb{P}(X^p > t) dt = \int_0^{+\infty} ps^{p-1} \mathbb{P}(X^p > s^p) ds,$$

ce qui donne la formule puisque $\{X^p > s^p\} = \{X > s\}$.

La fonction $t \mapsto \mathbb{P}(X > t)$ est décroissante et tend vers $\mathbb{P}(X = +\infty)$ lorsque $t \rightarrow +\infty$. Si X est finie p.s., $\mathbb{P}(X > t)$ décroît vers 0. Lorsque $\mathbb{E}[X^p] < +\infty$, on peut préciser cette décroissance vers 0 puisqu'on sait alors que $t^{p-1} \mathbb{P}(X > t)$ est intégrable. Cela donne une idée de la taille de X .

Remarquons également que $X \geq 0$ est intégrable si et seulement si la série de terme général $\mathbb{P}(X > n)$ est convergente. En effet, comme $\mathbb{P}(X > t)$ est décroissante,

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

Définition. Soit X une v.a. vectorielle ou numérique et $p > 0$. On dit que X possède un moment d'ordre p lorsque $\mathbb{E}[|X|^p] < +\infty$. On dit aussi que X est de puissance p^e -intégrable.

Lorsque X est un vecteur aléatoire, X possède un moment d'ordre p si et seulement s'il en est de même de chacune de ses composantes.

Remarquons que si X possède un moment d'ordre p alors X possède un moment d'ordre q pour tout $q \leq p$ puisque

$$|X|^q \leq |X|^q \mathbf{1}_{|X| \leq 1} + |X|^q \mathbf{1}_{|X| > 1} \leq 1 + |X|^p.$$

Ceci peut d'ailleurs se retrouver à l'aide de l'inégalité de Hölder : si $p \in]1, +\infty[$ et $p^{-1} + q^{-1} = 1$,

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

Rappelons aussi deux inégalités très souvent employées :

Inégalité de Minkowski : si $p \geq 1$, $\mathbb{E}[|X+Y|^p]^{\frac{1}{p}} \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}}$;

Inégalité de Jensen : soit $g :]a, b[\rightarrow \mathbf{R}$ une fonction convexe et X une v.a. à valeurs dans $]a, b[$, $-\infty \leq a < b \leq +\infty$. Si X et $g(X)$ sont intégrables, $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$.

Soit X une variable aléatoire réelle. Lorsque X est intégrable, on appelle $\mathbb{E}[X]$ la moyenne de X . Si X est de carré intégrable, le moment centré d'ordre deux s'appelle la variance de X ; on le note $\mathbb{V}(X)$ soit

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Il mesure la dispersion de X autour de sa moyenne.

Exercice. Soit X une v.a.r. de carré intégrable. Montrer que $c \mapsto \mathbb{E}[(X - c)^2]$ possède un minimum atteint pour $c = \mathbb{E}[X]$.

Si X et Y sont deux v.a. réelles de carré intégrable, on appelle covariance de X et Y le réel

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

On établit facilement la formule

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y).$$

D'autre part, $(X, Y) \mapsto \text{Cov}(X, Y)$ est une forme bilinéaire symétrique et

$$|\text{Cov}(X, Y)|^2 \leq \mathbb{V}(X)\mathbb{V}(Y).$$

Lemme 15 (Inégalité de Markov). *Soit X une v.a. positive. Pour tout $a > 0$,*

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

Démonstration. Puisque X est positive, $X \geq X\mathbf{1}_{X \geq a} \geq a\mathbf{1}_{X \geq a}$. Par suite, $\mathbb{E}[X] \geq a\mathbb{P}(X \geq a)$. \square

Corollaire 16 (Inégalité de Tchebycheff). *Soit X de carré intégrable. Pour tout $\varepsilon > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Ces notions s'étendent au cas de vecteurs aléatoires. Si X est une v.a. à valeurs dans \mathbf{R}^d intégrable i.e. $|X|$ est intégrable on appelle moyenne de X le vecteur (colonne) de \mathbf{R}^d , $(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$.

Si X est de carré intégrable i.e. $|X|^2$ est intégrable, l'application S_X définie par

$$S_X(u) = \mathbb{V}(u \cdot X), \quad u \in \mathbf{R}^d,$$

est une forme quadratique positive que l'on appelle variance du vecteur X . Sa matrice Γ_X dans la base canonique s'appelle la matrice de covariance de X ; $\Gamma_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d}$. On a

$$\Gamma_X = \mathbb{E} \left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t \right].$$

4.2. Moments et fonction caractéristique.

Nous allons voir que la régularité de la fonction caractéristique est liée à l'existence de moments.

Proposition 17. *Soit X une v.a. à valeurs dans \mathbf{R}^d . φ_X est uniformément continue.*

Si X a un moment d'ordre n , alors φ_X est de classe \mathcal{C}^n et, pour $k = (k_1, \dots, k_d)$ tel que $|k|_1 \leq n$,

$$\frac{\partial^{k_1}}{\partial x_1^{k_1}} \dots \frac{\partial^{k_d}}{\partial x_d^{k_d}} \varphi_X(t) = i^{|k|_1} \mathbb{E} \left[X_1^{k_1} \dots X_d^{k_d} e^{it \cdot X} \right].$$

Démonstration. La continuité uniforme, comme déjà dit vient de l'inégalité

$$\sup_{|t-s| \leq \eta} \left| e^{it \cdot X} - e^{is \cdot X} \right| \leq \min(2, \eta|X|).$$

Plaçons nous dans \mathbf{R} et supposons X intégrable. $t \mapsto e^{itX}$ est de classe \mathcal{C}^1 (pour tout $\omega \in \Omega$). On a, $\sup_{t \in \mathbf{R}} |iX e^{itX}| \leq |X|$ qui est intégrable par hypothèse. Les résultats sur les intégrales à paramètres montrent que φ_X est \mathcal{C}^1 et que $\varphi'_X(t) = i \mathbb{E} [X e^{itX}]$. On termine la preuve par récurrence. \square

Remarque. L'existence d'un moment d'ordre 1 d'une v.a.r. se voit sur sa fonction caractéristique puisque

$$\mathbb{E}[|X|] = \frac{1}{\pi} \int_0^{+\infty} \frac{1 - \text{Re}(\varphi_X(t))}{t^2} dt$$

En effet, le changement de variables $u = t|X|$, donne

$$\int_0^{+\infty} \frac{1 - \cos t|X|}{t^2} dt = |X| \int_0^{+\infty} \frac{1 - \cos u}{u^2} du.$$

Il reste à prendre l'espérance et utiliser Fubini.

Bibliographie.

- [Mal] F. Malrieu, *Cours d'intégration et probabilités*, URL : <http://name.math.univ-rennes1.fr/florent.malrieu/>.
- [Rud87] Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.

Chapitre II. La notion d'Indépendance

Nous abordons dans ce chapitre une notion fondamentale du calcul des probabilités : l'indépendance. L'objectif est de caractériser l'indépendance de tribus ou de variables aléatoires et d'étudier quelques propriétés et exemples liés à cette notion. La construction de suites de variables indépendantes est discutée en annexe.

1. Tribus indépendantes.

Dans toute la suite, on se place sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$; toutes les tribus qui interviendront seront des sous-tribus de \mathcal{F} . Nous utiliserons les conventions suivantes :

$$\bigcap_{i \in \emptyset} A_i = \Omega, \quad \bigcup_{i \in \emptyset} A_i = \emptyset, \quad \prod_{i \in \emptyset} \mathbb{P}(A_i) = 1.$$

D'autre part, nous noterons « $i \leq n$ » pour « $1 \leq i \leq n$ » ou « $0 \leq i \leq n$ » suivant le contexte.

Définition. Des tribus, $\mathcal{F}_1, \dots, \mathcal{F}_n$, sont indépendantes si

$$\forall (A_1, \dots, A_n) \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n, \quad \mathbb{P}\left(\bigcap_{i \leq n} A_i\right) = \prod_{i \leq n} \mathbb{P}(A_i).$$

Une suite de tribus, $(\mathcal{F}_n)_{n \in \mathbf{N}^*}$, est indépendante si, pour tout $n \in \mathbf{N}^*$, les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes.

Enfin, des événements $(A_n)_{n \in \mathbf{N}^*}$ sont indépendants lorsque les tribus $\sigma(A_n)$ le sont; on rappelle que $\sigma(A_n) = \{\emptyset, \Omega, A_n, A_n^c\}$.

Remarque. Insistons sur le fait que l'indépendance d'une suite est définie à partir de l'indépendance d'un nombre fini de tribus. Plus généralement, une famille quelconque de tribus – non nécessairement dénombrable – est indépendante si toute sous-famille finie est indépendante.

Exercice. Montrer que la tribu triviale $\{\emptyset, \Omega\}$ est indépendante de toute tribu.

On montre facilement – donc faites-le – que deux événements A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Plus généralement, on a le résultat suivant :

Proposition 1. n événements, A_1, \dots, A_n , sont indépendants si et seulement si

$$\forall I \subset \{1, \dots, n\}, \quad \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Démonstration. La condition est bien évidemment nécessaire; montrons qu'elle est aussi suffisante. On doit montrer que, pour tout $B_i \in \sigma(A_i)$, $i = 1, \dots, n$,

$$\mathbb{P}\left(\bigcap_{i \leq n} B_i\right) = \prod_{i \leq n} \mathbb{P}(B_i).$$

Or $B_i = \emptyset, \Omega, A_i$ ou A_i^c . Si l'un des $B_i = \emptyset$ la formule est correcte ($0=0$). On suppose donc qu'aucun des B_i n'est égal à l'ensemble vide. Lorsque l'un des $B_i = \Omega$, il ne compte pas : on

est donc ramené au cas d'une intersection de A_i et de A_j^c , le nombre total d'ensembles étant inférieur ou égal à n . On doit donc établir que, pour toute partie I de $\{1, \dots, n\}$ et toute partie J de $\{1, \dots, n\} \setminus I$,

$$\mathbb{P}\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} A_j^c\right) = \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J} \mathbb{P}(A_j^c).$$

La démonstration se fait par récurrence sur le nombre d'éléments de J noté $\#J$. Si $\#J = 0$ c'est l'hypothèse. Supposons la formule vraie lorsque $\#J = r - 1$ et montrons-la pour $\#J = r$. Soit donc $I \subset \{1, \dots, n\}$, et $J \subset \{1, \dots, n\} \setminus I$ tel que $\#J = r$. Notons j' le plus petit élément de J et $J' = J \setminus \{j'\}$ de sorte que $\#J' = r - 1$. On a, notant $B = \bigcap_{i \in I} A_i \cap \bigcap_{j \in J'} A_j^c$,

$$\mathbb{P}(B) = \mathbb{P}(A_{j'} \cap B) + \mathbb{P}(A_{j'}^c \cap B).$$

Mais puisque $\#J' = r - 1$ on a

$$\mathbb{P}(B) = \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J'} \mathbb{P}(A_j^c), \quad \mathbb{P}(A_{j'} \cap B) = \mathbb{P}(A_{j'}) \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J'} \mathbb{P}(A_j^c),$$

dont on déduit que

$$\mathbb{P}(A_{j'}^c \cap B) = [1 - \mathbb{P}(A_{j'})] \mathbb{P}(B) = \mathbb{P}(A_{j'}^c) \mathbb{P}(B) = \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J} \mathbb{P}(A_j^c).$$

Ceci termine la démonstration puisque $A_{j'}^c \cap B = \bigcap_{i \in I} A_i \cap \bigcap_{j \in J} A_j^c$. \square

Exercice. Que signifie que A , B et C sont indépendants ?

Une conséquence immédiate de la définition est que si $(\mathcal{F}_n)_{\mathbb{N}}$ est indépendante et si, pour tout $n \in \mathbb{N}$, \mathcal{A}_n est une sous-tribu de \mathcal{F}_n alors $(\mathcal{A}_n)_{\mathbb{N}}$ est indépendante.

Voici un résultat qui permet de montrer plus facilement que des tribus sont indépendantes.

Proposition 2. *Soit, pour $i = 1, \dots, n$, \mathcal{F}_i une tribu engendrée par \mathcal{C}_i où \mathcal{C}_i est un π -système tel que $\Omega \in \mathcal{C}_i$.*

Les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes si et seulement si

$$\forall (A_1, \dots, A_n) \in \mathcal{C}_1 \times \dots \times \mathcal{C}_n, \quad \mathbb{P}\left(\bigcap_{i \leq n} A_i\right) = \prod_{i \leq n} \mathbb{P}(A_i).$$

Démonstration. Nous devons montrer que, pour tout $(A_1, \dots, A_n) \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n$,

$$\mathbb{P}\left(\bigcap_{i \leq n} A_i\right) = \prod_{i \leq n} \mathbb{P}(A_i). \tag{1}$$

Considérons, pour $r = 0, \dots, n$, la propriété (P_r) suivante : l'égalité (1) est valable pour tous $A_i \in \mathcal{F}_i$, $1 \leq i \leq r$ et tous $A_i \in \mathcal{C}_i$, $r < i \leq n$.

Par hypothèse (P_0) est vraie. Montrons que (P_r) est vraie si (P_{r-1}) l'est. Considérons pour cela \mathcal{D} l'ensemble des $B \in \mathcal{F}_r$ tels que l'égalité (1) soit valable pour tout (A_1, \dots, A_n) avec $A_i \in \mathcal{F}_i$ si $1 \leq i \leq r - 1$, $A_i \in \mathcal{C}_i$ si $r < i \leq n$ et $A_r = B$. Puisque (P_{r-1}) est vraie, \mathcal{D} contient \mathcal{C}_r et donc Ω . Montrons que \mathcal{D} est un λ -système. Si B et C sont deux éléments de \mathcal{D} tels que $B \subset C$, on a, notant $A_- = A_1 \cap \dots \cap A_{r-1}$ et $A_+ = A_{r+1} \cap \dots \cap A_n$,

$$\mathbb{P}(A_- \cap (C \setminus B) \cap A_+) = \mathbb{P}(A_- \cap C \cap A_+) - \mathbb{P}(A_- \cap B \cap A_+);$$

comme B et C sont éléments de \mathcal{D} ,

$$\mathbb{P}(A_- \cap (C \setminus B) \cap A_+) = (\mathbb{P}(C) - \mathbb{P}(B)) \prod_{i \neq r} \mathbb{P}(A_i) = \mathbb{P}(C \setminus B) \prod_{i \neq r} \mathbb{P}(A_i).$$

Donc $C \setminus B$ appartient à \mathcal{D} .

Si $(B_n)_{\mathbf{N}}$ est une suite croissante de \mathcal{D} , notant $B = \cup_{n \geq 0} B_n$, on a

$$\mathbb{P}(A_- \cap B \cap A_+) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_- \cap B_n \cap A_+) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) \prod_{i \neq r} \mathbb{P}(A_i) = \mathbb{P}(B) \prod_{i \neq r} \mathbb{P}(A_i).$$

Par conséquent, $B \in \mathcal{D}$.

\mathcal{D} est donc un λ -système contenant le π -système \mathcal{C}_r . Il contient $\sigma(\mathcal{C}_r) = \mathcal{F}_r$ et donc (P_r) est vraie.

Par récurrence, la propriété (P_n) est vraie ce qui établit cette proposition. \square

Un corollaire de ce résultat est l'indépendance par paquets que j'énonce pour les suites mais qui est valable pour n'importe quelle famille. Commençons par un lemme technique.

Lemme 3. Soient $\{\mathcal{F}_i\}_{i \in I}$ une famille de tribus et

$$\mathcal{C} = \left\{ B = \bigcap_{j \in J} A_j : A_j \in \mathcal{F}_j, J \text{ partie finie de } I \right\}.$$

Alors \mathcal{C} est un π -système qui engendre la tribu $\sigma(\mathcal{F}_i, i \in I)$.

Démonstration. Montrons tout d'abord que \mathcal{C} est un π -système. Soient donc

$$B = \bigcap_{j \in J} A_j, \quad B' = \bigcap_{j \in J'} A'_j,$$

où J et J' sont deux parties finies de I et les $A_j \in \mathcal{F}_j$. On a,

$$B \cap B' = \bigcap_{j \in J \setminus J'} A_j \cap \bigcap_{j \in J' \setminus J} A'_j \cap \bigcap_{j \in J \cap J'} (A_j \cap A'_j) = \bigcap_{j \in J \cup J'} A''_j,$$

avec $A''_j = A_j$ si $j \in J \setminus J'$, $A''_j = A'_j$ si $j \in J' \setminus J$ et $A''_j = A_j \cap A'_j$ si $j \in J \cap J'$. Puisque $J \cup J'$ est une partie finie de I et $A''_j \in \mathcal{F}_j$ pour tout j , $B \cap B' \in \mathcal{C}$.

Montrons à présent que \mathcal{C} engendre la tribu $\sigma(\mathcal{F}_i, i \in I)$. On a $\cup_{i \in I} \mathcal{F}_i \subset \mathcal{C}$ et donc $\sigma(\mathcal{F}_i, i \in I) \subset \sigma(\mathcal{C})$. Réciproquement, si $B \in \mathcal{C}$, $B \in \sigma(\mathcal{F}_i, i \in I)$ puisqu'une tribu est stable par intersection. Par suite, $\sigma(\mathcal{C}) \subset \sigma(\mathcal{F}_i, i \in I)$. \square

Remarque. Un cas particulier du lemme précédent est le point suivant : si $(\mathcal{F}_n)_{\mathbf{N}}$ est une suite croissante de tribus, $\mathcal{C} = \cup_{n \in \mathbf{N}} \mathcal{F}_n$ est un π -système engendrant $\sigma(\mathcal{F}_n, n \in \mathbf{N})$.

Corollaire 4. Soient $(\mathcal{F}_n)_{\mathbf{N}}$ une suite de tribus indépendantes et $(I_k)_{k \in K}$ une partition de \mathbf{N} avec $K = \mathbf{N}^*$ ou $K = \{1, \dots, p\}$, $p \geq 2$. Alors, les tribus $\mathcal{U}_k = \sigma(\mathcal{F}_i, i \in I_k)$, $k \in K$, sont indépendantes.

Avant de donner la démonstration de ce résultat, regardons sa signification sur un exemple simple. Considérons 5 tribus indépendantes – on se ramène au cas d'une suite en prenant $\mathcal{F}_i = \{\emptyset, \Omega\}$ pour $i > 5$ – $\mathcal{F}_1, \dots, \mathcal{F}_5$. Les tribus $\sigma(\mathcal{F}_1, \mathcal{F}_3, \mathcal{F}_5)$ et $\sigma(\mathcal{F}_2, \mathcal{F}_4)$ sont indépendantes; de même pour les tribus $\sigma(\mathcal{F}_1, \mathcal{F}_4)$, \mathcal{F}_5 et $\sigma(\mathcal{F}_2, \mathcal{F}_3)$.

Démonstration. Nous devons montrer que les tribus \mathcal{U}_k , $k = 1, \dots, p$ sont indépendantes pour tout $p \geq 2$ si $K = \mathbf{N}^*$ et pour $p = \#K$ si K est fini. D'après le lemme précédent, pour tout $k \in K$, \mathcal{C}_k l'ensemble des $B_k \in \mathcal{F}$ de la forme

$$B_k = \bigcap_{j \in J_k} A_j, \quad A_j \in \mathcal{F}_j, \quad J_k \text{ partie finie de } I_k$$

est un π -système contenant Ω qui engendre \mathcal{U}_k . Les tribus \mathcal{F}_i étant indépendantes, on a

$$\mathbb{P}(B_k) = \prod_{j \in J_k} \mathbb{P}(A_j).$$

Soient $B_1 \in \mathcal{C}_1, \dots, B_p \in \mathcal{C}_p : B_k = \bigcap_{j \in J_k} A_j$. Les I_k étant deux à deux disjoints, les J_k le sont également. Par suite, on a

$$B_1 \cap \dots \cap B_p = \bigcap_{k \leq p} \bigcap_{j \in J_k} A_j = \bigcap_{j \in \bigcup_{k \leq p} J_k} A_j.$$

et via l'indépendance des \mathcal{F}_i ,

$$\mathbb{P}(B_1 \cap \dots \cap B_p) = \prod_{j \in \bigcup_{k \leq p} J_k} \mathbb{P}(A_j) = \prod_{k \leq p} \prod_{j \in J_k} \mathbb{P}(A_j) = \prod_{k \leq p} \mathbb{P}(B_k).$$

La Proposition 2 implique l'indépendance des \mathcal{U}_k , $k = 1, \dots, p$. □

Passons au second lemme de Borel–Cantelli. Rappelons que, si $\sum_{n \in \mathbf{N}} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup A_n) = 0$. Dans le cas indépendant, nous avons aussi

Lemme 5 (Borel–Cantelli, suite). *Soit $(A_n)_{\mathbf{N}}$ une suite d'événements indépendants. Si la série $\sum_{n \in \mathbf{N}} \mathbb{P}(A_n) = +\infty$, alors $\mathbb{P}(\limsup A_n) = 1$.*

Ceci signifie que pour des événements **indépendants**, on a l'alternative

$$\mathbb{P}(\limsup A_n) = 0 \text{ ou } 1 \text{ suivant que } \sum_{n \geq 0} \mathbb{P}(A_n) \text{ est finie ou pas.}$$

Rappelons d'autre part que $\omega \in \limsup A_n$ si et seulement si ω appartient à une infinité d'événements A_n c'est à dire si et seulement si $N(\omega) = \sum_{n \geq 0} \mathbf{1}_{A_n}(\omega) = +\infty$. Avec cette notation, on a $\mathbb{E}[N] = \sum_{n \geq 0} \mathbb{P}(A_n)$; pour des événements indépendants, la variable aléatoire N est soit finie \mathbb{P} -p.s. soit égale à $+\infty$ \mathbb{P} -p.s. suivant que son espérance est finie ou pas.

Démonstration. On a $\mathbb{P}(\limsup A_n) = 1 - \mathbb{P}(\liminf A_n^c)$; montrons que $\mathbb{P}(\liminf A_n^c) = 0$. Rappelons que $\liminf A_n^c = \bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} A_k^c$. Il s'agit donc de montrer que, pour tout $n \in \mathbf{N}$, $\bigcap_{k \geq n} A_k^c$ est de \mathbb{P} -mesure nulle. Fixons $n \in \mathbf{N}$. Par définition, les tribus $\sigma(A_n)$ sont indépendantes et donc les événements A_n^c sont aussi indépendants. Par conséquent,

$$\mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) = \lim_{p \rightarrow +\infty} \mathbb{P}\left(\bigcap_{n \leq k \leq p} A_k^c\right) = \lim_{p \rightarrow +\infty} \prod_{n \leq k \leq p} \mathbb{P}(A_k^c).$$

D'autre part, notant que $\mathbb{P}(A_k^c) = 1 - \mathbb{P}(A_k)$ et que $1 - x \leq e^{-x}$ pour $x \geq 0$, on a

$$\prod_{n \leq k \leq p} \mathbb{P}(A_k^c) = \prod_{n \leq k \leq p} (1 - \mathbb{P}(A_k)) \leq \prod_{n \leq k \leq p} e^{-\mathbb{P}(A_k)} = \exp\left\{-\sum_{n \leq k \leq p} \mathbb{P}(A_k)\right\}.$$

Pour finir, remarquons que, n étant fixé, $\sum_{n \leq k \leq p} \mathbb{P}(A_k)$ tend vers $+\infty$ si $p \rightarrow +\infty$. □

Exemple. Soit $(A_n)_{\mathbf{N}^*}$ une suite d'événements indépendants tels que $\mathbb{P}(A_n) = p \in]0, 1[$. Considérons τ le premier instant où A_n est réalisé c'est à dire

$$\tau(\omega) = \inf\{n \geq 1 : \omega \in A_n\}, \quad \inf \emptyset = +\infty.$$

Alors τ suit une loi géométrique de paramètre p . En particulier, τ est fini \mathbb{P} -presque sûrement. En effet, si $n \in \mathbf{N}^*$,

$$\mathbb{P}(\tau > n) = \mathbb{P}(A_1^c \cap \dots \cap A_n^c) = \prod_{i \leq n} \mathbb{P}(A_i^c) = (1 - p)^n,$$

et donc $\mathbb{P}(\tau = n) = \mathbb{P}(\tau > n - 1) - \mathbb{P}(\tau > n) = p(1 - p)^{n-1}$.

On peut remarquer également que puisque $\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty$ et les événements sont indépendants, $\mathbb{P}(\limsup A_n) = 1$. Ceci signifie que, \mathbb{P} -p.s., non seulement un des A_n est réalisé mais en réalité une infinité. En d'autres termes, si on définit par récurrence $\tau_0 = 0$ et, pour $r \in \mathbf{N}$,

$$\tau_{r+1}(\omega) = \inf\{n > \tau_r(\omega) : \omega \in A_n\}, \quad \inf \emptyset = +\infty,$$

chaque τ_r est \mathbb{P} -p.s. fini. Nous reprendrons cet exemple un peu plus loin.

Si on considère à présent, pour tout $n \in \mathbf{N}^*$, $B_n = A_{2n-1} \cap A_{2n}^c$ alors pour tout $n \in \mathbf{N}^*$, $\mathbb{P}(B_n) = \mathbb{P}(A_{2n-1} \cap A_{2n}^c) = p(1 - p)$. La suite $(B_n)_{n \in \mathbf{N}^*}$ est indépendante. En effet, pour tout $n \in \mathbf{N}^*$, B_n appartient à la tribu $\mathcal{U}_n = \sigma(\sigma(A_{2n-1}), \sigma(A_{2n}))$ et donc $\sigma(B_n) \subset \mathcal{U}_n$. Mais si on considère la partition de \mathbf{N}^* , $I_n = \{2n - 1, 2n\}$, le Corollaire 4 entraîne l'indépendance des \mathcal{U}_n et donc celle des B_n . Par conséquent, \mathbb{P} -presque sûrement, une infinité de B_n est réalisé. Ceci se généralise à une intersection fini à p éléments.

Pour donner un exemple concret, imaginez que vous schiez sur des mots-croisés à la question suivante : quel célèbre mathématicien français n'a jamais rien démontré d'important ? (Prénom Nom en 8 + 6 soit 15 caractères avec l'espace). Facile : mettez les 26 lettres de l'alphabet + l'espace dans un sac, tirez un caractère, reposez-le et recommencez. Tous les 15 caractères, regardez ce que vous avez obtenu. Nous venons de montrer que vous verrez apparaître une infinité de fois la bonne réponse. Le tout c'est de ne pas être pressé car $(1/27)^{15}$ c'est tout petit.

2. Variables indépendantes.

Nous avons, jusqu'à présent, abordé la notion d'indépendance pour des tribus ; nous allons transposer cette notion pour les variables aléatoires.

Deux tribus \mathcal{F}_1 et \mathcal{F}_2 sont indépendantes si $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$ pour tous $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$. Si X_1 et X_2 sont deux variables aléatoires à valeurs dans (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) , les événements qui dépendent de la variable aléatoire X_i sont les événements A_i de la forme $\{X_i \in B_i\}$, $B_i \in \mathcal{E}_i$. Il semble donc naturel de dire que X_1 et X_2 sont indépendantes si, pour tous $B_1 \in \mathcal{E}_1$, $B_2 \in \mathcal{E}_2$,

$$\mathbb{P}(\{X_1 \in B_1\} \cap \{X_2 \in B_2\}) = \mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1)\mathbb{P}(X_2 \in B_2).$$

Le résultat suivant montre comment passer d'une variable aléatoire à une tribu. Soit X une variable aléatoire à valeurs dans (E, \mathcal{E}) . Rappelons une dernière fois que, si $B \in \mathcal{E}$, $\{X \in B\} = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. Si \mathcal{A} est une sous-tribu de \mathcal{F} , on dit que X est \mathcal{A} -mesurable si, pour tout $B \in \mathcal{E}$, $\{X \in B\} \in \mathcal{A}$.

Définition. On appelle tribu engendrée par X , notée $\sigma(X)$ – ou $\mathcal{T}(X)$ – la plus petite tribu rendant X mesurable.

Cette définition repose, comme dans le cas de la tribu engendrée par une classe de parties, sur le fait qu’une intersection quelconque de tribus sur Ω est encore une tribu sur Ω : l’intersection de toutes les tribus rendant X mesurable est donc par construction la plus petite tribu rendant cette application mesurable. Si X est une variable aléatoire, X est \mathcal{F} –mesurable et par conséquent $\sigma(X) \subset \mathcal{F}$.

Remarque. La notation $\sigma(X)$ pour la tribu engendrée est un peu dangereuse car $\sigma(X)$ peut désigner aussi l’écart type de X c’est à dire la racine carrée positive de la variance de X . Pour éviter les confusions, nous noterons désormais $\mathbb{V}(X)$ pour la variance de X .

Proposition 6. Soit X une variable aléatoire à valeurs dans (E, \mathcal{E}) . Alors,

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

Si \mathcal{I} est un π –système engendrant \mathcal{E} , alors

$$X^{-1}(\mathcal{I}) = \{\{X \in B\} : B \in \mathcal{I}\}$$

est un π –système qui engendre $\sigma(X)$.

Démonstration. Puisque $X^{-1}(B^c) = X^{-1}(B)^c$ et $X^{-1}(\cup B_n) = \cup X^{-1}(B_n)$, $X^{-1}(\mathcal{E})$ est une tribu. D’autre part, si \mathcal{A} est une tribu sur Ω , X est \mathcal{A} –mesurable si et seulement si $X^{-1}(\mathcal{E}) \subset \mathcal{A}$. Par conséquent, $\sigma(X) = X^{-1}(\mathcal{E})$.

Si \mathcal{I} est un π –système, $X^{-1}(\mathcal{I})$ en est aussi un puisque $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$. Supposons que $\sigma(\mathcal{I}) = \mathcal{E}$. $X^{-1}(\mathcal{I})$ engendre $\sigma(X)$ car $X^{-1}[\sigma(\mathcal{I})] = \sigma(X^{-1}[\mathcal{I}])$. Justifions brièvement cette dernière égalité. $X^{-1}(\mathcal{I})$ est inclus dans la tribu $X^{-1}[\sigma(\mathcal{I})]$ et donc $\sigma(X^{-1}[\mathcal{I}]) \subset X^{-1}[\sigma(\mathcal{I})]$. Réciproquement, $\{B \in \sigma(\mathcal{I}) : X^{-1}(B) \in X^{-1}[\mathcal{I}]\}$ est une tribu contenant \mathcal{I} donc $\sigma(\mathcal{I})$. Le résultat s’en suit. \square

Munis de ce formalisme, l’indépendance de variables aléatoires se ramène à l’indépendance de tribus. Considérons, dans tout ce paragraphe, pour tout $n \in \mathbf{N}$, $X_n : (\Omega, \mathcal{F}) \rightarrow (E_n, \mathcal{E}_n)$ une variable aléatoire.

Définition. La suite $(X_n)_{n \in \mathbf{N}}$ est indépendante si les tribus $\mathcal{X}_n = \sigma(X_n)$, $n \in \mathbf{N}$, sont indépendantes.

Ceci signifie donc que pour tout $n \in \mathbf{N}$, les v.a. X_0, \dots, X_n sont indépendantes c’est à dire que les tribus $\sigma(X_0), \dots, \sigma(X_n)$ sont indépendantes soit encore, via la Proposition 6, que

$$\forall (B_0, \dots, B_n) \in \mathcal{E}_0 \times \dots \times \mathcal{E}_n, \quad \mathbb{P}(X_0 \in B_0, \dots, X_n \in B_n) = \prod_{i \leq n} \mathbb{P}(X_i \in B_i).$$

Remarque. Il résulte immédiatement de la définition que, si $(\mathcal{F}_n)_{\mathbf{N}}$ sont des tribus indépendantes et que, pour tout $n \in \mathbf{N}$, X_n est \mathcal{F}_n –mesurable, alors $(X_n)_{\mathbf{N}}$ sont des v.a. indépendantes.

Un cas particulier de cette remarque est le point suivant, très utile en pratique. Soit pour tout $n \in \mathbf{N}$ – et dans tout le paragraphe –, $f_n : (E_n, \mathcal{E}_n) \rightarrow (M_n, \mathcal{M}_n)$ une application mesurable.

Proposition 7. Si $(X_n)_{\mathbf{N}}$ est une suite de variables indépendantes, il en est de même de la suite $(f_n(X_n))_{\mathbf{N}}$.

Démonstration. Il suffit de remarquer que, pour tout $n \in \mathbf{N}$, $f_n(X_n)$ est une v.a. $\sigma(X_n)$ -mesurable puisque si $U_n \in \mathcal{M}_n$, $\{f_n(X_n) \in U_n\} = \{X_n \in f_n^{-1}(U_n)\}$. \square

Remarque. On peut constater qu'une variable aléatoire réelle est $\sigma(X)$ -mesurable si et seulement s'il existe une application $h : (E, \mathcal{E}) \rightarrow (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ mesurable telle que $Y = h(X)$. En effet, si $Y = h(X)$ nous venons de voir que Y était $\sigma(X)$ -mesurable. Pour la réciproque, on utilise la « méthode standard » de l'intégration.

(i) Si $Y = \mathbf{1}_A$ alors $A = \{Y = 1\} \in \sigma(X)$; d'après la Proposition 6, il existe $B \in \mathcal{E}$ tel que $A = X^{-1}(B)$ et $Y = \mathbf{1}_B(X)$.

(ii) Par linéarité, le résultat est valable pour les v.a. étagées.

(iii) Si Y est une v.a. positive alors elle est limite simple d'une suite croissante de v.a. étagées positives. D'après l'étape précédente, $Y_n = h_n(X)$. On a alors $Y = h(X)$ avec par exemple $h(x) = \sup h_n(x)$ si le sup est fini et $h(x) = 0$ dans le cas contraire. En effet, si $x = X(\omega) \in X(\Omega)$, $h_n(x)$ converge en croissant dans \mathbf{R} vers $Y(\omega)$ et donc $h(x) = Y(\omega)$.

(iv) Si Y est une v.a. numérique, $Y = Y^+ - Y^- = h^+(X) - h^-(X)$ et si Y est complexe, on applique ce dernier résultat à sa partie réelle et sa partie imaginaire.

Lorsque Y est une v.a. bornée on peut supposer que h l'est également.

Exercice. Montrer que, si A est un événement, $\sigma(A) = \sigma(\mathbf{1}_A)$. En déduire que $(A_n)_{\mathbf{N}}$ sont des événements indépendants si et seulement si $(\mathbf{1}_{A_n})_{\mathbf{N}}$ sont des v.a. indépendantes.

Remarque. Des variables aléatoires deux à deux indépendantes ne sont pas nécessairement indépendantes dans leur ensemble. Voici un exemple. Soient X et Y deux variables aléatoires indépendantes de même loi donnée par $\mathbb{P}(X = \pm 1) = 1/2$. On pose $Z = XY$. Les couples de variables X et Y , X et Z , Y et Z sont indépendants mais le triplet X, Y, Z n'est pas indépendant. Remarquons tout d'abord que Z a la même loi que X . En effet, Z est à valeurs dans $\{-1, 1\}$, et, comme X et Y sont indépendantes,

$$\mathbb{P}(Z = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = -1, Y = -1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Montrons que X et Z sont indépendantes. On a, pour ε_1 et ε_2 dans $\{-1, 1\}$.

$$\mathbb{P}(X = \varepsilon_1, Z = \varepsilon_2) = \mathbb{P}(X = \varepsilon_1, Y = \varepsilon_1 \varepsilon_2) = \frac{1}{4} = \mathbb{P}(X = \varepsilon_1) \mathbb{P}(Z = \varepsilon_2).$$

On obtient l'indépendance de Y et Z par symétrie; ces trois variables sont donc deux à deux indépendantes. Le triplet n'est pas indépendant puisque

$$\mathbb{P}(X = 1, Y = 1, Z = 1) = \mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) \mathbb{P}(Y = 1) = \frac{1}{4} \neq \frac{1}{8}.$$

Venons-en à un résultat très important.

Théorème 8. Soient X_1, \dots, X_n des variables aléatoires. On a équivalence entre :

1. X_1, \dots, X_n sont indépendantes;
2. pour tout $(B_1, \dots, B_n) \in \mathcal{I}_1 \times \dots \times \mathcal{I}_n$, \mathcal{I}_k π -système contenant E_k et engendrant \mathcal{E}_k ,

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i \leq n} \mathbb{P}(X_i \in B_i) ;$$

3. pour toutes applications mesurables et positives $f_i : E_i \rightarrow \mathbf{R}_+$ (resp. $f_i \in \mathcal{M}_b^+$),

$$\mathbb{E} \left[\prod_{i \leq n} f_i(X_i) \right] = \prod_{i \leq n} \mathbb{E} [f_i(X_i)] ;$$

4. la loi du vecteur (X_1, \dots, X_n) est égal au produit de ses marginales soit

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

Avant de démontrer ce résultat, commentons le dernier point. Rappelons tout d'abord que si μ_1 et μ_2 sont deux mesures σ -finies sur (E_1, \mathcal{E}_1) et (E_2, \mathcal{E}_2) , il existe une unique mesure, notée $\mu_1 \otimes \mu_2$ sur $\mathcal{E}_1 \otimes \mathcal{E}_2$ vérifiant

$$\forall (B_1, B_2) \in \mathcal{E}_1 \times \mathcal{E}_2, \quad \mu_1 \otimes \mu_2(B_1 \times B_2) = \mu_1(B_1)\mu_2(B_2).$$

Si μ_1 et μ_2 sont des probabilités, il en est de même de $\mu_1 \otimes \mu_2$. Si $f : E_1 \times E_2 \rightarrow \overline{\mathbf{R}}$ ou \mathbf{C} est une application mesurable, positive ou $\mu_1 \otimes \mu_2$ -intégrable,

$$\begin{aligned} \int_{E_1 \times E_2} f(x_1, x_2) [\mu_1 \otimes \mu_2][d(x_1, x_2)] &= \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1) \\ &= \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2). \end{aligned}$$

Le cas positif est le théorème de Tonelli, celui $\mu_1 \otimes \mu_2$ -intégrable le théorème de Fubini. Ces résultats justifient l'emploi de la notation $\mu_1(dx_1)\mu_2(dx_2)$ pour $[\mu_1 \otimes \mu_2][d(x_1, x_2)]$. Je vous renvoie à l'annexe « Construction de mesures » ou au cours de F. MALRIEU [Mal] pour les détails.

La généralisation au cas de n facteurs est quasi immédiate, les notations étant plus lourdes. Nous noterons E le produit cartésien $E_1 \times \dots \times E_n$ et $x = (x_1, \dots, x_n)$ les points de E ; désignons par \mathcal{R} l'ensemble des pavés (ou rectangles) mesurables c'est à dire l'ensemble des $B \subset E$ de la forme $B = B_1 \times \dots \times B_n$, $B_i \in \mathcal{E}_i$. On définit la tribu $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ comme la tribu engendrée par \mathcal{R} . On construit sur cette tribu une probabilité (ou mesure σ -finie dans le cas général), notée $\mu_1 \otimes \dots \otimes \mu_n$, qui est l'unique probabilité μ sur $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ vérifiant :

$$\forall (B_1, \dots, B_n) \in \mathcal{E}_1 \times \dots \times \mathcal{E}_n, \quad \mu(B_1 \times \dots \times B_n) = \mu_1(B_1) \times \dots \times \mu_n(B_n).$$

Pour ce qui concerne l'intégration par rapport à cette mesure, « la règle » est la suivante : pour une fonction $f(x_1, \dots, x_n)$ mesurable, si cette fonction est positive on intègre dans le sens qu'on veut ; si f n'est pas de signe constant, on commence par considérer $|f|$ pour laquelle on peut intégrer dans le sens qu'on veut : si le résultat est fini, on peut alors intégrer f dans le sens qu'on veut. On note encore $\mu_1(dx_1) \dots \mu_n(dx_n)$ pour $[\mu_1 \otimes \dots \otimes \mu_n][d(x_1, \dots, x_n)]$. En particulier,

$$\int_E f(x_1) \dots f(x_n) \mu_1(dx_1) \dots \mu_n(dx_n) = \prod_{i \leq n} \int_{E_i} f_i(x_i) \mu_i(dx_i), \quad (2)$$

dès que les applications mesurables $f_i : E_i \rightarrow \overline{\mathbf{R}}$ ou \mathbf{C} sont positives ou vérifient

$$\forall i = 1, \dots, n, \quad \int_{E_i} |f_i(x_i)| \mu_i(dx_i) < +\infty.$$

Rappelons d'autre part que deux mesures σ -finies μ et ν sur $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ sont égales si et seulement si

$$\forall f_i \in \mathcal{M}^+(E_i), \quad \int_E f_1(x_1) \dots f_n(x_n) \mu(dx) = \int_E f_1(x_1) \dots f_n(x_n) \nu(dx). \quad (3)$$

En effet, si $\mu = \nu$, alors pour toute $f \in \mathcal{M}^+$, $\int f(x) \mu(dx) = \int f(x) \nu(dx)$ et réciproquement si la formule (3) est valable μ et ν coïncident sur le π -système \mathcal{R} des pavés mesurables donc sur la tribu qu'il engendre.

Passons à présent à la preuve du Théorème 8. Dans la preuve de ce résultat comme dans celles de ses corollaires, nous noterons X le vecteur aléatoire (X_1, \dots, X_n) et μ la mesure $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$.

Démonstration. Nous allons montrer que $(2) \Rightarrow (1) \Rightarrow (4) \Rightarrow (3) \Rightarrow (2)$.

• $(2) \Rightarrow (1)$. Nous devons montrer que les tribus $\sigma(X_1), \dots, \sigma(X_n)$ sont indépendantes. Or nous savons que

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i \leq n} \mathbb{P}(X_i \in B_i),$$

pour tout $(B_1, \dots, B_n) \in \mathcal{I}_1 \times \dots \times \mathcal{I}_n$. D'après la seconde partie de la Proposition 6, $\{X_k \in B_k\}$, $B_k \in \mathcal{I}_k$, est un π -système engendrant $\sigma(X_k)$ qui contient Ω puisque $E_k \in \mathcal{I}_k$. On conclut alors à l'aide de la Proposition 2.

• $(1) \Rightarrow (4)$. Considérons les deux mesures de probabilités sur $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$, \mathbb{P}_X et $\mu = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$. On a, pour tout $B = B_1 \times \dots \times B_n \in \mathcal{R}$, vu l'indépendance des X_i ,

$$\mathbb{P}_X(B) = \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i \leq n} \mathbb{P}(X_i \in B_i) = \prod_{i \leq n} \mathbb{P}_{X_i}(B_i) = \mu(B).$$

Les deux probabilités \mathbb{P}_X et μ coïncident donc sur le π -système \mathcal{R} ; elles sont donc égales sur la tribu qu'il engendre c'est à dire $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$. C'est une conséquence du théorème de Dynkin.

• $(4) \Rightarrow (3)$. C'est le théorème de Tonelli. Par hypothèse,

$$\mathbb{E} \left[\prod_{i \leq n} f_i(X_i) \right] = \int_E f(x_1) \dots f(x_n) \mathbb{P}_X(dx) = \int_E f(x_1) \dots f(x_n) \mathbb{P}_{X_1}(dx_1) \dots \mathbb{P}_{X_n}(dx_n).$$

Or cette dernière intégrale n'est rien d'autre que – voir la formule (2) –

$$\prod_{i \leq n} \int_{E_i} f(x_i) \mathbb{P}_{X_i}(dx_i) = \prod_{i \leq n} \mathbb{E}[f_i(X_i)],$$

puisque la fonction à intégrer est mesurable et positive.

• $(3) \Rightarrow (2)$. Il suffit de prendre $f_k = \mathbf{1}_{B_k}$ avec $B_k \in \mathcal{I}_k$. On a alors

$$\mathbb{E} \left[\prod_{i \leq n} f_i(X_i) \right] = \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i \leq n} \mathbb{E}[f_i(X_i)] = \prod_{i \leq n} \mathbb{P}(X_i \in B_i).$$

Ceci termine la preuve du résultat. □

Remarque. Si les X_i sont discrètes le point (4) signifie simplement – il suffit de sommer – que

$$\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega), \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i \leq n} \mathbb{P}(X_i = x_i).$$

Exercice. À titre d'exercice, justifiez la propriété suivante que nous démontrerons plus loin d'une autre façon : si X_1, X_2, X_3 et X_4 sont indépendantes, alors (X_1, X_2) et (X_3, X_4) sont indépendantes.

Passons à quelques conséquences de notre résultat principal.

Corollaire 9. Soient X_i , $i = 1, \dots, n$ des v.a. indépendantes et $f_i = E_i \rightarrow \overline{\mathbf{R}}$ ou \mathbf{C} des applications mesurables. Si $f_i(X_i)$ est intégrable, pour tout $i = 1, \dots, n$, le produit $\prod_{i \leq n} f_i(X_i)$ est intégrable et dans ce cas,

$$\mathbb{E} \left[\prod_{i \leq n} f_i(X_i) \right] = \prod_{i \leq n} \mathbb{E}[f_i(X_i)] ;$$

Démonstration. Pour la première partie, il suffit de remarquer que, d'après le point (3) du Théorème 8,

$$\mathbb{E} \left[\left| \prod_{i \leq n} f_i(X_i) \right| \right] = \mathbb{E} \left[\prod_{i \leq n} |f_i(X_i)| \right] = \prod_{i \leq n} \mathbb{E}[|f_i(X_i)|].$$

La deuxième assertion est alors une réécriture de la formule (2) avec $\mu_i = \mathbb{P}_{X_i}$ puisque $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$. \square

Corollaire 10. *Soient X_1, \dots, X_n des variables aléatoires réelles. Elles sont indépendantes si et seulement si*

$$\forall (t_1, \dots, t_n) \in \mathbf{R}^n, \quad \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \prod_{i \leq n} \mathbb{P}(X_i \leq t_i) = \prod_{i \leq n} F_{X_i}(t_i).$$

Démonstration. Seule la suffisance demande une justification. $\mathcal{J}_k = \{]-\infty, t_k], t_k \in \mathbf{R}\}$ est un π -système qui engendre la tribu borélienne. Toutefois nous ne pouvons pas appliquer directement le second point du Théorème 8 car $\mathbf{R} \notin \mathcal{J}_k$. On remarque alors que, par limite croissante, la formule du Corollaire est valable en fait pour $t_k \in \mathbf{R} \cup \{+\infty\}$. On applique donc la partie (2) du théorème avec comme π -système $\mathcal{I}_k = \mathcal{J}_k \cup \{\mathbf{R}\}$ qui cette fois contient \mathbf{R} . \square

Corollaire 11. *Soient X_1, \dots, X_n des variables aléatoires à valeurs dans $\mathbf{R}^{d_1}, \dots, \mathbf{R}^{d_n}$. Elles sont indépendantes si et seulement si*

$$\forall (t_1, \dots, t_n) \in \mathbf{R}^{d_1} \times \dots \times \mathbf{R}^{d_n}, \quad \varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i \leq n} \varphi_{X_i}(t_i).$$

Démonstration. D'après le point (4) du Théorème 8, les variables X_1, \dots, X_n sont indépendantes si et seulement si $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n} = \mu$. Mais deux probabilités sont égales si et seulement si leurs transformées de Fourier le sont également. Par suite, ces variables sont indépendantes si et seulement si, pour tout $t = (t_1, \dots, t_n) \in \mathbf{R}^{d_1} \times \dots \times \mathbf{R}^{d_n}$,

$$\varphi_X(t) = \widehat{\mathbb{P}_X}(t) = \widehat{\mu}(t).$$

Or, on a, $E = \mathbf{R}^s$, $s = d_1 + \dots + d_n$,

$$\begin{aligned} \widehat{\mu}(t) &= \int_{\mathbf{R}^s} e^{i \sum_{k \leq n} t_k \cdot x_k} \mathbb{P}_{X_1}(dx_1) \dots \mathbb{P}_{X_n}(dx_n) \\ &= \int_{\mathbf{R}^s} \prod_{k \leq n} e^{it_k \cdot x_k} \mathbb{P}_{X_1}(dx_1) \dots \mathbb{P}_{X_n}(dx_n); \end{aligned}$$

la formule (2) – ici $\int |e^{it_k \cdot x_k}| \mathbb{P}_{X_k}(dx_k) = 1$ – donne alors

$$\widehat{\mu}(t) = \prod_{k \leq n} \int_{\mathbf{R}^{d_k}} e^{it_k \cdot x_k} \mathbb{P}_{X_k}(dx_k) = \prod_{k \leq n} \varphi_{X_k}(t_k).$$

Le résultat s'en suit immédiatement. \square

Nous considérons dans le résultat qui suit le cas de variables à densité.

Corollaire 12. *Soit $X = (X_1, \dots, X_n)$ une v.a. à valeurs dans $\mathbf{R}^{d_1} \times \dots \times \mathbf{R}^{d_n}$. On suppose X_i de densité $p_i(x_i)$, $i \leq n$. Les X_i sont indépendantes si et seulement si X a pour densité*

$$p(x) = p(x_1, \dots, x_n) = p_1(x_1) \dots p_n(x_n).$$

Démonstration. Puisque les X_i ont pour densités respectives p_i , elles sont indépendantes si et seulement si, pour toutes applications f_i mesurables positives,

$$\mathbb{E} \left[\prod_{i \leq n} f_i(X_i) \right] = \prod_{i \leq n} \mathbb{E}[f_i(X_i)] = \prod_{i \leq n} \int_{\mathbf{R}^{d_i}} f_i(x_i) p_i(x_i) dx_i.$$

Compte tenu de la formule (2), les X_i sont indépendantes si et seulement si, pour toutes applications f_i mesurables positives ($s = d_1 + \dots + d_n$),

$$\int_{\mathbf{R}^s} f(x_1) \dots f(x_n) \mathbb{P}_X(dx) = \int_{\mathbf{R}^s} f(x_1) \dots f(x_n) p(x) dx,$$

c'est à dire si et seulement si $\mathbb{P}_X(dx) = p(x) dx$. □

Remarque. En pratique, on utilise le fait suivant : X a pour densité p , les v.a. X_1, \dots, X_n sont indépendantes si et seulement si p est à variables séparées c'est à dire si, pour presque tout $x = (x_1, \dots, x_n)$, $p(x) = q_1(x_1) \dots q_n(x_n)$.

La démonstration est laissée en exercice.

Exemple. Revenons à l'exemple de la page 27. Nous avons vu que les v.a. entières τ_r sont presque sûrement finies. En fait, les variables $\tau_r - \tau_{r-1}$, $r \in \mathbf{N}^*$, sont indépendantes et identiquement distribuées suivant la loi géométrique de paramètre p . En effet, pour tout $r \in \mathbf{N}^*$, pour tous $n_1, \dots, n_r \in \mathbf{N}^*$, l'événement $\{\tau_1 - \tau_0 = n_1, \dots, \tau_r - \tau_{r-1} = n_r\}$ n'est rien d'autre que

$$A_1^c \cap \dots \cap A_{n_1-1}^c \cap A_{n_1} \cap \dots \cap A_{n_1+\dots+n_{r-1}+1}^c \cap \dots \cap A_{n_r-1}^c \cap A_{n_r},$$

et via l'indépendance des A_n ,

$$\mathbb{P}(\tau_1 - \tau_0 = n_1, \dots, \tau_r - \tau_{r-1} = n_r) = p(1-p)^{n_1-1} \dots p(1-p)^{n_r-1}.$$

Somme de variables indépendantes. Soient X et Y deux variables aléatoires à valeurs dans \mathbf{R}^d indépendantes ; X de loi μ et Y de loi ν . Notons S la somme $X + Y$. On a alors, pour tout $B \in \mathcal{B}(\mathbf{R}^d)$,

$$\mathbb{P}(S \in B) = \iint \mathbf{1}_B(x+y) \mu(dx) \nu(dy).$$

En effet, puisque (X, Y) a pour loi $\mu \otimes \nu$,

$$\mathbb{P}(S \in B) = \mathbb{E}[\mathbf{1}_B(S)] = \iint \mathbf{1}_B(x+y) \mu(dx) \nu(dy).$$

De plus, pour tout $t \in \mathbf{R}^d$, $\varphi_S(t) = \varphi_X(t) \varphi_Y(t)$. En effet, d'après le Corollaire 9,

$$\varphi_S(t) = \mathbb{E} \left[e^{it \cdot (X+Y)} \right] = \mathbb{E} \left[e^{it \cdot X} e^{it \cdot Y} \right] = \mathbb{E} \left[e^{it \cdot X} \right] \mathbb{E} \left[e^{it \cdot Y} \right] = \varphi_X(t) \varphi_Y(t).$$

On montre de la même façon, que si X et Y sont deux variables entières, pour tout $|z| \leq 1$, $G_S(z) = G_X(z) G_Y(z)$.

Proposition 13. Si X a pour densité p alors S a pour densité la fonction

$$u(s) = \int_{\mathbf{R}^d} p(s-y) \nu(dy).$$

En particulier, si Y a pour densité q , S a pour densité le produit de convolution $p \star q$ soit

$$u(s) = \int_{\mathbf{R}^d} p(s-y) q(y) dy.$$

Démonstration. Soit f mesurable et positive. Nous avons, comme X a pour densité p , via le théorème de Tonelli,

$$\mathbb{E}[f(S)] = \iint f(x+y) \mu(dx) \nu(dy) = \int \left(\int f(x+y) p(x) dx \right) \nu(dy).$$

Or, pour $y \in \mathbf{R}^d$, le changement de variable $s = x + y$ donne

$$\int f(x+y) p(x) dx = \int f(s) p(s-y) ds$$

et le théorème de Tonelli conduit à

$$\mathbb{E}[f(S)] = \int \left(\int f(s) p(s-y) ds \right) \nu(dy) = \int f(s) \left(\int p(s-y) \nu(dy) \right) ds = \int f(s) u(s) ds,$$

ce qui donne la première partie de cette proposition. La seconde est immédiate. \square

Exemple. Si X et Y sont deux gaussiennes indépendantes, X de loi $\mathcal{N}(\alpha, \sigma^2)$ et Y de loi $\mathcal{N}(\beta, \tau^2)$, alors $X + Y$ est une gaussienne de loi $\mathcal{N}(\alpha + \beta, \sigma^2 + \tau^2)$.

En effet, pour tout $t \in \mathbf{R}$,

$$\varphi_S(t) = \varphi_X(t) \varphi_Y(t) = e^{it\alpha - \frac{\sigma^2 t^2}{2}} e^{it\beta - \frac{\tau^2 t^2}{2}} = e^{it(\alpha+\beta) - \frac{t^2}{2}(\sigma^2 + \tau^2)};$$

c'est la fonction caractéristique de la gaussienne $\mathcal{N}(\alpha + \beta, \sigma^2 + \tau^2)$.

Exercice. Soient X et Y deux v.a.r. indépendantes. Montrer que $X + Y$ est intégrable si et seulement si X et Y le sont.

Deux variables aléatoires réelles X et Y de carré intégrable sont dites *non-corrélées* lorsque $\text{Cov}(X, Y) = 0$. Ceci revient à dire que $\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]$. On a dans ce cas

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Notons que deux v.a.r. indépendantes de carré intégrable sont non-corrélées. La réciproque est fautive comme le montre l'exemple suivant.

Exemple. Soient X une v.a.r. normale centrée réduite et ε indépendante de X de loi donnée par $\mathbb{P}(\varepsilon = \pm 1) = 1/2$. On note $Y = \varepsilon X$. X et Y sont non-corrélées. En effet, comme $\mathbb{E}[X] = \mathbb{E}[\varepsilon] = 0$, nous avons, vu l'indépendance de X et ε ,

$$\mathbb{E}[XY] = \mathbb{E}[\varepsilon X^2] = \mathbb{E}[\varepsilon] \mathbb{E}[X^2] = 0 = \mathbb{E}[X] \mathbb{E}[Y].$$

Déterminons la fonction caractéristique φ du couple (X, Y) . Pour tous réels s et t , nous avons, puisque X et ε sont indépendantes,

$$\varphi(s, t) = \mathbb{E} \left[e^{isX + itY} \right] = \mathbb{E} \left[\mathbf{1}_{\varepsilon=1} e^{i(s+t)X} + \mathbf{1}_{\varepsilon=-1} e^{i(s-t)X} \right] = \frac{1}{2} (\varphi_X(s+t) + \varphi_X(s-t)).$$

Or $\varphi_X(s) = \exp\left(-\frac{s^2}{2}\right)$ et donc $\varphi(s, t) = \exp\left(-\frac{s^2+t^2}{2}\right) \cosh(st)$. En particulier, Y est également normale centrée réduite puisque $\varphi_Y(t) = \varphi(0, t) = \exp\left(-\frac{t^2}{2}\right)$.

X et Y ne sont pas indépendantes car sinon on aurait $\varphi(s, t) = \varphi_X(s) \varphi_Y(t)$ c'est à dire $\cosh(st) = 1$ pour tout (s, t) .

On voit sur cet exemple que, bien que X et Y soient toutes deux gaussiennes, ni le vecteur (X, Y) ni la somme $X + Y$ ne sont gaussiens.

Remarque. Pour des variables aléatoires réelles de carré intégrable, X_1, \dots, X_n , deux à deux non-corrélées – pour tout $i \neq j$, $\text{Cov}(X_i, X_j) = 0$ – on a

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n).$$

3. Construction de variables indépendantes.

Nous avons étudié quelques propriétés des variables aléatoires indépendantes. Mais une question se pose : comment construire de telles variables ? existe-t-il des variables aléatoires indépendantes ?

Pour un nombre fini – deux pour la clarté de l'exposé –, la construction n'est pas très compliquée. Pour être un peu plus précis, considérons deux espaces probabilisés $(E_1, \mathcal{E}_1, \mu_1)$ et $(E_2, \mathcal{E}_2, \mu_2)$. Nous allons construire deux variables aléatoires X_1 et X_2 indépendantes, X_1 de loi μ_1 et X_2 de loi μ_2 . Pour cela, considérons l'espace $(\Omega, \mathcal{F}, \mathbb{P}) = (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2, \mu_1 \otimes \mu_2)$ et prenons pour X_1 la projection sur E_1 et X_2 la projection sur E_2 ; en d'autres termes

$$\forall \omega = (x_1, x_2) \in E_1 \times E_2, \quad X_1(\omega) = x_1, \quad X_2(\omega) = x_2.$$

X_1 et X_2 sont alors deux variables indépendantes à valeurs dans E_1 et E_2 ; X_1 a pour loi μ_1 et X_2 a pour loi μ_2 . En effet, pour tous $B_1 \in \mathcal{E}_1$ et $B_2 \in \mathcal{E}_2$,

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mu_1 \otimes \mu_2(B_1 \times B_2) = \mu_1(B_1)\mu_2(B_2).$$

Prenant $B_2 = E_2$ puis $B_1 = E_1$ dans la formule précédente, on s'aperçoit que X_i a pour loi μ_i et donc que

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1) \mathbb{P}(X_2 \in B_2).$$

Cette construction s'étend sans difficulté au cas d'un nombre fini de variables aléatoires.

Plus délicate est la construction d'une suite $(X_n)_{\mathbf{N}^*}$ de variables indépendantes telle que, pour tout $n \in \mathbf{N}^*$, X_n soit à valeurs dans (E_n, \mathcal{E}_n) de loi μ_n . La démonstration de ce résultat repose sur le théorème de Carathéodory et est donnée dans l'annexe « Construction de mesures ». Nous nous contentons ici de donner le résultat.

Construisons d'abord la tribu $\otimes_{n \geq 1} \mathcal{E}_n$. Notons E le produit cartésien $\prod_{n \geq 1} E_n$, X_n la projection canonique de E sur E_n et Π_n celle de E sur $E_1 \times \dots \times E_n$. Un ensemble B de E est cylindrique s'il existe $n \in \mathbf{N}^*$ et $C_n \in \mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$ tels que $B = \Pi_n^{-1}(C_n)$ soit $B = C_n \times \prod_{i > n} E_i$. La tribu $\otimes_{n \geq 1} \mathcal{E}_n$ est, par définition, la tribu engendrée par les ensembles cylindriques. On montre assez facilement que cette tribu est la plus petite tribu sur E rendant toutes les projections X_n mesurables. Elle est également engendrée par le π -système \mathcal{R} formé par les ensembles du type

$$B_1 \times \dots \times B_n \times \prod_{i > n} E_i, \quad B_i \in \mathcal{E}_i, \quad n \in \mathbf{N}^*.$$

On montre alors le résultat suivant :

Théorème 14. *Soit, pour tout $n \in \mathbf{N}^*$, $(E_n, \mathcal{E}_n, \mu_n)$ un espace probabilisé. Il existe une unique probabilité μ sur $(\prod_{n \geq 1} E_n, \otimes_{n \geq 1} \mathcal{E}_n)$, notée $\otimes_{n \geq 1} \mu_n$, telle que, pour tous $B_n \in \mathcal{E}_n$,*

$$\left(\otimes_{n \geq 1} \mu_n \right) \left(\prod_{n \geq 1} B_n \right) = \prod_{n \geq 1} \mu_n(B_n).$$

La suite $(X_n)_{n \in \mathbf{N}^}$ des projections est une suite de variables aléatoires indépendantes et, pour tout $n \in \mathbf{N}^*$, X_n a pour loi μ_n .*

4. Tribu asymptotique.

Soit $(\mathcal{F}_n)_{n \in \mathbf{N}^*}$ une suite de sous-tribus de \mathcal{F} . On note, pour tout $n \in \mathbf{N}^*$,

$$\mathcal{U}_n = \sigma(\mathcal{F}_i, i \leq n), \quad \mathcal{A}^n = \sigma(\mathcal{F}_i, i \geq n).$$

Définition. On appelle tribu asymptotique de la suite $(\mathcal{F}_n)_{n \in \mathbf{N}^*}$ la tribu \mathcal{A}^∞ définie par

$$\mathcal{A}^\infty = \bigcap_{n \geq 1} \mathcal{A}^n.$$

Un événement A appartient à la tribu asymptotique si et seulement si, pour tout $r \in \mathbf{N}^*$, il appartient à la tribu \mathcal{A}^r . On dit que A est asymptotique.

Exemple. Considérons, pour tout $n \in \mathbf{N}^*$, $A_n \in \mathcal{F}_n$. Alors l'événement $A = \limsup A_n$ est asymptotique. Rappelons tout d'abord que $A = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k$. Fixons $r \in \mathbf{N}^*$. Puisque les ensembles $B_n = \bigcup_{k \geq n} A_k$ sont décroissants, $A = \bigcap_{n \geq r} B_n$: il suffit donc de montrer que, pour tout $n \geq r$, $B_n \in \mathcal{A}^r$ pour montrer que $A \in \mathcal{A}^r$. Or, pour chaque $n \geq 1$, $B_n = \bigcap_{k \geq n} A_k$ appartient à \mathcal{A}^n puisque $A_k \in \mathcal{F}_k$. La suite de tribus \mathcal{A}^n étant décroissante, pour tout $n \geq r$, $B_n \in \mathcal{A}^r$. Donc $A \in \mathcal{A}^r$; ceci étant valable pour tout $r \in \mathbf{N}^*$, A est asymptotique.

Théorème 15 (Loi du « tout ou rien » de Kolmogorov). *Soit $(\mathcal{F}_n)_{n \in \mathbf{N}^*}$ une suite de tribus indépendantes. Alors,*

$$\forall A \in \mathcal{A}^\infty, \quad \mathbb{P}(A) = 0 \quad \text{ou} \quad \mathbb{P}(A) = 1.$$

Démonstration. D'après le Corollaire 4, pour tout $n \geq 1$, les tribus \mathcal{U}_n et \mathcal{A}^{n+1} sont indépendantes. Comme $\mathcal{A}^\infty \subset \mathcal{A}^{n+1}$, \mathcal{A}^∞ et \mathcal{U}_n sont indépendantes pour tout $n \in \mathbf{N}^*$. On a donc,

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

La suite de tribus \mathcal{U}_n étant croissante, $\bigcup_{n \geq 1} \mathcal{U}_n$ est un π -système. Il contient Ω et engendre la tribu $\mathcal{U}_\infty = \sigma(\mathcal{F}_n, n \geq 1)$; voir le Lemme 3 et la remarque de la page 25. La Proposition 2, implique donc que les tribus \mathcal{U}_∞ et \mathcal{A}^∞ sont indépendantes : on peut donc prendre, dans la formule précédente, $B \in \mathcal{U}_\infty$. Or la tribu \mathcal{A}^∞ est incluse dans \mathcal{U}_∞ : on peut donc prendre $B = A$. Par conséquent, pour tout $A \in \mathcal{A}^\infty$,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \mathbb{P}(A),$$

et donc $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$. □

Avant de passer à une conséquence de la loi du tout ou rien, revenons un instant sur les lemmes de Borel–Cantelli. Soit $(A_n)_{n \in \mathbf{N}}$ une suite d'événements indépendants. Par définition, les tribus $\sigma(A_n)$ sont indépendantes. Or nous avons vu que $\limsup A_n$ est un événement asymptotique de cette suite. Il est donc normal au vu du résultat précédent que $\mathbb{P}(\limsup A_n)$ vaille 0 ou 1 pour, répétons-le encore, des événements indépendants.

Corollaire 16. *Si $(\mathcal{F}_n)_{n \in \mathbf{N}^*}$ est une suite de tribus indépendantes, toute variable aléatoire réelle asymptotique (c'est à dire \mathcal{A}^∞ -mesurable) est presque sûrement constante.*

Démonstration. Soit X une v.a.r. asymptotique. Montrons qu'il existe une constante réelle c telle que $\mathbb{P}(X = c) = 1$. Notons F la fonction de répartition de X . La variable aléatoire X étant \mathcal{A}^∞ -mesurable, $\{X \leq t\}$ est un événement asymptotique pour tout réel t . La suite $(\mathcal{F}_n)_{n \in \mathbf{N}}$

étant indépendante, pour tout $t \in \mathbf{R}$, $F(t) = \mathbb{P}(X \leq t) = 0$ ou 1 . Considérons l'élément de $\overline{\mathbf{R}}$ $c = \inf\{u \in \mathbf{R} : F(u) = 1\}$ où $\inf \emptyset = +\infty$. Notons qu'il existe deux réels s et t , $s < t$, tels que $F(s) = 0$, $F(t) = 1$. En effet, comme $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$, il existe deux réels s et t tels que $F(s) \leq 1/2$ et $F(t) \geq 1/2$. On a donc $F(t) = 1$ et $F(s) = 0$. Par conséquent, $c \in \mathbf{R}$. F étant croissante, on a, par définition de la borne inférieure, $F(t) = 1$ si $t > c$ et $F(t) < 1$ si $t < c$, soit $F(t) = 0$ pour $t < c$. F est continue à droite donc $F(c) = 1$. Par suite, $F(t) = \mathbf{1}_{\mathbf{R}_+}(t - c)$ et $\mathbb{P}(X = c) = F(c) - F(c-) = 1 - 0 = 1$. \square

Cas des variables aléatoires. Comme d'habitude, le passage des tribus aux variables aléatoires se fait par l'intermédiaire de la tribu engendrée par une variable aléatoire. Plus précisément, nous noterons $\sigma(X_i, i \geq n)$ la tribu

$$\sigma(X_i, i \geq n) = \sigma(\sigma(X_i), i \geq n).$$

Pour une suite $(X_n)_{n \geq 1}$ de variables aléatoires, la tribu asymptotique de cette suite est donc

$$\mathcal{A}^\infty = \bigcap_{n \geq 1} \sigma(X_i, i \geq n),$$

ce qui correspond bien à la définition précédente avec $\mathcal{F}_n = \sigma(X_n)$, $\mathcal{A}^n = \sigma(X_i, i \geq n)$.

Si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes, nous venons de voir que, pour tout $A \in \mathcal{A}^\infty$, $\mathbb{P}(A) = 0$ ou 1 , et que, pour toute v.a.r. X \mathcal{A}^∞ -mesurable, X était presque sûrement constante.

Une question se pose alors : peut-on caractériser les événements (ou les variables réelles) asymptotiques d'une suite de variables aléatoires ? Il faut retenir la chose suivante : un événement ou une variable aléatoire est asymptotique si et seulement si on peut le décrire, pour tout $r \geq 1$, à l'aide des variables X_n , $n \geq r$. Avant de donner un sens plus précis à cette assertion, voyons sur des exemples sa signification.

Exemple. Soit $(X_n)_{\mathbf{N}^*}$ une suite de v.a.r. Montrons que l'événement

$$A = \{\omega \in \Omega : X_n(\omega) \text{ converge dans } \mathbf{R}\}$$

est un événement asymptotique et que la variable aléatoire $X(\omega) = \mathbf{1}_A(\omega) \lim_{n \rightarrow +\infty} X_n(\omega)$ l'est également. La raison fondamentale est : la convergence d'une suite comme la limite d'une suite ne dépend pas des premiers termes de cette suite. On a donc, pour tout $r \in \mathbf{N}^*$,

$$A = \{\omega \in \Omega : X_{r+n}(\omega) \text{ converge dans } \mathbf{R}\} ;$$

ce dernier ensemble appartient à \mathcal{A}^r puisque, \mathbf{R} étant complet,

$$A = \bigcap_{k \geq 1} \bigcup_{N \geq 1} \bigcap_{p \in \mathbf{N}} \{|X_{r+N} - X_{r+N+p}| \leq 2^{-k}\}.$$

Essayons de donner un argument qui s'écrit plus facilement. Commençons par remarquer que $\liminf X_n$ est une v.a. numérique asymptotique de la suite X_n . Rappelons tout d'abord que $\liminf X_n(\omega) = \sup_{n \geq 1} \inf_{k \geq n} X_k(\omega)$. La suite de v.a. numériques $\inf_{k \geq n} X_k$ étant croissante, on a, pour tout $r \in \mathbf{N}^*$,

$$\liminf X_n = \sup_{n \geq r} \inf_{k \geq n} X_k.$$

(pour tout r , on décrit la $\liminf X_n$ à l'aide des v.a.r. X_k , $k \geq r$.) Or, pour tout $k \geq r$, X_k est $\sigma(X_k)$ -mesurable et donc \mathcal{A}^r -mesurable puisque $\mathcal{A}^r = \sigma(X_i, i \geq r)$. La v.a. $\sup_{n \geq r} \inf_{k \geq n} X_k$ est

donc \mathcal{A}^r -mesurable. Finalement, pour tout $r \geq 1$, $\liminf X_n$ est \mathcal{A}^r -mesurable ce qui signifie que $\liminf X_n$ est asymptotique. On montre de la même manière que $\limsup X_n$ est asymptotique ; je vous conseille de le faire. L'ensemble $B = \{\omega \in \Omega : \liminf X_n(\omega) = \limsup X_n(\omega)\}$ est donc asymptotique ; or $\omega \in B$ si et seulement si la suite $X_n(\omega)$ converge dans $\overline{\mathbf{R}}$. Pour établir que A est asymptotique écrivons que A est l'intersection des trois événements asymptotiques

$$A = B \cap \{\liminf X_n > -\infty\} \cap \{\limsup X_n < +\infty\}.$$

La variable $X = \mathbf{1}_A \lim_{n \rightarrow +\infty} X_n$ est asymptotique puisqu'elle est égale à $\mathbf{1}_A \limsup X_n$. En résumé, si la suite $(X_n)_{n \geq 1}$ est indépendante, de deux choses l'une : ou bien, presque sûrement, elle ne converge pas dans \mathbf{R} , ou bien, presque sûrement, elle est convergente. Dans le second cas, la limite X est une v.a.r. asymptotique qui est donc presque sûrement constante.

Nous nous intéresserons plus loin dans le cours à la convergence de la suite de v.a.r. $n^{-1}S_n$ où $S_n = \sum_{i \leq n} X_i$. Montrons que l'ensemble $C = \{n^{-1}S_n \text{ converge dans } \mathbf{R}\}$ est un événement asymptotique de la suite X_n . Pour cela il suffit de remarquer que, $\liminf n^{-1}S_n$ et $\limsup n^{-1}S_n$ sont des variables asymptotiques de la suite X_n . En effet, nous avons – c'est pareil pour \liminf – puisque $n^{-1}S_{r-1}$ converge vers 0 si $n \rightarrow +\infty$, r étant fixé,

$$\limsup \frac{1}{n}S_n = \limsup \frac{1}{n}(S_{r-1} + S_n - S_{r-1}) = \limsup \frac{1}{n} \sum_{i=r}^n X_i.$$

Par conséquent, pour tout $r \in \mathbf{N}^*$, $\limsup n^{-1}S_n$ et $\liminf n^{-1}S_n$ sont des variables aléatoires $\sigma(X_i, i \geq r)$ -mesurables.

$$C = \{\liminf n^{-1}S_n = \limsup n^{-1}S_n\} \cap \{\liminf n^{-1}S_n > -\infty\} \cap \{\limsup n^{-1}S_n < +\infty\}$$

est donc un événement asymptotique. La variable aléatoire $\mathbf{1}_C \lim n^{-1}S_n = \mathbf{1}_C \liminf n^{-1}S_n$ l'est aussi. Si la suite $(X_n)_{n \geq 1}$ est indépendante, $n^{-1}S_n$ converge dans \mathbf{R} avec probabilité 1 ou diverge avec probabilité 1 et, dans le premier cas, la limite est déterministe.

Variable aléatoire à valeurs dans un produit dénombrable. Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires à valeurs dans (E_n, \mathcal{E}_n) . Nous avons, dans le paragraphe précédent, construit la tribu « produit dénombrable », $\otimes_{n \geq 1} \mathcal{E}_n$ sur le produit cartésien $E = \prod_{n \geq 1} E_n$. C'est la tribu engendrée par la famille \mathcal{R} des ensembles $B_1 \times \dots \times B_n \times \prod_{i > n} E_i$, $n \in \mathbf{N}$, $B_i \in \mathcal{E}_i$. On peut considérer toute la suite $X = (X_n)_{n \geq 1}$ comme une application de Ω dans le produit E et se demander si c'est une variable aléatoire, c'est à dire une application mesurable par rapport aux tribus \mathcal{F} et $\otimes_{n \geq 1} \mathcal{E}_n$. En fait, tout marche comme dans le cas fini, c'est à dire que X est une variable aléatoire si et seulement si chacune de ses composantes sont des variables aléatoires. En effet, si X est une v.a. $\{X_n \in B_n\} = \{X \in \prod_{i < n} E_i \times B_n \times \prod_{i > n} E_i\}$ appartient à \mathcal{F} pour tout $B_n \in \mathcal{E}_n$ et tout $n \in \mathbf{N}^*$. Réciproquement, si toutes les X_n sont mesurables, pour tout $R \in \mathcal{R}$, $R = B_1 \times \dots \times B_n \times \prod_{i > n} E_i$, $n \in \mathbf{N}$, $B_i \in \mathcal{E}_i$, $\{X \in R\} = \cap_{i \leq n} \{X_i \in B_i\}$ appartient à \mathcal{F} . Ceci montre que X est une v.a. puisque \mathcal{R} est un π -système engendrant $\otimes_{n \geq 1} \mathcal{E}_n$.

On peut alors essayer de comparer les deux tribus $\sigma(X_i, i \geq 1) = \sigma(\sigma(X_i), i \geq 1)$ et $\sigma(X)$ où X est à valeurs dans le produit E . En fait, ces deux tribus sont égales. En effet, d'après le second point de la Proposition 6, $\sigma(X)$ est engendrée par le π -système $X^{-1}(\mathcal{R})$. Mais via le Lemme 3, la tribu $\sigma(X_i, i \geq 1)$ est engendrée par le π -système

$$C = \left\{ \bigcap_{i \leq n} A_i, n \in \mathbf{N}^*, A_i \in \sigma(X_i) \right\}.$$

Or d'après le premier point de la Proposition 6, $A_i \in \sigma(X_i)$ si et seulement s'il existe $B_i \in \mathcal{E}_i$ tel que $A_i = \{X_i \in B_i\}$. Comme nous l'avons déjà dit, si $R = B_1 \times \dots \times B_n \times \prod_{i>n} E_i$, $\{X \in R\} = \cap_{i \leq n} \{X_i \in B_i\}$. On a donc $\mathcal{R} = \mathcal{C}$. Les deux tribus sont donc égales.

Essayons de voir quelques conséquences de ce résultat. Tout d'abord on peut remarquer la chose suivante.

Proposition 17. *Soit $(X_n)_{\mathbf{N}}$ une suite de variables aléatoires indépendantes et $(I_k)_{k \in K}$ une partition de \mathbf{N} avec $K = \mathbf{N}^*$ ou $K = \{1, \dots, p\}$. On définit, pour tout $k \in K$, la variable aléatoire – à valeurs dans un produit – $Y_k = (X_i)_{i \in I_k}$. Alors les variables aléatoires Y_k , $k \in K$, sont indépendantes.*

Démonstration. Nous devons montrer que les tribus $\sigma(Y_k)$, $k \in K$, sont indépendantes. Or d'après la caractérisation précédente, pour tout $k \in K$, la tribu $\sigma(Y_k)$ est égale à la tribu $\sigma(X_i, i \in I_k)$ qui par définition est la tribu $\sigma(\sigma(X_i), i \in I_k)$. Mais via le Corollaire 4, les tribus $\sigma(\sigma(X_i), i \in I_k)$, $k \in K$, sont indépendantes. \square

En particulier, dans le contexte de la proposition précédente, si f_n est une application mesurable alors les $f_n(Y_n)$ sont encore indépendantes. Voici un exemple d'application. On part d'une suite $(X_n)_{\mathbf{N}}$ de variables réelles indépendantes. On considère alors le temps d'arrêt $\tau = \inf\{n \geq 0 : X_{2n} \geq 1\}$ et $\rho = \sum_{n \geq 0} 2^{-n} (1 + |X_{2n+1}|)^{-1}$. Alors τ et ρ sont indépendantes puisque τ ne dépend que des X_i pour i pair et ρ seulement des X_i pour i impair : il suffit de considérer la partition formée par les pairs d'un côté et les impairs de l'autre.

Remarque. Le résultat précédent s'applique au cas d'un nombre fini n de variables aléatoires. Il suffit de prendre, pour $i > n$, X_i constante pour se ramener au cas d'une suite ; en effet, si X_i est constante $\sigma(X_i) = \{\emptyset, \Omega\}$.

Revenons à présent à la tribu asymptotique d'une suite $(X_n)_{\mathbf{N}}$. Une variable réelle Y est asymptotique si elle appartient à la tribu $\sigma(X_n, n \geq r)$ pour tout $r \in \mathbf{N}$. Or nous venons de voir que cette tribu est la tribu engendrée par la variable $X^{(r)} = (X_i)_{i \geq r}$. D'après la remarque de la page 29, il existe une fonction mesurable f_r telle que $Y = f_r(X^{(r)})$. On retrouve bien ici l'idée qu'une variable asymptotique s'exprime à partir des X_i , $i \geq r$, et ceci pour tout r .

Bibliographie.

[Mal] F. Malrieu, *Cours d'intégration et probabilités*, URL : <http://name.math.univ-rennes1.fr/florent.malrieu/>.

Chapitre III. Convergence d'une suite de variables aléatoires

Le but de ce paragraphe est d'étudier les différents types de convergence pour une suite de variables aléatoires $(X_n)_{n \in \mathbf{N}}$ disons pour simplifier à valeurs dans \mathbf{R}^d . Nous distinguerons quatre type de convergence qui se divisent en deux catégories. La première concerne les modes de convergence « trajectorielle » pour lesquels on peut déduire une information plus ou moins précise sur la convergence de la suite $X_n(\omega)$ à ω fixé. Dans ce cas de figure, les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. La convergence en loi, quant à elle, ne concerne pas directement la convergence de la suite d'applications mesurables X_n mais traite de la convergence de la suite de mesures de probabilité \mathbb{P}_{X_n} .

1. Convergence trajectorielle.

Dans tout ce paragraphe, on se place sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et on considère une suite $(X_n)_{n \in \mathbf{N}}$ de variables aléatoires à valeurs dans \mathbf{R}^d définies sur cet espace ainsi qu'une variable X également à valeurs dans \mathbf{R}^d . On désigne par $|x|$ la norme euclidienne de $x \in \mathbf{R}^d$; dans \mathbf{R}^d , toutes les normes sont équivalentes et le choix de la norme euclidienne est arbitraire.

Commençons par des notions que vous connaissez déjà.

Définition. La suite $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement vers X s'il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et

$$\forall \omega \in N^c, \quad \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega).$$

On dit aussi que $(X_n)_{n \in \mathbf{N}}$ converge vers X avec probabilité 1 puisque l'ensemble N^c précèdent sur lequel $X_n(\omega) \rightarrow X(\omega)$ est de mesure un.

Bien évidemment (justifiez-le) $X_n = (X_n^1, \dots, X_n^d)$ converge \mathbb{P} -p.s. vers $X = (X^1, \dots, X^d)$ si et seulement si, pour tout $i = 1, \dots, d$, X_n^i converge vers X^i presque sûrement.

La convergence presque sûre est facile à caractériser. En effet,

Proposition 1. $(X_n)_{n \in \mathbf{N}}$ converge vers X presque sûrement si et seulement si

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

Démonstration. La condition est nécessaire. Fixons $\varepsilon > 0$. Si $\omega \in \limsup\{|X_n - X| > \varepsilon\}$, il existe une infinité d'entiers n pour lesquels $|X_n(\omega) - X(\omega)| > \varepsilon$ et la suite $X_n(\omega)$ ne converge pas vers $X(\omega)$. En d'autres termes, $\limsup\{|X_n - X| > \varepsilon\} \subset \{X_n \rightarrow X\}^c$ et donc

$$\mathbb{P}(\limsup\{|X_n - X| > \varepsilon\}) \leq 1 - \mathbb{P}(X_n \rightarrow X) = 0.$$

La condition est également suffisante. Considérons $N = \bigcup_{r \in \mathbf{N}} \limsup\{|X_n - X| > 2^{-r}\}$. On a

$$\mathbb{P}(N) \leq \sum_{r \in \mathbf{N}} \mathbb{P}(\limsup\{|X_n - X| > 2^{-r}\}) = 0.$$

Si $\omega \in N^c$, pour tout $r \in \mathbf{N}$, $\omega \in \liminf\{|X_n - X| \leq 2^{-r}\}$ et par suite, il existe un entier n_ω tel que pour tout $k \geq n_\omega$, $|X_k(\omega) - X(\omega)| \leq 2^{-r}$. $X_n(\omega)$ converge donc vers $X(\omega)$ pour tout $\omega \in N^c$. \square

Corollaire 2. *Si, pour tout $\varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$ alors $(X_n)_{n \in \mathbf{N}}$ converge vers X presque sûrement.*

Démonstration. Il suffit d'appliquer le lemme de Borel–Cantelli. \square

Exercice. Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables indépendantes. Montrer que la suite $(X_n)_{n \in \mathbf{N}}$ converge vers 0 presque sûrement si et seulement si pour tout $\varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon) < +\infty$.

Remarque. Si $(X_n)_{n \in \mathbf{N}}$ est une suite de variables aléatoires identiquement distribuées, $n^{-1}X_n$ converge vers 0 dès que X_0 est intégrable. En effet, on a, pour tout $\varepsilon > 0$,

$$\sum_{n \geq 0} \mathbb{P}(|X_n| > n\varepsilon) = \sum_{n \geq 0} \mathbb{P}(|X_0| > n\varepsilon) = \sum_{n \geq 0} \mathbb{P}(\varepsilon^{-1}|X_0| > n);$$

cette quantité est finie si et seulement si $\mathbb{E}[\varepsilon^{-1}|X_0|] < +\infty$ c'est à dire si et seulement si $|X_0|$ est intégrable.

Si les X_n sont de plus indépendantes cette condition est également nécessaire. En effet, dans ce cas via les lemmes de Borel-Cantelli, $\mathbb{P}(\limsup\{|X_n| > n\varepsilon\}) = 0$ si et seulement si $\sum_{n \geq 0} \mathbb{P}(|X_n| > n\varepsilon) < +\infty$.

On a également un résultat analogue lorsqu'on contrôle les accroissements de la suite X_n . Plus précisément,

Lemme 3. *Soit $(\varepsilon_n)_{n \in \mathbf{N}}$ une suite de réels positifs telle que $\sum_{n \geq 0} \varepsilon_n < +\infty$. Supposons que $\sum_{n \geq 0} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$. Alors $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement.*

Démonstration. Écrivons pour établir la convergence que

$$X_n(\omega) = X_0(\omega) + \sum_{k=0}^{n-1} [X_{k+1}(\omega) - X_k(\omega)].$$

D'après le lemme de Borel-Cantelli, $\mathbb{P}(\limsup\{|X_{n+1} - X_n| > \varepsilon_n\}) = 0$. Or si ω appartient à $\liminf\{|X_{n+1} - X_n| \leq \varepsilon_n\}$, il existe un entier n_ω tel que $\forall k \geq n_\omega$, $|X_{k+1}(\omega) - X_k(\omega)| \leq \varepsilon_k$; la série définissant $X_n(\omega)$ est alors absolument convergente. \square

Voici quelques propriétés immédiates de la convergence presque sûre. Si $(X_n)_{n \in \mathbf{N}}$ converge vers X presque sûrement et si $f : \mathbf{R}^d \rightarrow \mathbf{R}^g$ est une fonction continue alors $(f(X_n))_{n \in \mathbf{N}}$ converge vers $f(X)$ presque sûrement.

Si $(X_n)_{n \in \mathbf{N}}$ converge vers X et Y presque sûrement alors X et Y sont égales presque sûrement; d'autre part, si pour tout $n \geq 0$, $X_n = Y_n$ presque sûrement, alors X_n converge vers X presque sûrement si et seulement si Y_n converge vers X presque sûrement.

On note \mathcal{L}^0 l'ensemble des variables aléatoires à valeurs dans \mathbf{R}^d et L^0 l'espace quotient \mathcal{L}^0 / \sim où \sim est la relation d'équivalence définie par $X \sim Y$ si $X = Y$ presque sûrement. Les deux remarques précédentes montrent que la convergence presque sûre se définit sur l'espace quotient L^0 .

Continuons avec une notion que vous connaissez : la convergence dans L^p . Soit $p \geq 1$ un réel. Une variable aléatoire X (dans \mathbf{R}^d) possède un moment d'ordre p si $\mathbb{E}[|X|^p] < +\infty$; ceci signifie que chacune de ses composantes sont de puissance p intégrable. On note $X \in \mathcal{L}^p$.

Définition. Soit $(X_n)_{n \in \mathbf{N}} \subset \mathcal{L}^p$. $(X_n)_{n \in \mathbf{N}}$ converge vers X en moyenne d'ordre p ou encore dans L^p si

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

Ceci est, comme pour la convergence presque sûre, équivalent à la convergence de chacune des composantes. Si $(X_n)_{n \in \mathbf{N}} \subset \mathcal{L}^p$ converge vers X en moyenne d'ordre p alors X appartient à \mathcal{L}^p .

Lorsque $p = 1$ on parle de convergence en moyenne, et pour $p = 2$ de convergence en moyenne quadratique.

La convergence en moyenne ne relève en réalité que des classes d'équivalences modulo l'égalité presque sûre. On note L^p l'espace quotient \mathcal{L}^p / \sim . Rappelons le résultat central que vous avez vu en intégration. Pour $X \in L^p$, on note $\|X\|_p = \mathbb{E} [|X|^p]^{\frac{1}{p}}$.

Théorème 4. Pour tout $p \geq 1$, $(L^p, \|\cdot\|_p)$ est un espace de Banach.

Finissons par une propriété propre au fait que nous travaillons sur un espace probabilisé.

Proposition 5. Soit $(X_n)_{n \in \mathbf{N}} \subset L^q$. Si $(X_n)_{n \in \mathbf{N}}$ converge vers X dans L^q alors $(X_n)_{n \in \mathbf{N}}$ converge vers X dans L^p pour tout $p \leq q$.

Démonstration. D'après l'inégalité de Hölder, on a

$$\mathbb{E} [|X_n - X|^p] \leq \mathbb{E} [|X_n - X|^q]^{\frac{p}{q}}.$$

Le résultat découle immédiatement de cette inégalité. \square

Signalons également que le théorème de convergence dominée permet de passer de la convergence presque sûre à la convergence en moyenne.

Théorème 6 (Convergence dominée). Soient $(X_n)_{n \in \mathbf{N}}$ et X des variables aléatoires. Supposons que $(X_n)_{n \in \mathbf{N}}$ converge vers X presque sûrement. S'il existe une variable aléatoire Y telle que,

$$\forall n \geq 0, \quad |X_n| \leq Y \quad \mathbb{P}\text{-p.s.}, \quad \text{et} \quad Y \in L^p,$$

alors $(X_n)_{n \in \mathbf{N}} \subset L^p$ et converge vers X dans L^p .

Remarque. On peut également considérer le cas $p = +\infty$. Rappelons que

$$\|X\|_\infty = \text{ess sup}|X| = \inf\{c \geq 0 : \mathbb{P}(|X| > c) = 0\} = \min\{c \geq 0 : \mathbb{P}(|X| > c) = 0\}.$$

$(L^\infty, \|\cdot\|_\infty)$ est également un espace de Banach.

Notons que, puisque nous travaillons avec une mesure de probabilité, si $(X_n)_{n \in \mathbf{N}}$ converge vers X dans L^∞ alors X_n converge vers X dans L^p pour tout $p \geq 1$. La convergence dans L^∞ entraîne également la convergence presque sûre.

Exemple. Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables réelles indépendantes. La loi de X_n est donnée par $\mathbb{P}(X_n = 0) = 1 - p_n$, $\mathbb{P}(X_n = x_n) = p_n$ où $0 < p_n < 1$ et $x_n \geq 1$.

Les variables étant indépendantes (cf. l'exercice de la page 42), X_n converge presque sûrement vers 0 si et seulement si, pour tout $\varepsilon > 0$, $\sum_{n \geq 0} \mathbb{P}(|X_n| > \varepsilon) < +\infty$ c'est à dire si et seulement si $\sum_{n \geq 0} p_n < +\infty$ ($x_n \geq 1$).

$(X_n)_{n \in \mathbf{N}}$ converge dans L^p vers 0 si et seulement si $\mathbb{E} [|X_n|^p] = x_n^p p_n \rightarrow 0$.

- $p_n = 2^{-n}$, $x_n = 2^n$: X_n converge vers 0 presque sûrement mais pas dans L^1 ;
- $p_n = n^{-1}$, $x_n = 1$: X_n converge vers 0 dans L^p pour tout $p \geq 1$ mais ne converge pas vers 0 presque sûrement ;
- $p_n = n^{-2}$, $x_n = n$: X_n converge vers 0 dans L^1 mais pas dans L^2 .

Convergence en probabilité.

Nous en venons à présent à une notion nouvelle de convergence : la convergence en probabilité. Rappelons encore une fois que les variables X_n et X sont à valeurs dans \mathbf{R}^d .

Définition. La suite $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité si :

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

Pour nous familiariser avec cette notion commençons par un résultat qui nous ramène au cas de variables réelles.

Proposition 7. $X_n = (X_n^1, \dots, X_n^d)$ converge en probabilité vers $X = (X^1, \dots, X^d)$ si et seulement si, pour tout $i = 1, \dots, d$, X_n^i converge en probabilité vers X^i .

Démonstration. La condition est nécessaire puisque, pour tout $i \leq d$ et tout $\varepsilon > 0$,

$$\mathbb{P}(|X_n^i - X^i| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon).$$

Pour établir la suffisance, observons que $\{|X_n - X| > \varepsilon\} \subset \bigcup_{i \leq d} \{|X_n^i - X^i| > \varepsilon/\sqrt{d}\}$. En effet, si, pour tout $i \leq d$, $|X_n^i(\omega) - X^i(\omega)| \leq \varepsilon/\sqrt{d}$ alors

$$|X_n(\omega) - X(\omega)| = \left(\sum_{i \leq d} |X_n^i(\omega) - X^i(\omega)|^2 \right)^{\frac{1}{2}} \leq \varepsilon ;$$

il suffit alors de passer au complémentaire. Par suite, on a

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \sum_{i \leq d} \mathbb{P}(|X_n^i - X^i| > \varepsilon/\sqrt{d}).$$

Le membre de droite tend vers 0 puisque c'est une somme finie de termes tendant vers 0. \square

Remarque. Si $(X_n)_{n \in \mathbf{N}}$ converge en probabilité vers X et vers Y , alors $X = Y$ presque sûrement. En effet, si $\varepsilon > 0$, utilisant le même argument que dans la démonstration précédente, on a, pour tout $n \geq 0$,

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|X_n - Y| > \varepsilon/2).$$

Comme le majorant de cette inégalité tend vers 0 si $n \rightarrow +\infty$, on en déduit que, pour tout $\varepsilon > 0$, $\mathbb{P}(|X - Y| > \varepsilon) = 0$ ce qui donne l'égalité presque sûre de X et Y .

Exercice. On suppose que $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité et que, pour tout $n \geq 0$, $|X_n| \leq Y$ presque sûrement. Montrer que $|X| \leq Y$ presque sûrement.

Indic. $\mathbb{P}(|X| > Y + \varepsilon) \leq \mathbb{P}(|X_n| > Y) + \mathbb{P}(|X_n - X| > \varepsilon)$.

Proposition 8. Soit $f : \mathbf{R}^d \rightarrow \mathbf{R}^q$ une application continue. Si $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité alors $(f(X_n))_{n \in \mathbf{N}}$ converge vers $f(X)$ en probabilité.

Démonstration. Nous allons utiliser l'uniforme continuité de f sur les compacts. Fixons $\varepsilon > 0$ et $a > 0$; il existe $\eta_{\varepsilon, a} > 0$ tel que

$$|x| \leq a \text{ et } |x - y| \leq \eta_{\varepsilon, a} \implies |f(x) - f(y)| \leq \varepsilon.$$

En particulier, $\{|X| \leq a\} \cap \{|X_n - X| \leq \eta_{\varepsilon,a}\} \subset \{|f(X_n) - f(X)| \leq \varepsilon\}$ et par passage au complémentaire

$$\{|f(X_n) - f(X)| > \varepsilon\} \subset \{|X| > a\} \cup \{|X_n - X| > \eta_{\varepsilon,a}\}.$$

Par conséquent, on a l'inégalité

$$\mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a) + \mathbb{P}(|X - X_n| > \eta_{\varepsilon,a}),$$

qui conduit à

$$\limsup \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a) + \limsup \mathbb{P}(|X - X_n| > \eta_{\varepsilon,a}).$$

Comme X_n converge vers X en probabilité et $\eta_{\varepsilon,a} > 0$, $\limsup \mathbb{P}(|X - X_n| > \eta_{\varepsilon,a}) = 0$ et finalement,

$$\forall a > 0, \quad \limsup \mathbb{P}(|f(X_n) - f(X)| > \varepsilon) \leq \mathbb{P}(|X| > a).$$

On conclut en remarquant que, X étant à valeurs dans \mathbf{R}^d , $\lim_{a \rightarrow +\infty} \mathbb{P}(|X| > a) = 0$. \square

Donnons des exemples d'application de ce résultat. On considère deux suites X_n et Y_n convergeant en probabilité vers X et Y et α_n une suite de réels convergeant vers α . Alors $X_n Y_n$ converge vers XY en probabilité et $X_n + \alpha_n Y_n$ converge en probabilité vers $X + \alpha Y$.

Proposition 9. *Si $(X_n)_{n \in \mathbf{N}}$ converge vers X presque sûrement ou dans L^1 alors $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité.*

Démonstration. Supposons tout d'abord que X_n converge vers X presque sûrement. Soit $\varepsilon > 0$. On a $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{E}[\mathbf{1}_{] \varepsilon, +\infty[}(|X_n - X|)]$; or $\mathbf{1}_{] \varepsilon, +\infty[}(|X_n - X|)$ converge vers 0 presque sûrement par hypothèse et $0 \leq \mathbf{1}_{] \varepsilon, +\infty[}(|X_n - X|) \leq 1$. Le résultat découle du théorème de convergence dominée.

Si X_n converge vers X en moyenne, l'inégalité de Markov

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\varepsilon}$$

donne le résultat. \square

Exemple. Reprenons l'exemple de la page 43. Pour $\varepsilon < 1$, $\mathbb{P}(|X_n| > \varepsilon) = p_n$. X_n converge vers 0 en probabilité si et seulement si $p_n \rightarrow 0$. Pour $p_n = n^{-1}$ et $x_n = n$, X_n converge en probabilité vers 0 sans converger presque sûrement ou en moyenne.

Comme nous venons de le voir la convergence en probabilité n'implique pas la convergence presque sûre. Toutefois, on peut montrer le résultat suivant.

Proposition 10. *Si $(X_n)_{n \in \mathbf{N}}$ converge en probabilité vers X , il existe une sous-suite $(X_{n_r})_{r \in \mathbf{N}}$ qui converge vers X presque sûrement.*

Démonstration. Pour tout $r \in \mathbf{N}$, $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > 2^{-r-1}) = 0$. Il existe donc un entier n_r tel que

$$\forall n \geq n_r, \quad \mathbb{P}(|X_n - X| > 2^{-r-1}) \leq 2^{-r-1}.$$

On peut supposer la suite n_r strictement croissante puisque si n_r convient $n_r + 1$ convient également. On a alors, pour tout $r \in \mathbf{N}$,

$$\mathbb{P}(|X_{n_{r+1}} - X_{n_r}| > 2^{-r}) \leq \mathbb{P}(|X_{n_{r+1}} - X| > 2^{-r-1}) + \mathbb{P}(|X_{n_r} - X| > 2^{-r-1}) \leq 2^{-r}.$$

Le Lemme 3 montre que la suite $(X_{n_r})_{r \in \mathbf{N}}$ converge presque sûrement. Appelons Y la limite. X_{n_r} converge vers Y presque sûrement donc en probabilité d'après la Proposition 9. Or X_{n_r} converge vers X en probabilité puisque c'est une sous-suite de X_n . Donc X et Y sont égales presque sûrement – cf. Remarque p. 44 – et X_{n_r} converge vers X presque sûrement. \square

Remarque. En fait, X_n converge vers X en probabilité si et seulement si de toute sous-suite X_{n_r} on peut extraire une sous-suite qui converge vers X presque sûrement. Justifions-le.

Si X_n converge vers X en probabilité, toute sous-suite X_{n_r} converge vers X en probabilité. On donc extraire de la suite X_{n_r} une sous-suite qui converge vers X presque sûrement : c'est le résultat précédent.

Réciproquement, supposons que X_n ne converge pas vers X en probabilité. Il existe $\varepsilon > 0$ et $\alpha > 0$ tel que

$$\forall n \in \mathbf{N}, \quad \exists p \geq n, \quad \mathbb{P}(|X_p - X| > \varepsilon) \geq \alpha.$$

On construit ainsi une suite strictement croissante d'entiers n' telle que $\mathbb{P}(|X_{n'} - X| > \varepsilon) \geq \alpha$. On peut extraire de la sous-suite $X_{n'}$ une sous-suite $X_{n'_r}$ convergeant vers X presque sûrement et donc en probabilité. Ceci est incompatible avec le fait que

$$\forall r \in \mathbf{N}, \quad \mathbb{P}(|X_{n'_r} - X| > \varepsilon) \geq \alpha.$$

Critère de convergence. Il est parfois bien utile de disposer de critères de type Cauchy pour établir la convergence d'une suite de variables aléatoires.

Proposition 11. Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires dans \mathbf{R}^d .

1. $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement si et seulement si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}\left(\sup_{r \geq 0} |X_{n+r} - X_n| > \varepsilon\right) = 0;$$

2. $(X_n)_{n \in \mathbf{N}}$ converge en probabilité si et seulement si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \sup_{r \geq 0} \mathbb{P}(|X_{n+r} - X_n| > \varepsilon) = 0;$$

3. $(X_n)_{n \in \mathbf{N}} \subset L^p$ converge en moyenne d'ordre p si et seulement si

$$\lim_{n \rightarrow +\infty} \sup_{r \geq 0} \mathbb{E}[|X_{n+r} - X_n|^p] = 0.$$

Démonstration. Le troisième point de la proposition n'est rien d'autre que la complétude de L^p . Je vous renvoie donc à votre cours d'intégration.

Passons à la convergence presque sûre. Supposons dans un premier temps que $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement. \mathbf{R}^d étant complet la suite $(X_n)_{n \in \mathbf{N}}$ est de Cauchy presque sûrement : ceci signifie que $\sup_{r \geq 0} |X_{n+r} - X_n|$ converge vers 0 presque sûrement lorsque $n \rightarrow +\infty$. Cette variable aléatoire converge donc aussi vers 0 en probabilité et par suite

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}\left(\sup_{r \geq 0} |X_{n+r} - X_n| > \varepsilon\right) = 0.$$

Réciproquement, supposons cette dernière condition satisfaite. Notons V_n la variable aléatoire $V_n = \sup_{p \geq n, q \geq n} |X_p - X_q|$. La suite $(V_n(\omega))_{n \in \mathbf{N}}$ est décroissante et minorée par 0 (pour tout $\omega \in \Omega$). V_n converge donc presque sûrement vers une v.a. disons V . On a d'autre part, d'après l'inégalité triangulaire, $V_n \leq 2 \sup_{r \geq 0} |X_{n+r} - X_n|$: la suite V_n converge donc vers 0 en probabilité puisque

$$\mathbb{P}(V_n > \varepsilon) \leq \mathbb{P}\left(2 \sup_{r \geq 0} |X_{n+r} - X_n| > \varepsilon\right) = \mathbb{P}\left(\sup_{r \geq 0} |X_{n+r} - X_n| > \varepsilon/2\right).$$

V_n converge vers V presque sûrement et vers 0 en probabilité ; donc $V = 0$ presque sûrement. Ceci montre que $(X_n)_{n \in \mathbf{N}}$ est presque sûrement de Cauchy et donc presque sûrement convergente puisque \mathbf{R}^d est complet.

Finissons par la convergence en probabilité. Si $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité, on a, pour tout $r \in \mathbf{N}$, puisque $|X_{n+r} - X_n| \leq |X_{n+r} - X| + |X_n - X|$,

$$\mathbb{P}(|X_{n+r} - X_n| > \varepsilon) \leq \mathbb{P}(|X_{n+r} - X| > \varepsilon/2) + \mathbb{P}(|X_n - X| > \varepsilon/2) \leq 2 \sup_{k \geq n} \mathbb{P}(|X_k - X| > \varepsilon/2) ;$$

le majorant tend vers 0 d'où le critère. Si à présent

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \sup_{r \geq 0} \mathbb{P}(|X_{n+r} - X_n| > \varepsilon) = 0,$$

on construit une suite strictement croissante d'entiers n_r telle que

$$\sup_{k \geq 0} \mathbb{P}(|X_{n_r+k} - X_{n_r}| > 2^{-r}) \leq 2^{-r}.$$

En particulier, $\mathbb{P}(|X_{n_{r+1}} - X_{n_r}| > 2^{-r}) \leq 2^{-r}$. D'après le Lemme 3, la sous-suite $(X_{n_r})_{r \in \mathbf{N}}$ converge presque sûrement et donc en probabilité disons vers X . On a alors

$$\begin{aligned} \mathbb{P}(|X_k - X| > \varepsilon) &\leq \mathbb{P}(|X_{n_k} - X_k| > \varepsilon/2) + \mathbb{P}(|X_{n_k} - X| > \varepsilon/2) \\ &\leq \sup_{r \geq 0} \mathbb{P}(|X_{k+r} - X_k| > \varepsilon/2) + \mathbb{P}(|X_{n_k} - X| > \varepsilon/2). \end{aligned}$$

Pour conclure, il suffit de remarquer que les deux termes de la majoration tendent vers 0 lorsque $k \rightarrow +\infty$. \square

Remarque. Cette proposition montre en particulier que la convergence en probabilité ne dépend que des classes d'équivalence des variables aléatoires. En fait, sur L^0 , si on définit

$$\delta(X, Y) = \inf \{c \geq 0, \mathbb{P}(|X - Y| > c) \leq c\},$$

on obtient une distance et (L^0, δ) est un espace métrique complet. De plus, la convergence pour la métrique δ est la convergence en probabilité.

2. Convergence étroite et convergence en loi.

Nous passons à présent à une notion de convergence d'un autre type : la convergence en loi. En fait, la convergence en loi d'une suite de variables aléatoires $(X_n)_{n \in \mathbf{N}}$ est la convergence étroite des mesures de probabilité \mathbb{P}_{X_n} . Nous commençons par ce point.

2.1. Convergence étroite.

On considère, dans tout ce paragraphe, une suite $(\mu_n)_{n \in \mathbf{N}}$ de probabilités sur \mathbf{R}^d et μ une probabilité sur \mathbf{R}^d . Rappelons que \mathcal{C}_b désigne l'ensemble des fonctions $f : \mathbf{R}^d \rightarrow \mathbf{R}$ continues et bornées, \mathcal{C}_c l'ensemble des fonctions continues à support compact, \mathcal{C}_0 celui des fonctions continues tendant vers 0 si $|x| \rightarrow +\infty$.

Définition. On dit que $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ si

$$\forall f \in \mathcal{C}_b(\mathbf{R}^d), \quad \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f(x) \mu_n(dx) = \int_{\mathbf{R}^d} f(x) \mu(dx).$$

Commençons par une proposition très similaire à la Proposition 10 du Chapitre I.

Proposition 12. Soit $\mathcal{H} \subset \mathcal{C}_b(\mathbf{R}^d)$ tel que $\mathcal{C}_c(\mathbf{R}^d) \subset \overline{\mathcal{H}}$ (l'adhérence pour $\|\cdot\|_\infty$). Une suite de probabilités $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers une probabilité μ si et seulement si

$$\forall f \in \mathcal{H}, \quad \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f(x) \mu_n(dx) = \int_{\mathbf{R}^d} f(x) \mu(dx).$$

Démonstration. Montrons d'abord le résultat dans le cas $\mathcal{H} = \mathcal{C}_c$. On utilise un argument de troncature. Considérons, pour $r > 0$, la fonction $\theta_r : \mathbf{R}_+ \rightarrow [0, 1]$ suivante : $\theta_r(x) = 1$ si $0 \leq x \leq r$, $\theta_r(x) = 0$ si $x \geq 2r$, θ_r est affine entre r et $2r$. Voir le graphe 1.

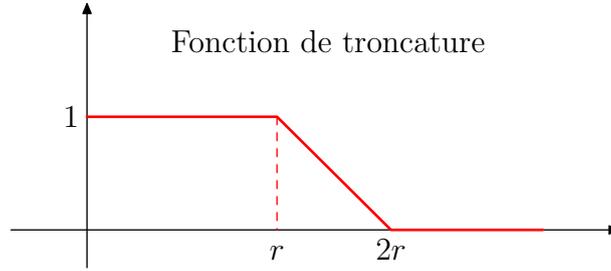


FIGURE 1 – Graphe de la fonction θ_r .

Soit f une fonction continue bornée. Si ν est une probabilité sur \mathbf{R}^d , on a

$$\int_{\mathbf{R}^d} f(x) \nu(dx) = \int_{\mathbf{R}^d} f(x) \theta_r(|x|) \nu(dx) + \int_{\mathbf{R}^d} f(x) [1 - \theta_r(|x|)] \nu(dx),$$

et

$$\left| \int_{\mathbf{R}^d} f(x) [1 - \theta_r(|x|)] \nu(dx) \right| \leq \|f\|_\infty \left(1 - \int_{\mathbf{R}^d} \theta_r(|x|) \nu(dx) \right).$$

Par suite, on obtient l'inégalité

$$\begin{aligned} \left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| &\leq \left| \int_{\mathbf{R}^d} f(x) \theta_r(|x|) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \theta_r(|x|) \mu(dx) \right| \\ &\quad + \|f\|_\infty \left(2 - \int_{\mathbf{R}^d} \theta_r(|x|) \mu_n(dx) - \int_{\mathbf{R}^d} \theta_r(|x|) \mu(dx) \right). \end{aligned}$$

Pour tout $r > 0$, les fonctions $x \mapsto \theta_r(|x|)$ et $x \mapsto f(x) \theta_r(|x|)$ sont continues à support compact de sorte que

$$\int_{\mathbf{R}^d} f(x) \theta_r(|x|) \mu_n(dx) \longrightarrow \int_{\mathbf{R}^d} f(x) \theta_r(|x|) \mu(dx), \quad \int_{\mathbf{R}^d} \theta_r(|x|) \mu_n(dx) \longrightarrow \int_{\mathbf{R}^d} \theta_r(|x|) \mu(dx).$$

Par conséquent, pour tout $r > 0$,

$$\limsup \left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq 2\|f\|_\infty \left(1 - \int_{\mathbf{R}^d} \theta_r(|x|) \mu(dx) \right).$$

Pour conclure remarquons que, par convergence dominée $-\lim_{r \rightarrow +\infty} \theta_r(|x|) = 1$ pour tout $x \in \mathbf{R}^d$ et $0 \leq \theta_r(|x|) \leq 1$,

$$\lim_{r \rightarrow +\infty} \int_{\mathbf{R}^d} \theta_r(|x|) \mu(dx) = \mu(\mathbf{R}^d) = 1.$$

Supposons à présent que le passage à la limite puisse se faire pour les fonctions appartenant \mathcal{H} où $\mathcal{C}_c \subset \overline{\mathcal{H}}$. Soit $f \in \mathcal{C}_c$. Pour toute fonction $h \in \mathcal{H}$, on a – cf. preuve de la Proposition 10 du Chapitre I –

$$\left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq \left| \int_{\mathbf{R}^d} h(x) \mu_n(dx) - \int_{\mathbf{R}^d} h(x) \mu(dx) \right| + 2\|f - h\|_\infty,$$

et donc, pour toute $h \in \mathcal{H}$,

$$\limsup \left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq 2\|f - h\|_\infty.$$

Comme $\mathcal{C}_c \subset \overline{\mathcal{H}}$, $\inf_{h \in \mathcal{H}} \|f - h\|_\infty = \text{dist}(f, \mathcal{H}) = 0$; cette dernière observation conduit au résultat. \square

La continuité de la fonction f dans la définition est essentielle. En particulier, si $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ , il est faux en général que $\mu_n(B)$ converge vers $\mu(B)$ lorsque B est un borélien. Par exemple, $\delta_{1/n}$ converge vers δ_0 étroitement et $\delta_{1/n}(\{0\}) = 0$ pour tout n tandis que $\delta_0(\{0\}) = 1$. Toutefois, on peut montrer le résultat suivant.

Théorème 13. Soient $(\mu_n)_{n \in \mathbf{N}}$ et μ des probabilités sur \mathbf{R}^d . On a équivalence entre

1. $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ ;
2. pour tout fermé F , $\limsup \mu_n(F) \leq \mu(F)$;
3. pour tout ouvert G , $\mu(G) \leq \liminf \mu_n(G)$;
4. pour tout borélien B tel que $\mu(\overline{B} \setminus B^\circ) = 0$, $\lim_{n \rightarrow +\infty} \mu_n(B) = \mu(B)$;
5. pour toute fonction f bornée telle que $\mu(D_f) = 0$ – D_f désigne l'ensemble des points de discontinuités de f –

$$\lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f(x) \mu_n(dx) = \int_{\mathbf{R}^d} f(x) \mu(dx).$$

Démonstration. • Supposons que $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ et prenons un fermé F de \mathbf{R}^d . La fonction $f_k(x) = (1 + d(x, F))^{-k}$ est continue et bornée pour tout $k \in \mathbf{N}^*$; de plus $f_k(x)$ décroît vers $\mathbf{1}_F(x)$. On a, pour tout $k \geq 0$,

$$\limsup \mu_n(F) \leq \limsup \int_{\mathbf{R}^d} f_k(x) \mu_n(dx) = \int_{\mathbf{R}^d} f_k(x) \mu(dx).$$

Par convergence dominée, $\lim_{k \rightarrow +\infty} \int f_k(x) \mu(dx) = \mu(F)$ et donc $\limsup \mu_n(F) \leq \mu(F)$.

- Les points 2 et 3 sont équivalents par passage au complémentaire.

- Les points 2 et 3 entraînent la 4^e assertion. En effet, on a $B^\circ \subset B \subset \overline{B}$. Par conséquent,

$$\mu(B^\circ) \leq \liminf \mu_n(B^\circ) \leq \liminf \mu_n(B) \leq \limsup \mu_n(B) \leq \limsup \mu_n(\overline{B}) \leq \mu(\overline{B}).$$

Si $\mu(\overline{B} \setminus B^\circ) = 0$, $\mu(B) = \mu(B^\circ) = \mu(\overline{B})$ et $\mu(B) = \liminf \mu_n(B) = \limsup \mu_n(B)$.

• La partie plus délicate consiste à montrer que cette dernière propriété implique le point 5. Soit $f : \mathbf{R}^d \rightarrow \mathbf{R}$ une fonction bornée. L'ensemble $\{t \in \mathbf{R} : \mu(\{x \in \mathbf{R}^d : f(x) = t\}) > 0\}$ est un ensemble au plus dénombrable : c'est l'ensemble des points de discontinuités de la fonction de répartition de la variable aléatoire f définie sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d), \mu)$. Son complémentaire D est donc dense. Soit $c > 0$ tel que $\|f\|_\infty \leq c$. Soit $\varepsilon > 0$; il existe un nombre fini de points $t_1 < t_2 < t_3 < \dots < t_r$ de D tel que $t_1 < -c$, $t_r > c$ et $\max_{i \leq r} |t_i - t_{i-1}| \leq \varepsilon$. Considérons la fonction $g(x) = \sum_{i=1}^{r-1} t_i \mathbf{1}_{[t_i, t_{i+1}[}(f(x))$. On a

$$\forall x \in \mathbf{R}^d, \quad |f(x) - g(x)| \leq \max_{i \leq r} |t_i - t_{i-1}| \leq \varepsilon,$$

de sorte que

$$\left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq \left| \int_{\mathbf{R}^d} g(x) \mu_n(dx) - \int_{\mathbf{R}^d} g(x) \mu(dx) \right| + 2\varepsilon.$$

D'autre part, pour tout $n \geq 0$,

$$\int_{\mathbf{R}^d} g(x) \mu_n(dx) = \sum_{i=1}^{r-1} t_i \mu_n(\{x \in \mathbf{R}^d : f(x) \in [t_i, t_{i+1}[}\}),$$

et la même formule est valable pour μ . Il reste donc à montrer pour conclure que, pour tout $i \leq r-1$, la frontière de $B = \{x \in \mathbf{R}^d : f(x) \in [t_i, t_{i+1}[}\}$ est de μ -mesure nulle. Remarquons pour cela que

$$\{x \in D_f^c : t_i < f(x) < t_{i+1}\} \subset B^\circ, \quad \overline{B} \subset \{x \in D_f^c : t_i \leq f(x) \leq t_{i+1}\} \cup D_f;$$

par suite, $\overline{B} \setminus B^\circ \subset \cup_{i=1}^r \{x \in \mathbf{R}^d : f(x) = t_i\} \cup D_f$. Comme les t_i sont dans D^c , le résultat s'en suit.

- Le fait que cette dernière propriété implique la convergence étroite est immédiat. \square

Nous passons à résultat reliant la convergence étroite à celle des transformées de Fourier des probabilités. Nous admettrons partiellement ce théorème dû à Paul Lévy.

Théorème 14 (Paul Lévy). *Soit $(\mu_n)_{n \in \mathbf{N}}$ une suite de probabilités sur \mathbf{R}^d .*

Si la suite de fonctions $(\widehat{\mu}_n)_{n \in \mathbf{N}}$ converge simplement vers une fonction φ continue au point 0, il existe une probabilité μ sur \mathbf{R}^d telle que $\varphi = \widehat{\mu}$ et $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ . De plus, la convergence de $\widehat{\mu}_n$ vers $\widehat{\mu}$ est uniforme sur tout compact de \mathbf{R}^d .

Nous nous contenterons d'établir le corollaire suivant très utilisé en pratique.

Corollaire 15. *Soient $(\mu_n)_{n \in \mathbf{N}}$ et μ des probabilités sur \mathbf{R}^d . $(\mu_n)_{n \in \mathbf{N}}$ converge étroitement vers μ si et seulement si $(\widehat{\mu}_n)_{n \in \mathbf{N}}$ converge simplement vers $\widehat{\mu}$.*

Il suffit donc d'établir que $\lim_{n \rightarrow +\infty} \widehat{\mu}_n(t) = \widehat{\mu}(t)$ pour montrer la convergence étroite de μ_n vers μ . Nous verrons des exemples un peu plus loin.

Démonstration. Comme $x \mapsto e^{it \cdot x}$ est continue bornée, la convergence étroite implique la convergence simple des transformées de Fourier.

Supposons la convergence simple de $\widehat{\mu}_n$ vers $\widehat{\mu}$. Rappelons que d'après la Proposition 12 du Chapitre I, on a, pour toute fonction $f \in \mathcal{C}_c$, et tout $\varepsilon > 0$,

$$\left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq \left| \int_{\mathbf{R}^d} f(x) U_n^\varepsilon(x) dx \right| + 2\omega_f(\varepsilon\sqrt{d}), \quad (1)$$

où $\omega_f(\eta) = \sup_{|x-y| \leq \eta} |f(x) - f(y)|$ et

$$U_n^\varepsilon(x) = \int_{\mathbf{R}^d} e^{-it \cdot x} H(\varepsilon t) (\widehat{\mu}_n(t) - \widehat{\mu}(t)) dx,$$

avec $H(t) = h(t_1) \dots h(t_d)$ et $h(s) = \frac{1}{\pi} \frac{1 - \cos s}{s^2}$, $h(0) = \frac{1}{2\pi}$. En particulier, $t \mapsto H(\varepsilon t)$ est intégrable sur \mathbf{R}^d . Montrons que, ε étant fixé, $U_n^\varepsilon(x)$ converge vers 0, lorsque $n \rightarrow +\infty$, pour tout $x \in \mathbf{R}^d$. Fixons x ; on a

$$\forall t \in \mathbf{R}^d, \quad \left| e^{-it \cdot x} H(\varepsilon t) (\widehat{\mu}_n(t) - \widehat{\mu}(t)) \right| \leq |H(\varepsilon t)| \min(2, |\widehat{\mu}_n(t) - \widehat{\mu}(t)|).$$

Cette inégalité permet d'appliquer le théorème de convergence dominée, pour obtenir

$$\forall x \in \mathbf{R}^d, \quad U_n^\varepsilon(x) \longrightarrow 0, \quad \|U_n^\varepsilon\|_\infty \leq 2 \|H(\varepsilon \cdot)\|_1.$$

Puisque f est continue et à support compact, $f \|H(\varepsilon \cdot)\|_1$ est intégrable; on peut à nouveau appliquer le théorème de convergence dominée, pour obtenir

$$\lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f(x) U_n^\varepsilon(x) dx = 0.$$

Par conséquent, revenant à la majoration (1), on a

$$\forall \varepsilon > 0, \quad \limsup \left| \int_{\mathbf{R}^d} f(x) \mu_n(dx) - \int_{\mathbf{R}^d} f(x) \mu(dx) \right| \leq 2\omega_f(\varepsilon\sqrt{d}).$$

f étant uniformément continue, $\lim_{\varepsilon \rightarrow 0^+} \omega_f(\varepsilon\sqrt{d}) = 0$ ce qui conclut la démonstration via la Proposition 12. \square

2.2. Convergence en loi.

Soient $(X_n)_{n \in \mathbf{N}}$ et X des variables aléatoires dans \mathbf{R}^d .

Définition. La suite $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi si la suite de probabilités $(\mathbb{P}_{X_n})_{n \in \mathbf{N}}$ converge vers \mathbb{P}_X étroitement.

La signification de cette définition est la suivante : $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi si, pour toute fonction $f : \mathbf{R}^d \rightarrow \mathbf{R}$ continue et bornée,

$$\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

D'après les résultats du paragraphe précédent, X_n converge vers X en loi si et seulement si l'égalité précédente a lieu pour toute fonction f appartenant à un ensemble \mathcal{H} tel que $\mathcal{H} \subset \mathcal{C}_b$ et $\mathcal{C}_c \subset \overline{\mathcal{H}}$. Le théorème de Paul Lévy nous dit que la convergence en loi est équivalente à la convergence simple des fonctions caractéristiques soit :

$$X_n \longrightarrow X \text{ en loi} \iff \forall t \in \mathbf{R}^d, \quad \varphi_{X_n}(t) = \mathbb{E} \left[e^{it \cdot X_n} \right] \longrightarrow \varphi_X(t) = \mathbb{E} \left[e^{it \cdot X} \right].$$

Remarque. On déduit de cette dernière caractérisation que le vecteur X_n converge en loi vers X si et seulement si $t \cdot X_n$ converge en loi vers $t \cdot X$ pour tout $t \in \mathbf{R}^d$. Ceci résulte du fait que, pour tout vecteur aléatoire X et tout réel c , $\varphi_{t \cdot X}(c) = \varphi_X(ct)$. En particulier, si le vecteur X_n converge en loi vers X , ses marginales convergent en loi vers celles de X . Attention, la réciproque est fautive : voir l'exemple de la page 53.

Remarque. Il faut bien noter que la convergence en loi n'est pas une notion portant sur la convergence des $X_n(\omega)$. Par exemple, les v.a. X_n peuvent être définies sur des espaces probabilisés différents.

Nous relierons à présent la convergence en loi à celle des fonctions de répartition.

Proposition 16. *Soient $(X_n)_{n \in \mathbf{N}}$ et X des variables aléatoires réelles. $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi si et seulement si, pour tout $t \in \mathbf{R}$ où F_X est continue, $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$.*

Démonstration. Supposons dans un premier temps que X_n converge vers X en loi. Nous allons utiliser un argument d'approximation en substituant à la fonction $\mathbf{1}_{]-\infty, t]}(x)$ une fonction continue. Si $s < u$ sont deux réels, on note $\psi_{s,u}$ la fonction réelle valant 1 lorsque $x \leq s$, 0 lorsque $x \geq u$ et affine entre s et u : voir le graphe 2. Fixons $t \in \mathbf{R}$, et prenons $s < t < u$.

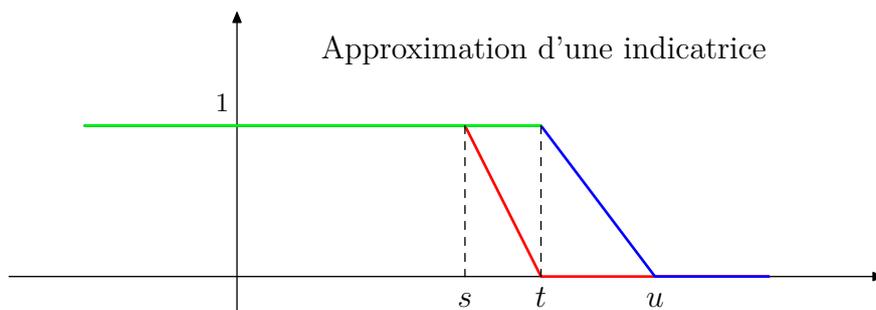


FIGURE 2 – L'encadrement $\psi_{s,t} \leq \mathbf{1}_{]-\infty, t]} \leq \psi_{t,u}$.

On a, pour tout réel x , $\psi_{s,t}(x) \leq \mathbf{1}_{]-\infty, t]}(x) \leq \psi_{t,u}(x)$, de sorte que

$$\mathbb{E}[\psi_{s,t}(X_n)] \leq F_{X_n}(t) = \mathbb{E}[\mathbf{1}_{]-\infty, t]}(X_n)] \leq \mathbb{E}[\psi_{t,u}(X_n)].$$

Les fonctions $\psi_{s,t}$ et $\psi_{t,u}$ étant continues et bornées, on a

$$\mathbb{E}[\psi_{s,t}(X)] = \lim_{n \rightarrow +\infty} \mathbb{E}[\psi_{s,t}(X_n)] = \liminf \mathbb{E}[\psi_{s,t}(X_n)], \quad \mathbb{E}[\psi_{t,u}(X)] = \limsup \mathbb{E}[\psi_{t,u}(X_n)],$$

et par suite, comme $\mathbf{1}_{]-\infty, s]} \leq \psi_{s,t}$ et $\psi_{t,u} \leq \mathbf{1}_{]-\infty, u]}$,

$$F_X(s) \leq \mathbb{E}[\psi_{s,t}(X)] \leq \liminf F_{X_n}(t) \leq \limsup F_{X_n}(t) \leq \mathbb{E}[\psi_{t,u}(X)] \leq F_X(u).$$

Cette dernière inégalité étant valable pour tout $s < t$ et tout $u > t$, on obtient lorsque $s \rightarrow t-$ et $u \rightarrow t+$, comme F_X est continue à droite,

$$F_X(t-) \leq \liminf F_{X_n}(t) \leq \limsup F_{X_n}(t) \leq F_X(t).$$

Si donc F_X est continue au point t , $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$.

Supposons à présent que F_{X_n} converge vers F_X en tout point de continuité de F_X . Soit $f : \mathbf{R} \rightarrow \mathbf{R}$ une fonction \mathcal{C}_c^∞ . On a

$$\forall x \in \mathbf{R}, \quad f(x) = \int_{-\infty}^x f'(t) dt = \int_{\mathbf{R}} f'(t) \mathbf{1}_{]-\infty, x[}(t) dt.$$

En particulier, d'après le théorème de Fubini – la fonction f' est intégrable puisque continue à support compact –, si Y est une variable réelle,

$$\mathbb{E}[f(Y)] = \mathbb{E} \left[\int_{\mathbf{R}} f'(t) \mathbf{1}_{]-\infty, Y[}(t) dt \right] = \int_{\mathbf{R}} f'(t) \mathbb{E} \left[\mathbf{1}_{]-\infty, Y[}(t) \right] dt = \int_{\mathbf{R}} f'(t) [1 - F_Y(t)] dt.$$

L'ensemble des points de discontinuité de F_X est au plus dénombrable donc F_{X_n} converge vers F_X presque partout sur \mathbf{R} . D'après le théorème de convergence dominée – pour la domination $|f'(t) [1 - F_{X_n}(t)]| \leq |f'(t)|$ – on a

$$\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \lim_{n \rightarrow +\infty} \int_{\mathbf{R}} f'(t) [1 - F_{X_n}(t)] dt = \int_{\mathbf{R}} f'(t) [1 - F_X(t)] dt = \mathbb{E}[f(X)].$$

Les fonctions \mathcal{C}_c^∞ sont denses dans \mathcal{C}_c ; d'après la Proposition 12, on a convergence en loi de X_n vers X . \square

Exercice. Soit D une partie dense dans \mathbf{R} . On suppose que $F_{X_n}(t)$ converge vers $F_X(t)$ pour tout $t \in D$. Montrer que X_n converge vers X en loi.

Exercice. Soient $(X_n)_{n \in \mathbf{N}}$ et X des variables aléatoires à valeurs dans \mathbf{N} . Montrer que les assertions suivantes sont équivalentes :

- $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi ;
- pour tout $k \in \mathbf{N}$, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$;
- G_{X_n} converge vers G_X simplement.

Nous montrons à présent que la convergence en loi est stable pour l'image continue.

Proposition 17. *Si $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi et si $g : \mathbf{R}^d \rightarrow \mathbf{R}^q$ est une application continue, alors $(g(X_n))_{n \in \mathbf{N}}$ converge vers $g(X)$ en loi.*

Démonstration. La preuve est très simple. Soit $f : \mathbf{R}^q \rightarrow \mathbf{R}$ une fonction continue et bornée. La fonction $f \circ g : \mathbf{R}^d \rightarrow \mathbf{R}$ est alors continue et bornée. Comme X_n converge vers X en loi, nous avons

$$\lim_{n \rightarrow +\infty} \mathbb{E}[f(g(X_n))] = \mathbb{E}[f(g(X))] ;$$

c'est exactement ce qu'il fallait établir. \square

Remarque. En particulier, si le **couple** (X_n, Y_n) converge en loi vers (X, Y) , alors X_n converge en loi vers X et Y_n converge en loi vers Y ; de même, $X_n + Y_n$ et $X_n Y_n$ converge en loi respectivement vers $X + Y$ et XY .

Exemple. Une question se pose alors : que signifie la convergence en loi du couple (X_n, Y_n) ? Pour cette convergence, les variables aléatoires X_n et Y_n peuvent converger en loi sans que le couple (X_n, Y_n) converge. Voici un exemple très simple dans laquelle cette situation se produit. Prenons une variable aléatoire X de loi donnée par $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$. Prenons $X_n = X$ si n est pair, $X_n = -X$ si n est impair et $Y_n = X$ pour tout n . Comme X et $-X$ ont même loi, X_n converge vers X en loi tout comme Y_n . Par contre, le couple (X_n, Y_n) ne converge pas en loi. Si tel était le cas, $X_n Y_n$ convergerait en loi ; ceci n'est manifestement pas le cas puisque $X_n Y_n = \pm 1$ presque sûrement suivant que n est pair ou pas.

La proposition suivante montre que la convergence en loi est plus faible que tous les autres modes de convergence.

Proposition 18. *Si $(X_n)_{n \in \mathbf{N}}$ converge vers X en probabilité alors $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi.*

Démonstration. Signalons que toutes les variables sont définies sur le même espace et montrons que φ_{X_n} converge vers φ_X simplement. Nous avons, pour tout $t \in \mathbf{R}^d$,

$$|\varphi_{X_n}(t) - \varphi_X(t)| \leq \mathbb{E} \left[\left| e^{it \cdot X_n} - e^{it \cdot X} \right| \right] \leq \mathbb{E} [\min(2, |t| |X_n - X|)],$$

de sorte que, pour tout $\varepsilon > 0$, en écrivant $1 = \mathbf{1}_{[0, \varepsilon]}(|X_n - X|) + \mathbf{1}_{] \varepsilon, +\infty[}(|X_n - X|)$,

$$|\varphi_{X_n}(t) - \varphi_X(t)| \leq \varepsilon |t| \mathbb{P}(|X_n - X| \leq \varepsilon) + 2 \mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon |t| + 2 \mathbb{P}(|X_n - X| > \varepsilon);$$

par suite, pour tout $\varepsilon > 0$, $\limsup |\varphi_{X_n}(t) - \varphi_X(t)| \leq \varepsilon |t|$, ce qui donne le résultat. \square

Exemple. La réciproque est fausse. Si on reprend l'exemple de la page 53, on voit que X_n converge vers X en loi mais ne converge pas en probabilité puisque pour tout réel $\varepsilon < 1$, $\mathbb{P}(|X_n - X_{n+1}| > \varepsilon) = \mathbb{P}(2|X| > \varepsilon) = 1$.

Nous avons seulement la réciproque lorsque la limite est déterministe.

Proposition 19. *On suppose les variables $(X_n)_{n \in \mathbf{N}}$ définies sur le même espace probabilisé. Soit $c \in \mathbf{R}^d$. Si $(X_n)_{n \in \mathbf{N}}$ converge vers c en loi, alors la convergence a lieu aussi en probabilité.*

Démonstration. On se ramène au cas réel en considérant les composantes de X_n . Soit $\varepsilon > 0$.

$$\mathbb{P}(|X_n - c| > \varepsilon) = \mathbb{P}(X_n < c - \varepsilon) + \mathbb{P}(X_n > c + \varepsilon) \leq F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon).$$

Puisque X_n converge vers c en loi, d'après la Proposition 16, pour tout $t \neq c$, $F_{X_n}(t)$ converge vers $\mathbf{1}_{\mathbf{R}_+}(t - c)$. Ceci montre que $\mathbb{P}(|X_n - c| > \varepsilon)$ tend vers 0 pour tout $\varepsilon > 0$. \square

Nous avons vu que la convergence en loi des marginales n'impliquait pas la convergence en loi du vecteur. Toutefois, le fait que la limite soit constante permet de conclure.

Lemme 20 (Slutsky). *Soit $((X_n, Y_n))_{n \in \mathbf{N}}$ des vecteurs aléatoires. On suppose que $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi et $(Y_n)_{n \in \mathbf{N}}$ converge vers c en probabilité. Alors $((X_n, Y_n))_{n \in \mathbf{N}}$ converge vers (X, c) en loi.*

Démonstration. Notons que la fonction caractéristique de (X, c) est $\varphi_{(X, c)}(s, t) = \varphi_X(s)e^{it \cdot c}$. On a alors,

$$\begin{aligned} \left| \varphi_{(X_n, Y_n)}(s, t) - \varphi_X(s)e^{it \cdot c} \right| &= \left| \mathbb{E} \left[e^{is \cdot X_n} \left(e^{it \cdot Y_n} - e^{it \cdot c} \right) \right] + e^{it \cdot c} (\varphi_{X_n}(s) - \varphi_X(s)) \right| \\ &\leq \mathbb{E} \left[\left| e^{it \cdot Y_n} - e^{it \cdot c} \right| \right] + |\varphi_{X_n}(s) - \varphi_X(s)|. \end{aligned}$$

Utilisant l'inégalité utilisé dans la preuve de la Proposition 18, on a, pour tout $\varepsilon > 0$,

$$\left| \varphi_{(X_n, Y_n)}(s, t) - \varphi_X(s)e^{it \cdot c} \right| \leq \varepsilon |t| + 2 \mathbb{P}(|Y_n - c| > \varepsilon) + |\varphi_{X_n}(s) - \varphi_X(s)|,$$

et il suffit de prendre \limsup pour conclure. \square

Remarque. Rappelons que, d'après la Proposition 19, la convergence en loi de Y_n vers c implique la convergence en probabilité de Y_n vers c .

Exemple. Une des applications les plus fréquente de ce résultat est le point suivant : si X_n converge vers X en loi et $Y_n - X_n$ converge vers 0 en probabilité alors Y_n converge vers X en loi. Il suffit d'écrire $Y_n = X_n + (Y_n - X_n)$ et d'appliquer le résultat précédent ainsi que la Proposition 17 avec la fonction $g(x, d) = x + d$.

Nous finissons ce catalogue sommaire des propriétés de la convergence en loi par une généralisation du lemme de Fatou. On obtient ainsi un critère garantissant l'intégrabilité d'une limite en loi.

Proposition 21. *Supposons que $(X_n)_{n \in \mathbf{N}}$ converge vers X en loi. Alors*

$$\mathbb{E}[|X|] \leq \liminf \mathbb{E}[|X_n|].$$

Démonstration. Commençons par remarquer que d'après la Proposition 17, $|X_n|$ converge vers $|X|$ en loi. Pour tout entier k ,

$$\mathbb{E}[\min(|X|, k)] = \lim_{n \rightarrow +\infty} \mathbb{E}[\min(|X_n|, k)] = \liminf \mathbb{E}[\min(|X_n|, k)] \leq \liminf \mathbb{E}[|X_n|].$$

Par convergence monotone, on obtient

$$\mathbb{E}[|X|] = \sup_{k \geq 0} \mathbb{E}[\min(|X|, k)] \leq \liminf \mathbb{E}[|X_n|]$$

qui est l'inégalité requise. □

Chapitre IV. La Loi des Grands Nombres

L'objectif de ce chapitre est d'établir « la loi forte des grands nombres » de Kolmogorov qui est l'un des résultats fondamentaux de la théorie des probabilités. Ce résultat est très utilisé en pratique comme par exemple dans la méthode de Monte-Carlo ou bien en Statistique. De quoi s'agit-il ? On considère une suite de variables aléatoires réelles $(X_n)_{n \in \mathbf{N}^*}$ et on note, pour tout $n \geq 1$,

$$S_n = \sum_{i=1}^n X_i, \quad M_n = n^{-1} S_n, \quad S_n^* = \sup_{1 \leq i \leq n} |S_i|.$$

La loi des grands nombres consiste à étudier la limite de la suite $(M_n)_{n \in \mathbf{N}^*}$. On parle de loi faible lorsqu'on étudie la convergence en probabilité et de loi forte lorsqu'on s'intéresse à la convergence presque sûre.

Nous mènerons cette étude dans le cas d'une suite de variables aléatoires réelles $(X_n)_{n \in \mathbf{N}^*}$ indépendantes et identiquement distribuées. D'après la loi du tout ou rien de Kolmogorov – voir les exemples de la page 37 du Chapitre II – la limite, lorsqu'elle existe, est déterministe.

1. Le cadre L^2 .

Désignons par (H) l'hypothèse suivante :

(H) : $(X_n)_{n \in \mathbf{N}^*}$ est une suite de variables aléatoires réelles indépendantes et identiquement distribuées ; X_1 est de carré intégrable.

On note $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}(X_1)$.

Théorème 1 (Loi faible des grands nombres). *Sous (H), la suite $(M_n)_{n \in \mathbf{N}^*}$ converge en probabilité et dans L^2 vers $m = \mathbb{E}[X_1]$.*

Démonstration. Les variables $(X_n)_{n \in \mathbf{N}^*}$ étant indépendantes et identiquement distribuées, nous avons

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = nm, \quad \mathbb{V}(S_n) = \sum_{i=1}^n \mathbb{V}(X_i) = n\sigma^2.$$

Par conséquent, $\mathbb{E}[M_n] = m$ et $\mathbb{V}(M_n) = n^{-2}\mathbb{V}(S_n) = n^{-1}\sigma^2$. Il s'en suit que

$$\mathbb{E} \left[|M_n - m|^2 \right] = \mathbb{V}(M_n) = n^{-1}\sigma^2.$$

Ceci montre que $(M_n)_{n \in \mathbf{N}^*}$ converge vers m dans L^2 et donc en probabilité. □

Remarque. On peut établir directement la convergence en probabilité via l'inégalité de Tchebycheff puisque, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|M_n - m| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

En fait, l'hypothèse (H) suffit pour établir la loi forte des grands nombres.

Théorème 2 (Loi forte des grands nombres). *Sous (H), la suite $(M_n)_{n \in \mathbf{N}^*}$ converge vers $m = \mathbb{E}[X_1]$ presque sûrement.*

Démonstration. Supposons dans un premier temps que, pour tout $n \geq 1$, $X_n \geq 0$. Pour montrer que $(M_n)_{n \in \mathbf{N}^*}$ converge vers m , nous allons d'abord établir que la sous-suite $(M_{n^2})_{n \in \mathbf{N}^*}$ converge presque sûrement vers m puis nous passerons à la convergence de la suite toute entière. Comme déjà dit, on a, pour tout $n \geq 1$, $\mathbb{E}[M_n] = m$, $\mathbb{V}(M_n) = n^{-1}\sigma^2$. En particulier, pour tout $n \geq 1$, notant $Z_n = M_{n^2}$, $\mathbb{E}[Z_n] = m$, $\mathbb{V}(Z_n) = n^{-2}\sigma^2$. D'après le Corollaire 2 du Chapitre III, il suffit de montrer que

$$\forall \varepsilon > 0, \quad \sum_{n \geq 1} \mathbb{P}(|Z_n - m| > \varepsilon) < +\infty$$

pour établir la convergence presque sûre de $(Z_n)_{n \in \mathbf{N}^*}$ vers m . Or, d'après l'inégalité de Tchebycheff, nous avons, pour tout $\varepsilon > 0$,

$$\sum_{n \geq 1} \mathbb{P}(|Z_n - m| > \varepsilon) \leq \sum_{n \geq 1} \varepsilon^{-2} \mathbb{V}(Z_n) = \varepsilon^{-2} \sum_{n \geq 1} \sigma^2 n^{-2} < +\infty.$$

Montrons à présent que la suite $(M_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers m . Nous allons ici nous servir de la positivité des variables. Pour tout $n \geq 1$, $[\sqrt{n}] \leq \sqrt{n} < [\sqrt{n}] + 1$ et donc, notant $q_n = [\sqrt{n}]$, $q_n^2 \leq n \leq (q_n + 1)^2$. Comme les variables sont positives, on obtient l'inégalité

$$S_{q_n^2} \leq S_n \leq S_{(q_n+1)^2}, \quad \text{et donc} \quad n^{-1} S_{q_n^2} \leq M_n \leq n^{-1} S_{(q_n+1)^2}.$$

Par conséquent,

$$n^{-1} q_n^2 Z_{q_n} \leq M_n \leq n^{-1} (q_n + 1)^2 Z_{q_n+1}.$$

On a $\sqrt{n} \geq q_n \geq \sqrt{n} - 1$; donc $\lim_{n \rightarrow +\infty} q_n = +\infty$ et en fait $\lim_{n \rightarrow +\infty} q_n / \sqrt{n} = 1$. Par conséquent, comme $(Z_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers m , les deux suites $n^{-1} q_n^2 Z_{q_n}$ et $n^{-1} (q_n + 1)^2 Z_{q_n+1}$ convergent vers m presque sûrement. Il en va de même de $(M_n)_{n \in \mathbf{N}^*}$ via l'encadrement précédent.

Déduisons le cas général du cas des variables positives. Pour cela, on écrit X_n comme la différence de sa partie positive et négative : $X_n = X_n^+ - X_n^-$. Les suites $(X_n^+)_{n \in \mathbf{N}^*}$ et $(X_n^-)_{n \in \mathbf{N}^*}$ sont deux suites de variables i.i.d. de carré intégrable. D'après l'étape précédente, les suites

$$M_n^+ = n^{-1} \sum_{i=1}^n X_i^+, \quad M_n^- = n^{-1} \sum_{i=1}^n X_i^-$$

convergent presque sûrement respectivement vers $\mathbb{E}[X_1^+]$ et $\mathbb{E}[X_1^-]$. $M_n = M_n^+ - M_n^-$ converge donc presque sûrement vers $\mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X_1]$. \square

Exemple. Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi μ . Soit B un borélien. On s'intéresse au comportement de $N_n(B)$, le nombre de variables aléatoires qui sont dans B parmi X_1, \dots, X_n soit

$$N_n(B) = \#\{i : 1 \leq i \leq n, X_i \in B\} = \sum_{i=1}^n \mathbf{1}_B(X_i).$$

Les variables $\mathbf{1}_B(X_i)$ sont indépendantes et identiquement distribuées suivant la loi de Bernoulli de paramètre $\mu(B)$; $N_n(B)$ suit donc la loi binomiale de paramètres n et $\mu(B)$. Le résultat

précédent montre que la fréquence de l'événement « tomber dans B », c'est à dire le nombre de fois où on tombe dans B sur le nombre total d'épreuves, soit encore

$$\frac{N_n(B)}{n} = n^{-1} \sum_{i=1}^n \mathbf{1}_B(X_i)$$

converge presque sûrement vers $\mathbb{E}[\mathbf{1}_B(X_1)] = \mathbb{P}(X_1 \in B) = \mu(B)$.

Ceci donne en quelque sorte une justification du modèle probabiliste : l'idée naïve que l'on se fait de la probabilité d'un événement est celle de fréquence – on regarde le nombre de fois où on tombe dans B sur le nombre total de réalisations de l'expérience –, le modèle mathématique quant à lui fournit la probabilité $\mu(B)$. Si l'échantillon est grand, ces deux quantités sont très proches.

Cet exemple se poursuit au travers du résultat suivant connu sous le nom de « Théorème fondamental de la Statistique ».

Théorème 3. *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi μ sur \mathbf{R}^d . Pour $n \in \mathbf{N}^*$, on note*

$$\mu_n^\omega = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}, \quad \omega \in \Omega,$$

la mesure empirique.

Alors, presque sûrement, $(\mu_n)_{n \in \mathbf{N}^*}$ converge étroitement vers μ .

Avant d'établir ce résultat, précisons sa signification. Commençons par la mesure empirique. Pour tout borélien $B \in \mathbf{R}^d$, et tout $\omega \in \Omega$,

$$\mu_n^\omega(B) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}(B) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_B(X_i(\omega)) ;$$

c'est à dire $\mu_n(B) = n^{-1} \sum_{i=1}^n \mathbf{1}_B(X_i)$. Si $f : \mathbf{R}^d \rightarrow \mathbf{R}$ est borélienne bornée ou borélienne et positive

$$\int_{\mathbf{R}^d} f(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)).$$

Le théorème précédent affirme qu'il existe un ensemble $N \in \mathcal{F}$ avec $\mathbb{P}(N) = 0$ tel que, pour tout $\omega \in N^c$, $(\mu_n^\omega)_{n \in \mathbf{N}^*}$ converge vers μ étroitement soit

$$\forall \omega \in N^c, \quad \forall f \in \mathcal{C}_b, \quad \int_{\mathbf{R}^d} f(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)) \longrightarrow \int_{\mathbf{R}^d} f(x) \mu(dx).$$

Démonstration. \mathcal{C}_0 est séparable : il existe donc une suite de fonctions $\mathcal{H} = (h_r)_{r \in \mathbf{N}} \subset \mathcal{C}_0$ dense dans \mathcal{C}_0 pour $\|\cdot\|_\infty$. Soit $r \in \mathbf{N}$. Nous avons

$$\int_{\mathbf{R}^d} h_r(x) \mu_n^\omega(dx) = \frac{1}{n} \sum_{i=1}^n h_r(X_i(\omega)).$$

Les variables aléatoires réelles $(h_r(X_n))_{n \in \mathbf{N}^*}$ sont indépendantes et identiquement distribuées puisqu'il en est ainsi de la suite $(X_n)_{n \in \mathbf{N}}$. De plus, $h_r(X_1)$ est bornée et par conséquent de carré

intégrable. D'après le Théorème 2, avec probabilité 1, c'est à dire pour tout $\omega \in N_r^c$ avec $N_r \in \mathcal{F}$ et $\mathbb{P}(N_r) = 0$,

$$\lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} h_r(x) \mu_n^\omega(dx) = \mathbb{E}[h_r(X_1)] = \int_{\mathbf{R}^d} h_r(x) \mu(dx).$$

Notons $N = \cup_{r \in \mathbf{N}} N_r$ de sorte que $\mathbb{P}(N) \leq \sum_{r \geq 0} \mathbb{P}(N_r) = 0$. Si $\omega \in N^c = \cap_{r \geq 0} N_r^c$, on a

$$\forall r \in \mathbf{N}, \quad \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} h_r(x) \mu_n^\omega(dx) = \int_{\mathbf{R}^d} h_r(x) \mu(dx).$$

D'après la Proposition 12 du Chapitre III, on a, pour tout $\omega \in N^c$,

$$\forall f \in \mathcal{C}_b, \quad \lim_{n \rightarrow +\infty} \int_{\mathbf{R}^d} f(x) \mu_n^\omega(dx) = \int_{\mathbf{R}^d} f(x) \mu(dx);$$

c'est exactement ce qu'il fallait montrer. \square

Lorsque les variables aléatoires $(X_n)_{n \in \mathbf{N}^*}$ sont réelles, on peut montrer un résultat sur la convergence des fonctions de répartition empiriques.

Théorème 4. *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées. On note F la fonction de répartition de X_1 et, pour tout $n \geq 1$, F_n la fonction de répartition empirique soit*

$$\forall \omega \in \Omega, \quad \forall t \in \mathbf{R}, \quad F_n^\omega(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, t]}(X_i(\omega)).$$

Alors, presque sûrement, $(F_n)_{n \in \mathbf{N}^}$ converge vers F uniformément sur \mathbf{R} .*

Faisons quelques commentaires à propos de ce résultat. Tout d'abord, il signifie qu'il existe $N \in \mathcal{F}$ tel que $\mathbb{P}(N) = 0$ et

$$\forall \omega \in N^c, \quad \lim_{n \rightarrow +\infty} \sup_{t \in \mathbf{R}} |F_n^\omega(t) - F(t)| = 0.$$

D'autre part, la fonction aléatoire F_n n'est rien d'autre que la fonction de répartition de la mesure empirique (aléatoire elle aussi) μ_n : pour tout $\omega \in \Omega$, F_n^ω est la fonction de répartition de μ_n^ω .

Démonstration. Pour tout $t \in \mathbf{R}$, $F_n(t)$ converge vers $F(t)$ presque sûrement. En effet, les variables aléatoires $(\mathbf{1}_{]-\infty, t]}(X_n))_{n \in \mathbf{N}^*}$ sont indépendantes, identiquement distribuées suivant la loi de Bernoulli de paramètre $F(t) = \mathbb{P}(X_1 \leq t)$ et bornées donc de carré intégrable. D'après le Théorème 2, on a la convergence presque sûre de $F_n(t)$ vers $\mathbb{E}[\mathbf{1}_{]-\infty, t]}(X_1)] = \mathbb{P}(X_1 \leq t) = F(t)$. Il existe donc $N_t \in \mathcal{F}$ tel que $\mathbb{P}(N_t) = 0$ et, pour tout $\omega \in N_t^c$, $\lim_{n \rightarrow +\infty} F_n^\omega(t) = F(t)$. De la même manière,

$$F_n(t-) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, t[}(X_i)$$

converge presque sûrement vers $F(t-) = \mathbb{P}(X_1 < t)$ pour tout réel t fixé. Il existe $N_t' \in \mathcal{F}$ tel que $\mathbb{P}(N_t') = 0$ et, pour tout $\omega \in N_t'^c$, $\lim_{n \rightarrow +\infty} F_n^\omega(t-) = F(t-)$.

Poursuivons par un petit calcul. Soient F et F_n deux fonctions de répartitions et τ une subdivision finie : $\tau = (t_i)_{i=1, \dots, p}$, $t_1 \leq t_2 \leq \dots \leq t_p$, $p \in \mathbf{N}^*$. On note

$$\delta_F(\tau) = \max \left\{ 1 - F(t_p), (F(t_p-) - F(t_{p-1}))^+, \dots, (F(t_2-) - F(t_1))^+, F(t_1-) \right\},$$

$$R_n(\tau) = \max_{i=1, \dots, p} \left\{ (F(t_i) - F_n(t_i))^+, (F_n(t_i-) - F(t_i-))^+ \right\}.$$

On a alors

$$\|F - F_n\|_\infty = \sup_{t \in \mathbf{R}} |F(t) - F_n(t)| \leq \delta_F(\tau) + R_n(\tau). \quad (1)$$

En effet, F et F_n étant croissantes,

- Si $t < t_1$, $F(t) - F_n(t) \leq F(t_1-) \leq \delta_F(\tau)$ et

$$F_n(t) - F(t) \leq F_n(t_1-) \leq F_n(t_1-) - F(t_1-) + F(t_1-) \leq R_n(\tau) + \delta_F(\tau).$$

- Si $t \in [t_{i-1}, t_i[$, $i = 2, \dots, p$,

$$F(t) - F_n(t) \leq F(t_i-) - F_n(t_{i-1}) \leq F(t_i-) - F(t_{i-1}) + F(t_{i-1}) - F_n(t_{i-1}) \leq \delta_F(\tau) + R_n(\tau),$$

$$F_n(t) - F(t) \leq F_n(t_i-) - F(t_{i-1}) \leq F_n(t_i-) - F(t_i-) + F(t_i-) - F(t_{i-1}) \leq R_n(\tau) + \delta_F(\tau).$$

- Finalement, si $t \geq t_p$, $F_n(t) - F(t) \leq 1 - F(t_p) \leq \delta_F(\tau)$ et

$$F(t) - F_n(t) \leq 1 - F_n(t_p) \leq 1 - F(t_p) + F(t_p) - F_n(t_p) \leq \delta_F(\tau) + R_n(\tau).$$

Notons, pour tout $x \in]0, 1[$,

$$C(x) = \inf\{u \in \mathbf{R} : F(u) \geq x\}.$$

Soit $x \in]0, 1[$. Puisque $\lim_{t \rightarrow +\infty} F(t) = 1$, $A_x = \{u \in \mathbf{R} : F(u) \geq x\}$ est non vide. Comme $\lim_{t \rightarrow -\infty} F(t) = 0$, A_x est minoré. D'où l'existence du réel $C(x)$. La fonction C est clairement croissante sur $]0, 1[$. F étant croissante, A_x est une demi-droite et, comme F est de plus continue à droite, $F(C(x)) \geq x$ soit $C(x) \in A_x$. Finalement, $A_x = [C(x), +\infty[$ ou en d'autres termes

$$C(x) \leq t \iff x \leq F(t).$$

En particulier, $F(C(x)-) \leq x$ puisque pour $s < C(x)$, $F(s) < x$.

Considérons l'ensemble N suivant :

$$N = \bigcup_{q \in \mathbf{Q} \cap]0, 1[} \left(N_{C(q)} \cup N'_{C(q)} \right).$$

N est négligeable comme union dénombrable de négligeables. D'autre part, pour tout $\omega \in N^c$,

$$\forall q \in \mathbf{Q} \cap]0, 1[, \quad F_n^\omega(C(q)) \longrightarrow F(C(q)), \quad F_n^\omega(C(q)-) \longrightarrow F(C(q)-).$$

Nous allons montrer que, pour tout $\omega \in N^c$,

$$\sup_{t \in \mathbf{R}} |F_n^\omega(t) - F(t)| \longrightarrow 0.$$

Soit donc $\omega \in N^c$ fixé; nous noterons F_n pour F_n^ω . Soient $p \in \mathbf{N}^*$ fixé et pour $i = 1, \dots, p$, $t_i = C(i(p+1)^{-1})$. Remarquons que $F(t_i-) \leq i(p+1)^{-1}$ et $F(t_i) \geq i(p+1)^{-1}$ pour tout $i = 1, \dots, p$.

Par conséquent, $\delta_F(\tau) \leq (p+1)^{-1}$ et via l'inégalité (1)

$$\|F_n - F\|_\infty \leq (p+1)^{-1} + R_n(\tau),$$

On sait que, pour tout $\omega \in N^c$, à $p \in \mathbf{N}^*$ fixé, $R_n(\tau) \longrightarrow 0$ si $n \rightarrow +\infty$ et par suite et donc

$$\limsup_{n \rightarrow +\infty} \|F - F_n\|_\infty \leq (p+1)^{-1} + \limsup_{n \rightarrow +\infty} R_n(p) = (p+1)^{-1}.$$

Il suffit de faire tendre p vers $+\infty$ pour conclure. □

Remarque. On peut obtenir la convergence simple sans avoir recours à la fonction C qui s'appelle le pseudo-inverse de F . Notons D les points de discontinuités de la fonction F ; D est au plus dénombrable. Soit $N_1 = \cup_{t \in \mathbf{Q} \cup D} N_t$. N_1 est négligeable puisque c'est une union dénombrable de négligeables; nous allons montrer que

$$\forall \omega \in N_1^c, \quad \forall t \in \mathbf{R}, \quad \lim_{n \rightarrow +\infty} F_n^\omega(t) = F(t).$$

Par définition de N_1 , nous avons

$$\forall \omega \in N_1^c, \quad \forall t \in \mathbf{Q} \cup D, \quad \lim_{n \rightarrow +\infty} F_n^\omega(t) = F(t)$$

de sorte qu'il suffit de montrer que $\lim_{n \rightarrow +\infty} F_n^\omega(t) = F(t)$ pour tout $t \in D^c$ et tout $\omega \in N_1^c$.

Soient $t \in \mathbf{R}$, $(q, r) \in \mathbf{Q}^2$ tels que $q \leq t \leq r$. Puisque F_n^ω est croissante sur \mathbf{R} , on a

$$F_n^\omega(q) \leq F_n^\omega(t) \leq F_n^\omega(r),$$

de sorte que si $\omega \in N_1^c$,

$$F(q) = \liminf_{n \rightarrow +\infty} F_n^\omega(q) \leq \liminf_{n \rightarrow +\infty} F_n^\omega(t) \leq \limsup_{n \rightarrow +\infty} F_n^\omega(t) \leq \limsup_{n \rightarrow +\infty} F_n^\omega(r) = F(r).$$

Par conséquent, si $\omega \in N_1^c$, pour tout couple de rationnels (q, r) tel que $q \leq t \leq r$,

$$0 \leq \limsup F_n^\omega(t) - \liminf F_n^\omega(t) \leq F(r) - F(q).$$

Les rationnels étant denses dans \mathbf{R} , il existe une suite strictement décroissante de rationnels $(r_k)_{k \in \mathbf{N}}$ de limite t ainsi qu'une suite strictement croissante de rationnels $(q_k)_{k \in \mathbf{N}}$ de limite t . On a alors, F étant croissante,

$$0 \leq \limsup F_n^\omega(t) - \liminf F_n^\omega(t) \leq \inf_{k \in \mathbf{N}} (F(r_k) - F(q_k)) = F(t+) - F(t-) = F(t) - F(t-).$$

Cette inégalité montre que $F_n^\omega(t)$ converge vers $F(t)$ pour tout $t \in D^c$ et tout $\omega \in N_1^c$. On a donc bien convergence simple.

Lorsque la fonction F est continue, on peut alors obtenir la convergence uniforme à l'aide du théorème de Dini suivant.

Lemme 5 (Dini). Soient $(F_n)_{n \in \mathbf{N}}$ et F des fonctions de répartition. On suppose que F est continue sur \mathbf{R} .

Si $(F_n)_{n \in \mathbf{N}}$ converge simplement vers F , alors la convergence est uniforme.

Démonstration. Considérons, pour $p \in \mathbf{N}^*$, τ la subdivision régulière de $[-a, a]$ de pas $2a/p$ soit $t_i = -a + (i-1)2a/p$, $i = 1, \dots, p+1$. On a alors, F étant continue,

$$\delta_F(\tau) = \max\{1 - F(a), \omega_F(2a/p), F(-a)\},$$

où $\omega_F(r) = \sup_{|u-s| \leq r} |F(u) - F(s)|$. F est uniformément continue sur \mathbf{R} puisque c'est une fonction continue possédant des limites finies en $+\infty$ et $-\infty$; donc $\lim_{r \rightarrow 0^+} \omega_F(r) = 0$. D'autre part, F étant continue, $(F_n(t_i-) - F(t_i-))^+ = (F_n(t_i-) - F(t_i))^+ \leq (F_n(t_i) - F(t_i))^+$ et par conséquent,

$$R_n(\tau) \leq \max\{|F_n(t_i) - F(t_i)|, i = 1, \dots, p+1\}.$$

La majoration (1) donne

$$\|F_n - F\|_\infty \leq \max(1 - F(a), F(-a)) + \max_{i=0, \dots, p} |F_n(t_i) - F(t_i)| + \omega_F(2a/p).$$

Pour tous $a > 0$ et $p \in \mathbf{N}^*$, la convergence simple de F_n vers F entraîne

$$\lim_{n \rightarrow +\infty} \max_{i=0, \dots, p} |F_n(t_i) - F(t_i)| = 0$$

puisqu'il y a $p + 1$ points. Donc,

$$\limsup_{n \rightarrow +\infty} \|F_n - F\|_\infty \leq \max(1 - F(a), F(-a)) + \omega_F(2a/p).$$

Il reste à faire tendre p vers $+\infty$ puis a vers $+\infty$ pour conclure. \square

2. Séries de variables indépendantes.

Dans ce paragraphe, nous nous intéressons à la convergence des séries de variables aléatoires réelles indépendantes (pas forcément de même loi). Précisons que l'expression « S_n converge » est employée pour « S_n converge dans \mathbf{R} ».

Lemme 6 (Inégalité de Lévy–Ottoviani). *Soient ξ_1, \dots, ξ_p des variables indépendantes. On note, pour $r = 1, \dots, p$, $Z_r = \sum_{1 \leq i \leq r} \xi_i$. Pour $\eta > 0$, $\delta \geq 0$,*

$$\inf_{1 \leq r < p} \mathbb{P}(|Z_p - Z_r| \leq \delta) \times \mathbb{P}\left(\sup_{1 \leq r \leq p} |Z_r| > \eta + \delta\right) \leq \mathbb{P}(|Z_p| > \eta).$$

Démonstration. Notons $\tau = \inf\{i = 1, \dots, p : |Z_i| > \eta + \delta\}$, $\inf \emptyset = +\infty$. On cherche à majorer la probabilité de l'événement $\{\sup_{1 \leq r \leq p} |Z_r| > \eta + \delta\} = \{\tau \leq p\}$ par celle de l'événement $\{|Z_p| > \eta\}$. Remarquons que $\{\tau = 1\} = \{|Z_1| > \eta + \delta\}$ et, pour $1 < r \leq p$,

$$\{\tau = r\} = \{|Z_1| \leq \eta + \delta\} \cap \dots \cap \{|Z_{r-1}| \leq \eta + \delta\} \cap \{|Z_r| > \eta + \delta\}. \quad (2)$$

On a alors $\{\tau = p\} \subset \{|Z_p| > \eta + \delta\} \subset \{|Z_p| > \eta\}$ et, pour $r = 1, \dots, p - 1$,

$$\{\tau = r\} \cap \{|Z_p - Z_r| \leq \delta\} \subset \{|Z_r| > \eta + \delta\} \cap \{|Z_p - Z_r| \leq \delta\} \subset \{|Z_p| > \eta\},$$

puisque $|Z_p| \geq |Z_r| - |Z_p - Z_r|$. Il s'en suit que

$$\mathbb{P}(|Z_p| > \eta) \geq \mathbb{P}(\tau = p) + \sum_{r=1}^{p-1} \mathbb{P}(\tau = r, |Z_p - Z_r| \leq \delta).$$

Remarquons également que les événements $\{\tau = r\}$ et $\{|Z_p - Z_r| \leq \delta\}$ sont indépendants : le premier dépend des variables ξ_1, \dots, ξ_r – cf. (2) – tandis que le second dépend des variables ξ_{r+1}, \dots, ξ_p . Par conséquent, on obtient l'inégalité

$$\mathbb{P}(|Z_p| > \eta) \geq \mathbb{P}(\tau = p) + \sum_{r=1}^{p-1} \mathbb{P}(\tau = r) \mathbb{P}(|Z_p - Z_r| \leq \delta) \geq \alpha \sum_{r=1}^p \mathbb{P}(\tau = r)$$

avec $\alpha = \inf_{1 \leq r < p} \mathbb{P}(|Z_p - Z_r| \leq \delta)$. Il suffit de remarquer que $\mathbb{P}(\tau \leq p) = \sum_{r=1}^p \mathbb{P}(\tau = r)$ pour conclure. \square

La principale application est le théorème suivant dû à Paul Lévy.

Théorème 7 (Paul Lévy). *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes. Pour $n \geq 1$, $S_n = X_1 + \dots + X_n$. On a équivalence entre :*

1. $(S_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers une variable aléatoire réelle ;
2. $(S_n)_{n \in \mathbf{N}^*}$ converge en probabilité vers une variable aléatoire réelle ;
3. $(S_n)_{n \in \mathbf{N}^*}$ converge en loi vers une variable aléatoire réelle.

Démonstration. Montrons tout d'abord que si $(S_n)_{n \in \mathbf{N}^*}$ converge en probabilité alors $(S_n)_{n \in \mathbf{N}^*}$ converge presque sûrement. D'après la Proposition 11 du Chapitre III, il s'agit de montrer que, pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon \right) = 0.$$

On a, par monotonie,

$$\mathbb{P} \left(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon \right) = \lim_{p \rightarrow +\infty} \mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right).$$

Nous allons appliquer l'inégalité de Lévy–Ottoviani. Pour cela, observons que

$$\forall 1 \leq r \leq p, \quad S_{n+r} - S_n = \sum_{j=n+1}^{n+r} X_j = \sum_{i=1}^r X_{n+i} = \sum_{i=1}^r \xi_i,$$

avec $\xi_i = X_{i+n}$. Avec ces notations, $S_{n+r} - S_n$ est égal à Z_r de l'inégalité de Lévy–Ottoviani que l'on applique au couple $(\eta, \delta) = (\varepsilon/2, \varepsilon/2)$. On obtient alors,

$$\inf_{1 \leq r < p} \mathbb{P}(|Z_p - Z_r| \leq \varepsilon/2) \times \mathbb{P} \left(\sup_{1 \leq r \leq p} |Z_r| > \varepsilon \right) \leq \mathbb{P}(|Z_p| > \varepsilon/2) ;$$

puisque $Z_r = S_{n+r} - S_n$, cette inégalité se réécrit

$$\inf_{1 \leq r < p} \mathbb{P}(|S_{n+p} - S_{n+r}| \leq \varepsilon/2) \times \mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right) \leq \mathbb{P}(|S_{n+p} - S_n| > \varepsilon/2).$$

Notons $\beta_n = \sup \{ \mathbb{P}(|S_{q+n} - S_{p+n}| > \varepsilon/2) : p \geq 0, q \geq 0 \}$. On a, pour tout $p \geq 1$,

$$\mathbb{P}(|S_{n+p} - S_n| > \varepsilon/2) \leq \beta_n, \quad \inf_{1 \leq r < p} \mathbb{P}(|S_{n+p} - S_{n+r}| \leq \varepsilon/2) \geq 1 - \beta_n,$$

et par suite, $(1 - \beta_n) \mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right) \leq \beta_n$. D'autre part, comme

$$\mathbb{P}(|S_{n+p} - S_{n+q}| > \varepsilon/2) \leq \mathbb{P}(|S_{n+p} - S_n| > \varepsilon/4) + \mathbb{P}(|S_{n+q} - S_n| > \varepsilon/4),$$

$\beta_n \leq 2 \sup_{r \geq 0} \mathbb{P}(|S_{n+r} - S_n| > \varepsilon/4)$. Pour conclure, observons que, $(S_n)_{n \in \mathbf{N}^*}$ convergeant en probabilité, la Proposition 11 du Chapitre III et la majoration précédente impliquent que $\lim_{n \rightarrow +\infty} \beta_n = 0$. Si n est assez grand, de sorte que $1 - \beta_n > 0$, on a, pour tout $p \geq 1$,

$$\mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right) \leq \frac{\beta_n}{1 - \beta_n}$$

et donc

$$\mathbb{P} \left(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon \right) = \sup_{p \geq 1} \mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right) \leq \frac{\beta_n}{1 - \beta_n}.$$

Ceci montre que $(S_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers une variable réelle.

Montrons à présent que si $(S_n)_{n \in \mathbf{N}^*}$ converge en loi alors $(S_n)_{n \in \mathbf{N}^*}$ converge également en probabilité. Procédons par l'absurde. D'après la Proposition 11 du Chapitre III, si $(S_n)_{n \in \mathbf{N}^*}$ ne converge pas en probabilité, il existe $\varepsilon > 0$ et $\alpha > 0$ tels que

$$\forall n \geq 1, \quad \exists (p_n, q_n) \in \mathbf{N}^2, \quad n \leq p_n < q_n, \quad \mathbb{P}(|S_{q_n} - S_{p_n}| > \varepsilon) > \alpha. \quad (3)$$

Posons $Z_n = S_{q_n} - S_{p_n}$ et montrons que $(Z_n)_{n \in \mathbf{N}^*}$ converge en loi vers 0. Puisque S_{p_n} est indépendante de Z_n , on a, écrivant $S_{q_n} = S_{p_n} + Z_n$,

$$\forall t \in \mathbf{R}, \quad \varphi_{S_{q_n}}(t) = \varphi_{S_{p_n}}(t) \varphi_{Z_n}(t).$$

Puisque $(S_n)_{n \in \mathbf{N}^*}$ converge en loi disons vers S_∞ , φ_{S_n} converge simplement vers la fonction caractéristique φ de S_∞ . La fonction φ est continue sur \mathbf{R} et $\varphi(0) = 1$. Il existe donc $c > 0$ tel que, pour tout $|t| \leq c$, $|\varphi(t)| > 0$. Comme $n \leq p_n < q_n$, $\varphi_{S_{p_n}}$ et $\varphi_{S_{q_n}}$ convergent simplement vers φ lorsque $n \rightarrow +\infty$, et par conséquent, pour tout $|t| \leq c$, $\lim_{n \rightarrow +\infty} \varphi_{Z_n}(t) = 1$. Puisque, pour tout $x \in \mathbf{R}$, $1 - \cos(2x) \leq 4(1 - \cos x)$, on a,

$$\forall t \in \mathbf{R}, \quad 0 \leq 1 - \operatorname{Re}(\varphi_{Z_n}(2t)) \leq 4[1 - \operatorname{Re}(\varphi_{Z_n}(t))].$$

Par suite, pour tout réel t , $\lim_{n \rightarrow +\infty} \operatorname{Re}(\varphi_{Z_n}(t)) = 1$ et finalement, puisque $|\varphi_{Z_n}(t)| \leq 1$, $\lim_{n \rightarrow +\infty} \varphi_{Z_n}(t) = 1$. La fonction constante égale à un est la transformée de Fourier de la probabilité δ_0 . $(Z_n)_{n \in \mathbf{N}^*}$ converge donc en loi vers 0. Via la Proposition 19 du Chapitre III, $(Z_n)_{n \in \mathbf{N}^*}$ converge vers 0 en probabilité. Ceci contredit (3). La suite $(S_n)_{n \in \mathbf{N}^*}$ converge en probabilité.

Pour finir, rappelons que la convergence presque sûre implique la convergence en loi. \square

Comme conséquence du théorème de Paul Lévy, nous donnons un critère simple permettant d'obtenir la convergence presque sûre d'une série de variables aléatoires indépendantes de carré intégrable et centrées.

Proposition 8 (Séries centrées). *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes; on suppose que, pour tout $n \in \mathbf{N}^*$, X_n est de carré intégrable et que $\mathbb{E}[X_n] = 0$.*

Si $\sum_{n \geq 1} \mathbb{E}[X_n^2] < +\infty$, alors $(S_n)_{n \in \mathbf{N}^}$ converge presque sûrement et dans L^2 vers une variable aléatoire réelle.*

Démonstration. Puisque les variables aléatoires X_n sont centrées, $\mathbb{E}[X_n^2] = \mathbb{V}(X_n)$. On a, pour tous $n \in \mathbf{N}^*$, $r \in \mathbf{N}^*$, via l'indépendance de $(X_n)_{n \in \mathbf{N}^*}$,

$$\mathbb{E}[|S_{n+r} - S_n|^2] = \mathbb{V}\left(\sum_{i=n+1}^{n+r} X_i\right) = \sum_{i=n+1}^{n+r} \mathbb{V}(X_i) = \sum_{i=n+1}^{n+r} \mathbb{E}[X_i^2] \leq \sum_{i>n} \mathbb{E}[X_i^2],$$

qui est le reste d'une série convergente. La suite $(S_n)_{n \in \mathbf{N}^*}$ est de Cauchy dans L^2 qui est complet; elle converge donc dans L^2 vers la variable réelle S_∞ . La convergence a lieu à fortiori en probabilité et d'après le théorème de Paul Lévy $(S_n)_{n \in \mathbf{N}^*}$ converge presque sûrement, disons vers S'_∞ . Pour finir, $S_\infty = S'_\infty$ presque sûrement puisque $(S_n)_{n \in \mathbf{N}^*}$ converge en probabilité vers S_∞ et S'_∞ ; cf. Remarque page 44 et Proposition 9 du Chapitre III. \square

Finissons ce paragraphe par une inégalité classique due à Kolmogorov.

Proposition 9 (Inégalité de Kolmogorov). *Soit $(X_n)_{n \in \mathbf{N}^*}$ des variables aléatoires réelles indépendantes ; on suppose que, pour tout $n \in \mathbf{N}^*$, X_n est de carré intégrable et que $\mathbb{E}[X_n] = 0$.*

Pour tout $a > 0$ et tout $n \in \mathbf{N}^$,*

$$\mathbb{P} \left(\sup_{1 \leq k \leq n} |S_k| \geq a \right) \leq \frac{\mathbb{E}[S_n^2]}{a^2}.$$

Démonstration. Introduisons la variable aléatoire $\tau = \inf\{k \geq 1 : |S_k| \geq a\}$, $\inf \emptyset = +\infty$. $\{\tau \leq n\} = \{S_n^* \geq a\}$ et l'on souhaite majorer la probabilité de cet événement. On a, pour tout $1 \leq k \leq n$,

$$\mathbb{E} \left[S_n^2 \mathbf{1}_{\tau=k} \right] = \mathbb{E} \left[(S_n - S_k)^2 \mathbf{1}_{\tau=k} \right] + 2 \mathbb{E} \left[(S_n - S_k) S_k \mathbf{1}_{\tau=k} \right] + \mathbb{E} \left[S_k^2 \mathbf{1}_{\tau=k} \right].$$

Rappelons que $\{\tau = 1\} = \{|S_1| \geq a\}$, et que, pour $k \geq 2$,

$$\{\tau = k\} = \{|S_1| < a\} \cap \dots \cap \{|S_{k-1}| < a\} \cap \{|S_k| \geq a\}.$$

En particulier, $S_k^2 \mathbf{1}_{\tau=k} \geq a^2 \mathbf{1}_{\tau=k}$ et les variables $S_n - S_k$ et $S_k \mathbf{1}_{\tau=k}$ sont indépendantes. Par suite, les variables étant centrées,

$$\mathbb{E} \left[S_n^2 \mathbf{1}_{\tau=k} \right] \geq 2 \mathbb{E}[S_n - S_k] \mathbb{E}[S_k \mathbf{1}_{\tau=k}] + a^2 \mathbb{P}(\tau = k) = a^2 \mathbb{P}(\tau = k).$$

Il reste à sommer de $k = 1$ à $k = n$ pour obtenir

$$\mathbb{E} \left[S_n^2 \right] \geq \mathbb{E} \left[S_n^2 \mathbf{1}_{\tau \leq n} \right] \geq a^2 \mathbb{P}(\tau \leq n)$$

qui est l'inégalité requise. □

Remarque. On peut démontrer le résultat sur les séries centrées seulement à partir de l'inégalité de Kolmogorov sans utiliser le théorème de Paul Lévy. J'esquisse la démonstration. On a montré que la suite $(S_n)_{n \in \mathbf{N}^*}$ convergeait vers S_∞ dans L^2 et donc en probabilité. Nous allons montrer que cette suite converge presque sûrement en appliquant le critère de Cauchy (Proposition 11 du Chapitre III). Soit donc $\varepsilon > 0$. On a, d'après l'inégalité de Kolmogorov, pour tout $n \in \mathbf{N}^*$ et tout $p \in \mathbf{N}^*$,

$$\mathbb{P} \left(\sup_{1 \leq r \leq p} |S_{n+r} - S_n| > \varepsilon \right) \leq \frac{\mathbb{E}[(S_{n+p} - S_n)^2]}{\varepsilon^2} \leq \varepsilon^{-2} \sum_{i>n} \mathbb{E} \left[X_i^2 \right].$$

En prenant le sup en p , on obtient $\mathbb{P} \left(\sup_{r \geq 0} |S_{n+r} - S_n| > \varepsilon \right) \leq \varepsilon^{-2} \sum_{i>n} \mathbb{E} \left[X_i^2 \right]$ qui donne le résultat.

3. Loi forte de Kolmogorov.

Dans ce dernier paragraphe, nous démontrons que la loi forte des grands nombres est valable pour les suites i.i.d. de variables intégrables ; ce résultat est dû à Kolmogorov.

Nous commençons par quelques rappels sur les séries réelles.

Lemme 10. *Soit $(b_n)_{n \in \mathbf{N}^*}$ une suite croissante de réels strictement positifs – on pose $b_0 = 0$ – telle que $\lim_{n \rightarrow +\infty} b_n = +\infty$. Soit $(x_n)_{n \in \mathbf{N}^*}$ une suite de réels.*

Césaro. *Si $(x_n)_{n \in \mathbf{N}^*}$ converge vers x , $\lim_{n \rightarrow +\infty} b_n^{-1} \sum_{i=1}^n (b_i - b_{i-1}) x_i = x$.*

Kronecker. Si la série $\sum b_n^{-1}x_n$ converge dans \mathbf{R} , alors $\lim_{n \rightarrow +\infty} b_n^{-1} \sum_{i=1}^n x_i = 0$.

Démonstration. Pour la preuve du lemme de Césaro, notons $u_n = b_n^{-1} \sum_{i \leq n} (b_i - b_{i-1})x_i - x$ et observons que

$$|u_n| = \left| b_n^{-1} \sum_{i \leq n} (b_i - b_{i-1})(x_i - x) \right| \leq b_n^{-1} \sum_{i \leq n} (b_i - b_{i-1})|x_i - x|.$$

La suite $(x_n)_{n \in \mathbf{N}^*}$ est convergente dans \mathbf{R} donc bornée disons par x^* . On a alors pour tous $n \geq 1, k \geq 1$,

$$|u_{n+k}| \leq 2x^* b_{n+k}^{-1} b_k + b_{n+k}^{-1} (b_{n+k} - b_k) \sup_{i \geq k+1} |x_i - x| \leq 2x^* b_{n+k}^{-1} b_k + \sup_{i \geq k+1} |x_i - x|.$$

Par conséquent, pour tout $k \geq 1$,

$$\limsup |u_n| = \limsup_{n \rightarrow +\infty} |u_{n+k}| \leq \sup_{i \geq k+1} |x_i - x|.$$

Il reste à prendre la limite lorsque k tend vers $+\infty$ pour conclure.

Passons à la démonstration du lemme de Kronecker. Notons $R_i = \sum_{k \geq i} b_k^{-1} x_k$; par hypothèse $\lim_{i \rightarrow +\infty} R_i = 0$. On a $b_i(R_i - R_{i+1}) = x_i$ de sorte que, comme $b_0 = 0$,

$$\sum_{i=1}^n x_i = \sum_{i=1}^n b_i(R_i - R_{i+1}) = \sum_{i=1}^n b_i R_i - \sum_{i=1}^{n+1} b_{i-1} R_i = \sum_{i=1}^n (b_i - b_{i-1}) R_i - b_n R_{n+1}.$$

Il reste à diviser par b_n et appliquer le lemme de Césaro pour conclure. \square

Remarque. On applique souvent ces deux lemmes avec $b_n = n$.

Puisque nous sommes dans les rappels, mentionnons la majoration classique du reste des séries de Riemann.

Lemme 11. Pour $\alpha > 1$ et $k \geq 1$, $\sum_{n \geq k+1} n^{-\alpha} \leq k^{1-\alpha}/(\alpha - 1)$.

Démonstration. La fonction $x \rightarrow x^{-\alpha}$ est décroissante sur $]0, +\infty[$. Pour tout $n > 1$, et tout $x \in [n-1, n]$, $n^{-\alpha} \leq x^{-\alpha}$. Par conséquent, $n^{-\alpha} \leq \int_{n-1}^n x^{-\alpha} dx$ et

$$\sum_{n \geq k+1} n^{-\alpha} \leq \sum_{n \geq k+1} \int_{n-1}^n x^{-\alpha} dx = \int_k^{+\infty} x^{-\alpha} dx = k^{1-\alpha}/(\alpha - 1).$$

\square

Nous passons au résultat principal de ce chapitre : la loi des grands nombres.

Théorème 12 (Loi des grands nombres, Kolmogorov). Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées; on note, pour tout entier $n \geq 1$, $M_n = n^{-1}(X_1 + \dots + X_n)$.

1. Si X_1 est intégrable, $(M_n)_{n \in \mathbf{N}^*}$ converge presque sûrement et dans L^1 vers $m = \mathbb{E}[X_1]$.
2. Si X_1 n'est pas intégrable, au moins un des deux événements $\{\limsup M_n = +\infty\}$ et $\{\liminf M_n = -\infty\}$ a pour probabilité un.

Remarque. En particulier, pour une suite $(X_n)_{n \in \mathbf{N}^*}$ i.i.d., $(M_n)_{n \in \mathbf{N}^*}$ converge presque sûrement dans \mathbf{R} si et seulement si X_1 est intégrable.

Démonstration. 1. Supposons que X_1 est intégrable et, pour commencer, que $m = \mathbb{E}[X_1] = 0$. Introduisons quelques notations. Pour tout $n \geq 1$,

$$\widehat{X}_n = X_n \mathbf{1}_{|X_n| < n}, \quad \widehat{M}_n = \frac{1}{n} \sum_{i=1}^n \widehat{X}_i, \quad \widetilde{X}_n = \widehat{X}_n - \mathbb{E}[\widehat{X}_n], \quad \widetilde{M}_n = \frac{1}{n} \sum_{i=1}^n \widetilde{X}_i.$$

Notons que \widetilde{X}_n est bien définie puisque \widehat{X}_n est bornée donc intégrable et que $(\widehat{X}_n)_{n \in \mathbf{N}^*}$ et $(\widetilde{X}_n)_{n \in \mathbf{N}^*}$ sont deux suites de variables i.i.d.

Pour montrer que $(M_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers 0, nous allons procéder en deux étapes :

(a) tout d'abord nous établirons que

$$M_n \longrightarrow 0 \text{ } \mathbb{P}\text{-p.s.} \iff \widehat{M}_n \longrightarrow 0 \text{ } \mathbb{P}\text{-p.s.} \iff \widetilde{M}_n \longrightarrow 0 \text{ } \mathbb{P}\text{-p.s.}$$

(b) puis nous montrerons que la suite $(\widetilde{M}_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers 0.

(a) Pour établir la première équivalence, montrons que $M_n - \widehat{M}_n$ converge presque sûrement vers 0. On a, pour tout $n \geq 1$, $M_n - \widehat{M}_n = n^{-1} \sum_{i=1}^n X_i \mathbf{1}_{|X_i| \geq i}$. On a, les variables $(X_n)_{n \in \mathbf{N}^*}$ étant identiquement distribuées,

$$\sum_{n \geq 1} \mathbb{P}(|X_n| \geq n) = \sum_{n \geq 1} \mathbb{P}(|X_1| \geq n) \leq \sum_{n \geq 0} \mathbb{P}(|X_1| > n) \leq 1 + \mathbb{E}[|X_1|] < +\infty.$$

D'après le lemme de Borel-Cantelli, $\mathbb{P}(\limsup\{|X_n| \geq n\}) = 0$. Si $\omega \in (\limsup\{|X_n| \geq n\})^c = \liminf\{|X_n| < n\}$, il existe un entier $n_\omega \geq 1$, tel que pour tout $n \geq n_\omega$, $|X_n(\omega)| < n$ soit $X_n(\omega) = \widehat{X}_n(\omega)$. Pour tout $\omega \in (\limsup\{|X_n| \geq n\})^c$,

$$\forall n \geq n_\omega, \quad M_n(\omega) - \widehat{M}_n(\omega) = \frac{1}{n} \sum_{i=1}^{n_\omega} X_i \mathbf{1}_{|X_i| \geq i};$$

il s'agit donc d'une somme finie divisée par n . La première équivalence est donc établie.

Passons à la seconde équivalence en utilisant la même démarche. On a, pour tout $n \geq 1$, $\widehat{M}_n - \widetilde{M}_n = n^{-1} \sum_{i=1}^n \mathbb{E}[X_i \mathbf{1}_{|X_i| < i}] = n^{-1} \sum_{i=1}^n \mathbb{E}[X_1 \mathbf{1}_{|X_1| < i}]$ puisque les variables sont identiquement distribuées. X_1 est intégrable donc finie presque sûrement : $X_1 \mathbf{1}_{|X_1| < i}$ converge vers X_1 presque sûrement lorsque $i \rightarrow +\infty$; de plus $\sup_{i \geq 1} |X_1 \mathbf{1}_{|X_1| < i}| \leq |X_1|$ qui est intégrable. Le théorème de convergence dominée montre que $\mathbb{E}[X_1 \mathbf{1}_{|X_1| < i}] \rightarrow \mathbb{E}[X_1] = 0$ si $i \rightarrow +\infty$ et le lemme de Césaro entraîne que $\widehat{M}_n - \widetilde{M}_n = n^{-1} \sum_{i=1}^n \mathbb{E}[X_1 \mathbf{1}_{|X_1| < i}]$ converge également vers 0.

(b) Montrons à présent que $(\widetilde{M}_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers 0. Pour cela nous allons appliquer le lemme de Kronecker : pour montrer que $n^{-1} \sum_{i=1}^n \widetilde{X}_i$ converge vers 0 presque sûrement, il suffit de montrer que $\sum_{i=1}^n i^{-1} \widetilde{X}_i$ converge dans \mathbf{R} presque sûrement. Or les variables aléatoires $(n^{-1} \widetilde{X}_n)_{n \in \mathbf{N}^*}$ sont indépendantes et, pour tout $n \in \mathbf{N}^*$, $n^{-1} \widetilde{X}_n$ est bornée (par deux) donc de carré intégrable et finalement $\mathbb{E}[n^{-1} \widetilde{X}_n] = 0$. D'après le résultat sur les séries centrées (Proposition 8), il suffit de vérifier que $\sum_{n \geq 1} n^{-2} \mathbb{E}[\widetilde{X}_n^2] < +\infty$ pour obtenir la convergence presque sûre de $\sum_{n \geq 1} n^{-1} \widetilde{X}_n$ dans \mathbf{R} . On a

$$\sum_{n \geq 1} n^{-2} \mathbb{E}[\widetilde{X}_n^2] = \sum_{n \geq 1} n^{-2} \mathbb{E}[(\widehat{X}_n - \mathbb{E}[\widehat{X}_n])^2] = \sum_{n \geq 1} n^{-2} \mathbb{V}(\widehat{X}_n) \leq \sum_{n \geq 1} n^{-2} \mathbb{E}[\widehat{X}_n^2].$$

Or les variables $(X_n)_{n \in \mathbf{N}^*}$ étant identiquement distribuées, on a

$$\forall n \geq 1, \quad \mathbb{E}[\widehat{X}_n^2] = \mathbb{E}\left[X_n^2 \mathbf{1}_{|X_n| < n}\right] = \mathbb{E}\left[X_1^2 \mathbf{1}_{|X_1| < n}\right],$$

et par convergence monotone,

$$\sum_{n \geq 1} n^{-2} \mathbb{E}[\widehat{X}_n^2] \leq \sum_{n \geq 1} n^{-2} \mathbb{E}\left[X_1^2 \mathbf{1}_{|X_1| < n}\right] = \mathbb{E}\left[\sum_{n \geq 1} n^{-2} X_1^2 \mathbf{1}_{|X_1| < n}\right].$$

Or pour tout $x \geq 0$, on a, notant $[x]$ la partie entière du réel x , via le Lemme 11,

$$\sum_{n \geq 1} n^{-2} x^2 \mathbf{1}_{x < n} = x^2 \sum_{n \geq [x]+1} n^{-2} = \frac{x^2}{([x]+1)^2} + x^2 \sum_{n \geq [x]+2} n^{-2} \leq \frac{x^2}{([x]+1)^2} + \frac{x^2}{[x]+1} \leq 2x.$$

Par conséquent, $\sum_{n \geq 1} n^{-2} \mathbb{E}[\widehat{X}_n^2] \leq 2 \mathbb{E}[|X_1|] < +\infty$. On a donc convergence presque sûre de $(\widetilde{M}_n)_{n \in \mathbf{N}^*}$ vers 0 et par suite celle de $(M_n)_{n \in \mathbf{N}^*}$.

Montrons que la suite $(M_n)_{n \in \mathbf{N}^*}$ converge vers 0 également dans L^1 . Soit $k \in \mathbf{N}^*$. Écrivons que $|M_n| = \min(|M_n|, k) + (|M_n| - k)^+$. La fonction $x \mapsto (x - k)^+$ est convexe et croissante; par suite $(|M_n| - k)^+ \leq n^{-1} \sum_{i=1}^n (|X_i| - k)^+$ et, puisque les variables sont identiquement distribuées,

$$\mathbb{E}[|M_n|] \leq \mathbb{E}[\min(|M_n|, k)] + n^{-1} \sum_{i=1}^n \mathbb{E}\left[(|X_i| - k)^+\right] = \mathbb{E}[\min(|M_n|, k)] + \mathbb{E}\left[(|X_1| - k)^+\right].$$

$\min(|M_n|, k)$ converge presque sûrement vers 0 si $n \rightarrow +\infty$ et $0 \leq \min(|M_n|, k) \leq k$; par convergence dominée, $\lim_{n \rightarrow +\infty} \mathbb{E}[\min(|M_n|, k)] = 0$. Il vient alors

$$\forall k \in \mathbf{N}^*, \quad \limsup_{n \rightarrow +\infty} \mathbb{E}[|M_n|] \leq \mathbb{E}\left[(|X_1| - k)^+\right].$$

Or, comme $|X_1|$ est intégrable, $(|X_1| - k)^+$ converge presque sûrement vers 0 lorsque $k \rightarrow +\infty$ et $0 \leq (|X_1| - k)^+ \leq |X_1|$. Par convergence dominée, on a

$$\limsup_{n \rightarrow +\infty} \mathbb{E}[|M_n|] \leq \lim_{k \rightarrow +\infty} \mathbb{E}\left[(|X_1| - k)^+\right] = 0.$$

Ceci termine la démonstration dans le cas $m = 0$.

Considérons le cas général. Observons que, notant $\overline{X}_n = X_n - m$ pour tout $n \geq 1$, on a

$$M_n - m = n^{-1} \sum_{i=1}^n (X_i - m) = n^{-1} \sum_{i=1}^n \overline{X}_i = \overline{M}_n;$$

les variables $(\overline{X}_n)_{n \in \mathbf{N}^*}$ sont i.i.d; \overline{X}_1 est intégrable et $\mathbb{E}[\overline{X}_1] = 0$. D'après le cas $m = 0$, \overline{M}_n converge vers 0 presque sûrement et dans L^1 : M_n converge donc vers m presque sûrement et dans L^1 .

2. Plaçons nous dans le cas où X_1 n'est pas intégrable. Comme nous l'avons vu au Chapitre II – voir les exemples de la page 37 –, $\liminf M_n$ et $\limsup M_n$ sont des variables asymptotiques de la suite de variables indépendantes $(X_n)_{n \in \mathbf{N}^*}$. D'après le Théorème 15 et le Corollaire 16 du Chapitre II, les événements $\{\liminf M_n = +\infty\}$ et $\{\liminf M_n = -\infty\}$ ont une probabilité égale à 0 ou 1 et en fait il existe deux éléments c^* et c_* de $\overline{\mathbf{R}}$ tels que, presque sûrement, $\liminf M_n = c_*$, $\limsup M_n = c^*$.

Supposons que les deux événements $\{\limsup M_n = +\infty\}$ et $\{\liminf M_n = -\infty\}$ sont négligeables; on a $-\infty < c_* \leq c^* < +\infty$. Écrivons alors que, $\frac{X_n}{n} = M_n - \frac{n-1}{n}M_{n-1}$, de sorte que

$$\limsup \frac{X_n}{n} \leq c^* - c_*, \quad \liminf \frac{X_n}{n} \geq c_* - c^*.$$

Soit $c > c^* - c_*$. On a $\limsup\{X_n \geq cn\} \subset \left\{\limsup \frac{X_n}{n} \geq c\right\}$; d'où $\mathbb{P}(\limsup\{X_n \geq cn\}) = 0$. Comme les variables $(X_n)_{n \in \mathbf{N}^*}$ sont i.i.d., le lemme de Borel–Cantelli (2) implique,

$$\sum_{n \geq 1} \mathbb{P}(X_1^+ \geq cn) = \sum_{n \geq 1} \mathbb{P}(X_1 \geq cn) = \sum_{n \geq 1} \mathbb{P}(X_n \geq cn) < +\infty.$$

X_1^+/c et donc X_1^+ est intégrable. De même, $\limsup\{X_n \leq -cn\} \subset \left\{\liminf \frac{X_n}{n} \leq -c\right\}$, et utilisant les mêmes arguments, on obtient

$$\sum_{n \geq 1} \mathbb{P}(X_1^- \geq cn) = \sum_{n \geq 1} \mathbb{P}(X_1 \leq -cn) = \sum_{n \geq 1} \mathbb{P}(X_n \leq -cn) < +\infty.$$

X_1^- est intégrable. Par conséquent, $\mathbb{E}[|X_1|] = \mathbb{E}[X_1^+] + \mathbb{E}[X_1^-] < +\infty$. X_1 est intégrable. \square

Remarque. Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires positives i.i.d. avec $\mathbb{E}[X_1] = +\infty$. Alors, presque sûrement, $\lim_{n \rightarrow +\infty} M_n = +\infty$.

En effet, pour tout $k \in \mathbf{N}^*$, $\liminf \frac{1}{n} \sum_{i=1}^n X_i \geq \liminf \frac{1}{n} \sum_{i=1}^n \min(X_i, k)$. D'après la loi forte des grands nombres, pour tout $k \in \mathbf{N}^*$, il existe $N_k \in \mathcal{F}$ tel que $\mathbb{P}(N_k) = 0$ et, pour tout $\omega \in N_k^c$, $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \min(X_i(\omega), k) = \mathbb{E}[\min(X_1, k)]$. Posons $N = \cup_{k \in \mathbf{N}^*} N_k$; $\mathbb{P}(N) = 0$ et, pour tout $\omega \in N^c$,

$$\forall k \in \mathbf{N}^*, \quad \liminf \frac{1}{n} \sum_{i=1}^n X_i(\omega) \geq \liminf \frac{1}{n} \sum_{i=1}^n \min(X_i(\omega), k) = \mathbb{E}[\min(X_1, k)].$$

Par convergence monotone, $\lim_{k \rightarrow +\infty} \mathbb{E}[\min(X_1, k)] = +\infty$ et par suite,

$$\forall \omega \in N^c, \quad \liminf \frac{1}{n} \sum_{i=1}^n X_i(\omega) \geq \lim_{k \rightarrow +\infty} \mathbb{E}[\min(X_1, k)] = +\infty.$$

Chapitre V. Théorème Limite Central

Nous démontrons dans ce chapitre un résultat fondamental du calcul des probabilités : le « Théorème Limite Central ». Signalons que l'adjectif « central » se reporte à « théorème » et non à « limite » ; il s'agit donc d'un théorème limite qui joue un rôle central en théorie des probabilités. Nous utiliserons l'abréviation, communément utilisée, TCL qui nous vient directement de la traduction anglaise.

De quoi s'agit-il ? Commençons par une application de la loi des grands nombres : la méthode de Monte-Carlo. Soit $f : [0, 1] \rightarrow \mathbf{R}$ une application mesurable, intégrable sur $[0, 1]$ par rapport à la mesure de Lebesgue. Imaginons que l'on veuille calculer numériquement l'intégrale $I = \int_0^1 f(x) dx$. Pour cela, considérons $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables i.i.d. suivant la loi uniforme sur $[0, 1]$ et, pour tout $n \in \mathbf{N}^*$,

$$I_n(\omega) = \frac{1}{n} \sum_{i=1}^n f[X_i(\omega)], \quad \omega \in \Omega.$$

Les variables aléatoires $f(X_i)$ sont i.i.d. et $f(X_1)$ est intégrable puisque

$$\mathbb{E}[|f(X_1)|] = \int_0^1 |f(x)| dx < +\infty.$$

On obtient, en appliquant la loi des grands nombres – cf. Théorème 12 du Chapitre IV –, la convergence presque sûre de la suite $(I_n(\omega))_{n \in \mathbf{N}^*}$ vers I . D'un point de vue pratique, pour obtenir une approximation de I , il « suffit » d'avoir un échantillon de la loi uniforme sur $[0, 1]$ et de calculer $I_n(\omega)$. Une question se pose alors : quelle est la précision de cette approximation ? Faut-il choisir n grand ?

Pour tenter de répondre à ces questions, on étudie la convergence de la suite $n^\alpha (I_n - I)$ en cherchant une limite non triviale.

1. TCL pour les variables réelles.

Replaçons-nous dans le contexte du chapitre précédent en considérant une suite de v.a.r. $(X_n)_{n \in \mathbf{N}^*}$ i.i.d. Nous noterons toujours $S_n = X_1 + \dots + X_n$ et $M_n = n^{-1}S_n$. Lorsque X_1 est intégrable, $(M_n)_{n \in \mathbf{N}^*}$ converge presque sûrement vers $m = \mathbb{E}[X_1]$; nous allons voir que, lorsque X_1 est de carré intégrable, $\sqrt{n}(M_n - m)$ converge en loi vers une variable aléatoire gaussienne centrée de variance $\mathbb{V}(X_1)$.

Commençons par un lemme technique.

Lemme 1. Soit $(z_n)_{n \in \mathbf{N}^*}$ une suite de complexes.

Si $\lim_{n \rightarrow +\infty} nz_n = z$, alors $\lim_{n \rightarrow +\infty} (1 + z_n)^n = e^z$.

Démonstration. Notons $w_n = e^{\frac{z}{n}}$. On a

$$(1 + z_n)^n - e^z = (1 + z_n)^n - w_n^n = (1 + z_n - w_n) \sum_{k=0}^{n-1} (1 + z_n)^{n-1-k} w_n^k ;$$

par conséquent,

$$\begin{aligned} & |(1 + z_n)^n - e^z| \\ & \leq n |1 + z_n - w_n| \sup_{k \leq n-1} \left\{ |1 + z_n|^{n-1-k} |w_n|^k \right\} \leq n |1 + z_n - w_n| (1 + |z_n|)^n e^{|z|}, \end{aligned}$$

et, comme $\ln(1 + x) \leq x$ pour tout $x > -1$,

$$|(1 + z_n)^n - e^z| \leq n |1 + z_n - w_n| e^{n|z_n| + |z|}.$$

On a d'autre part, pour tout $|z| \leq 1$,

$$|e^z - 1 - z| = \left| \sum_{n \geq 2} \frac{z^n}{n!} \right| \leq |z|^2 \sum_{n \geq 2} \frac{1}{n!} \leq |z|^2 \sum_{n \geq 2} 2^{-(n-1)} = |z|^2.$$

Il vient alors, pour tout $n \geq |z|$, $|1 + z_n - w_n| \leq |z_n - z/n| + |z|^2/n^2$, et par suite

$$|(1 + z_n)^n - e^z| \leq \left(|nz_n - z| + n^{-1}|z|^2 \right) e^{n|z_n| + |z|}.$$

Le résultat s'en suit immédiatement. \square

Passons à présent au résultat essentiel de ce paragraphe sous une forme qui se généralisera au cas vectoriel.

Théorème 2. *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées avec X_1 de carré intégrable ; on note $m = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}(X_1)$. Considérons, pour tout $n \in \mathbf{N}^*$,*

$$T_n = \sqrt{n} \left(\frac{S_n}{n} - m \right), \quad \text{où } S_n = X_1 + \dots + X_n.$$

Alors la suite $(T_n)_{n \in \mathbf{N}^}$ converge en loi, lorsque $n \rightarrow +\infty$, vers une variable aléatoire réelle de loi $\mathcal{N}(0, \sigma^2)$.*

Démonstration. Rappelons que la fonction caractéristique d'une v.a.r. de loi $\mathcal{N}(0, \sigma^2)$ est $t \mapsto e^{-\frac{\sigma^2 t^2}{2}}$. D'après le théorème de Paul Lévy, il suffit donc de vérifier que

$$\forall t \in \mathbf{R}, \quad \lim_{n \rightarrow +\infty} \varphi_{T_n}(t) = e^{-\frac{\sigma^2 t^2}{2}}.$$

Commençons par remarquer que, pour tout $n \in \mathbf{N}^*$, notant $\bar{X}_i = X_i - m$ et $\bar{S}_n = \sum_{i=1}^n \bar{X}_i$,

$$T_n = \sqrt{n} \left(\frac{S_n}{n} - m \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{X}_i = \frac{\bar{S}_n}{\sqrt{n}}.$$

Soit φ la fonction caractéristique de \bar{X}_1 ; les variables $(\bar{X}_n)_{n \in \mathbf{N}^*}$ étant i.i.d., $\varphi_{\bar{S}_n} = \varphi^n$ et nous avons, pour tout réel t ,

$$\varphi_{T_n}(t) = \mathbb{E} \left[e^{it \frac{\bar{S}_n}{\sqrt{n}}} \right] = \varphi_{\bar{S}_n}(t/\sqrt{n}) = \varphi(t/\sqrt{n})^n.$$

Puisque X_1 est de carré intégrable, il en est de même de \overline{X}_1 . D'après la Proposition 17 du Chapitre I, φ est de classe \mathcal{C}^2 sur \mathbf{R} et de plus

$$\varphi(0) = 1, \quad \varphi'(0) = i \mathbb{E}[\overline{X}_1] = 0, \quad \varphi''(0) = -\mathbb{E}[\overline{X}_1^2] = -\mathbb{V}(X_1) = -\sigma^2.$$

La formule de Taylor donne $\varphi(t) = 1 - \frac{t^2\sigma^2}{2} + t^2 \varepsilon(t)$ où $\lim_{t \rightarrow 0} \varepsilon(t) = 0$ et

$$\varphi_{T_n}(t) = \left(1 - \frac{t^2\sigma^2}{2n} + \frac{t^2}{n} \varepsilon(t/\sqrt{n}) \right)^n.$$

Comme $n \left(-\frac{t^2\sigma^2}{2n} + \frac{t^2}{n} \varepsilon(t/\sqrt{n}) \right) \rightarrow -\frac{t^2\sigma^2}{2}$ si $n \rightarrow +\infty$, on obtient, via le Lemme 1,

$$\forall t \in \mathbf{R}, \quad \lim_{n \rightarrow +\infty} \varphi_{T_n}(t) = e^{-\frac{t^2\sigma^2}{2}}$$

ce qui termine la démonstration. □

Pour les variables réelles, on utilise plutôt une variante du résultat précédent.

Corollaire 3. *Soit $(X_n)_{n \in \mathbf{N}^*}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées avec X_1 de carré intégrable ; on note $m = \mathbb{E}[X_1]$, $\sigma^2 = \mathbb{V}(X_1)$ et on suppose que $\sigma > 0$. Considérons, pour tout $n \in \mathbf{N}^*$,*

$$Z_n = \frac{1}{\sigma\sqrt{n}}(S_n - nm), \quad \text{où } S_n = X_1 + \dots + X_n.$$

La suite $(Z_n)_{n \in \mathbf{N}^}$ converge en loi, lorsque $n \rightarrow +\infty$, vers une variable aléatoire réelle G de loi $\mathcal{N}(0, 1)$.*

Démonstration. Remarquons que $Z_n = T_n/\sigma$. D'après le théorème précédent, nous savons que T_n converge en loi vers une variable réelle G_{σ^2} de loi $\mathcal{N}(0, \sigma^2)$. La fonction $x \mapsto x/\sigma$ étant continue sur \mathbf{R} , Z_n converge en loi vers G_{σ^2}/σ qui suit la loi normale centrée réduite $\mathcal{N}(0, 1)$. □

Conservons les hypothèses ($\sigma > 0$) et notations du résultat précédent et désignons par Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$ soit

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

Φ est une fonction continue sur \mathbf{R} : la fonction de répartition de Z_n , F_{Z_n} , converge vers Φ uniformément sur \mathbf{R} d'après le lemme de Dini – Lemme 5 du Chapitre IV. Remarquons que, pour tout $t \in \mathbf{R}$,

$$|F_{Z_n}(t-) - \Phi(t)| = \lim_{s \rightarrow t-} |F_{Z_n}(s) - \Phi(s)| \leq \|F_{Z_n} - \Phi\|_{\infty}.$$

En particulier, on a, pour tous $s \leq t$,

$$|\mathbb{P}(s \leq Z_n \leq t) - \mathbb{P}(s \leq G \leq t)| = |[F_{Z_n}(t) - F_{Z_n}(s-)] - [\Phi(t) - \Phi(s)]| \leq 2 \|F_{Z_n} - \Phi\|_{\infty} ;$$

on vérifie sans peine que l'inégalité précédente demeure valable pour tout intervalle réel non nécessairement borné. Si donc I est un intervalle, on a

$$\left| \mathbb{P}(Z_n \in I) - \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{x^2}{2}} dx \right| \leq 2 \|F_{Z_n} - \Phi\|_{\infty}.$$

La loi forte des grands nombres donne le limite de $M_n = \frac{S_n}{n}$ lorsque n tend vers $+\infty$; le TCL quant à lui donne une information sur les déviations d'ordre $\frac{1}{\sqrt{n}}$ de M_n par rapport à sa limite puisque

$$\lim_{n \rightarrow +\infty} \sup_{t \geq 0} |\mathbb{P}(|M_n - m| > \sigma t / \sqrt{n}) - 2[1 - \Phi(t)]| = 0.$$

En effet, observons que

$$\mathbb{P}(|M_n - m| > \sigma t / \sqrt{n}) = \mathbb{P}(|Z_n| > t) = F_{Z_n}(-t-) + 1 - F_{Z_n}(t),$$

et que $\Phi(-t) = 1 - \Phi(t)$ de sorte que $F_{Z_n}(-t-) + 1 - F_{Z_n}(t)$ converge uniformément sur \mathbf{R} vers $2[1 - \Phi(t)]$.

Historiquement, le TCL a d'abord été démontré pour la loi de Bernoulli ; plus précisément,

Théorème 4 (de Moivre–Laplace). *Si S_n est une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ avec $p \in]0, 1[$, alors*

$$\frac{S_n - np}{\sqrt{np(1-p)}} = \sqrt{\frac{n}{p(1-p)}} \left(\frac{S_n}{n} - p \right)$$

converge en loi, lorsque $n \rightarrow +\infty$, vers une variable aléatoire de loi $\mathcal{N}(0, 1)$.

Démonstration. Il suffit de remarquer que S_n a la loi de $X_1 + \dots + X_n$, où les X_i sont i.i.d. de loi de Bernoulli de paramètre p et d'appliquer le Corollaire 3 en notant que $\mathbb{E}[X_1] = p$ et $\mathbb{V}(X_1) = p(1-p)$. \square

Exemple (Intervalle de confiance). Une application du théorème de De Moivre–Laplace est la construction d'intervalles de confiance pour l'estimation d'une probabilité inconnue $p \in]0, 1[$ à partir de l'observation d'un échantillon de n variables de Bernoulli indépendantes de paramètre p .

Par exemple, si μ est une probabilité sur \mathbf{R}^d et B un borélien de \mathbf{R}^d , on peut estimer $\mu(B)$ à l'aide d'un n -échantillon de loi μ via la mesure empirique $\mu_n(B)$ du borélien B :

$$\mu_n(B) = n^{-1} N_n(B) = n^{-1} \sum_{i=1}^n \mathbf{1}_B(X_i) ;$$

voir au Chapitre IV l'exemple de la page 58 ainsi que le Théorème 3.

Considérons, pour $t > 0$, l'événement

$$A_n(t) = \left\{ \omega \in \Omega : -t \leq \sqrt{\frac{n}{p(1-p)}} \left(\frac{S_n(\omega)}{n} - p \right) \leq t \right\}.$$

D'après le théorème de de Moivre–Laplace, si n est suffisamment grand, la probabilité de $A_n(t)$ est approximativement $2\Phi(t) - 1$: plus précisément, $\mathbb{P}(A_n(t)) = 2\Phi(t) - 1 + \varepsilon_n(t)$ avec $\sup_{t \in \mathbf{R}} |\varepsilon_n(t)| \rightarrow 0$. Ceci se réécrit

$$\mathbb{P} \left(\frac{S_n}{n} - t \sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{S_n}{n} + t \sqrt{\frac{p(1-p)}{n}} \right) = 2\Phi(t) - 1 + \varepsilon_n(t).$$

On ignore la valeur de p et donc à fortiori celle de $p(1-p)$. Toutefois, on peut majorer $p(1-p)$ par $1/4$ et on a $\mathbb{P}(B_n(t)) \geq 2\Phi(t) - 1 + \varepsilon_n(t)$ avec

$$B_n(t) = \left\{ \omega \in \Omega : \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leq p \leq \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}} \right\}.$$

En pratique, n est fixé et on a observé des valeurs numériques explicites x_1, \dots, x_n que l'on interprète comme les valeurs de $X_1(\omega), \dots, X_n(\omega)$ pour un même ω tiré au sort suivant \mathbb{P} : $m_n = (x_1 + \dots + x_n)/n$ représente alors $S_n(\omega)/n$. Lorsqu'on propose, pour le paramètre p , l'intervalle de confiance

$$I_n(t) = \left[m_n - \frac{t}{2\sqrt{n}}, m_n + \frac{t}{2\sqrt{n}} \right]$$

on fait le pari que le ω observé se trouve dans l'ensemble $B_n(t)$. La probabilité de gagner ce pari est minorée par $2\Phi(t) - 1 + \varepsilon_n(t)$. On dit que $I_n(t)$ est un intervalle de confiance de p avec un niveau de confiance d'au moins $2\Phi(t) - 1 + \varepsilon_n(t)$. Bien évidemment, en pratique, on oublie le $\varepsilon_n(t)$ et on détermine t à l'aide d'une tabulation de Φ . Par exemple pour un niveau de confiance de 95%, on doit avoir $\Phi(t) = 1,95/2$; on obtient $t \simeq 1,96$ ce qui donne l'intervalle $\left[m_n - \frac{1,96}{2\sqrt{n}}, m_n + \frac{1,96}{2\sqrt{n}} \right]$ au niveau de confiance 95%.

On utilise souvent une variante de cette méthode pour obtenir des intervalles de confiance de longueur plus petite. Au lieu de majorer $p(1-p)$ par $1/4$, on remplace $p(1-p)$ par un estimateur. Puisque $M_n(\omega) = S_n(\omega)/n$ converge vers p pour presque tout ω , on remplace $p(1-p)$ par $M_n(\omega)(1-M_n(\omega))$. En pratique on dispose de la valeur m_n et on propose, au niveau de confiance 95%, l'intervalle

$$\left[m_n - 1,96\sqrt{\frac{m_n(1-m_n)}{n}}, m_n + 1,96\sqrt{\frac{m_n(1-m_n)}{n}} \right].$$

2. Vecteurs Gaussiens.

Rappelons tout d'abord qu'une variable aléatoire réelle X suit la loi normale (ou gaussienne) de moyenne $m \in \mathbf{R}$ et de variance $\sigma^2 \geq 0$, notée $\mathcal{N}(m, \sigma^2)$, si :

- elle a pour densité la fonction $x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ lorsque $\sigma^2 > 0$;
- elle est presque sûrement égale à m dans le cas où $\sigma^2 = 0$.

Si X a pour loi $\mathcal{N}(m, \sigma^2)$, $\mathbb{E}[X] = m$, $\mathbb{V}(X) = \sigma^2$ et on dit que X est une variable gaussienne. On vérifie facilement qu'une v.a.r. X est gaussienne si et seulement si

$$\forall t \in \mathbf{R}, \quad \varphi_X(t) = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

Avant d'étudier les vecteurs gaussiens, précisons quelques notations. On se place dans la base canonique $(e_i)_{i \leq d}$ de \mathbf{R}^d . Si X est une variable aléatoire à valeurs dans \mathbf{R}^d , on considérera toujours X comme un vecteur colonne (même si X est écrit en ligne pour des raisons typographiques) i.e., notant u^* la transposée de u , $X = (X_1, \dots, X_d)^*$.

Lorsque X appartient à L^2 , c'est à dire quand toutes ses composantes sont de carré intégrable, l'espérance du vecteur X est simplement le vecteur $m = \mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])^*$ et la matrice de covariance de X , notée $\Gamma = \Gamma_X$ si besoin – ou encore $\text{Cov}(X)$ est la matrice dont les coefficients sont

$$\Gamma_{i,j} = \text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])], \quad 1 \leq i \leq d, 1 \leq j \leq d.$$

On peut voir facilement que $\Gamma = \text{Cov}(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^*]$.

Si A est une matrice réelle de taille $q \times d$ et b un vecteur de \mathbf{R}^q , le vecteur aléatoire $Y = AX + b$ appartient à L^2 dès qu'il en est de même pour X . La linéarité de l'espérance conduit facilement aux formules suivantes

$$\mathbb{E}[AX + b] = A\mathbb{E}[X] + b, \quad \text{Cov}(AX + b) = ACov(X)A^*$$

qui généralisent celles du cas réel. En particulier, si $u \in \mathbf{R}^d$, la variable réelle u^*X a pour moyenne $\mathbb{E}[u^*X] = u^*\mathbb{E}[X]$ et pour variance $\mathbb{V}(u^*X) = u^*\text{Cov}(X)u$. Par conséquent, $\text{Cov}(X)$ est une matrice symétrique semi-définie positive.

Ces notations étant fixées, passons à la définition d'un vecteur gaussien.

Définition (Vecteur Gaussien). Soit X une application de Ω dans \mathbf{R}^d . X est un vecteur gaussien de \mathbf{R}^d si, pour tout $t \in \mathbf{R}^d$, t^*X est une variable aléatoire réelle gaussienne.

Remarque. Si X est un vecteur gaussien, alors pour tout $i = 1, \dots, d$, X_i est une v.a.r. gaussienne. En particulier, X appartient à L^2 .

Par contre, la réciproque est fautive : si X et ε sont deux v.a.r. indépendantes, X de loi $\mathcal{N}(0, 1)$ et $\mathbb{P}(\varepsilon = \pm 1) = 1/2$, alors εX suit la loi $\mathcal{N}(0, 1)$ mais le couple $(X, Y)^*$ n'est pas un vecteur gaussien ; revoir l'exemple page 34 du Chapitre II.

Néanmoins, si les variables réelles X_i , $i = 1, \dots, d$, sont gaussiennes et indépendantes alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien. En effet, si $t \in \mathbf{R}^d$, $t^*X = \sum_{i=1}^d t_i X_i$ est une variable gaussienne comme somme de variables réelles indépendantes et gaussiennes ; cf. Chapitre II.

Théorème 5. X est un vecteur gaussien de \mathbf{R}^d si et seulement si sa fonction caractéristique est de la forme

$$t \longmapsto \exp \left\{ it^*m - \frac{t^*\Gamma t}{2} \right\},$$

où $m \in \mathbf{R}^d$ et Γ est une matrice réelle symétrique semi-définie positive.

On a, dans ce cas, $m = \mathbb{E}[X]$ et $\Gamma = \text{Cov}(X)$. On dit que X suit la loi $\mathcal{N}(m, \Gamma)$.

Démonstration. Si X est un vecteur gaussien, t^*X est une v.a.r. gaussienne pour tout $t \in \mathbf{R}^d$. On a donc

$$\varphi_X(t) = \mathbb{E} \left[e^{it^*X} \right] = \varphi_{t^*X}(1) = \exp \left\{ i\mathbb{E}[t^*X] - \frac{\mathbb{V}(t^*X)}{2} \right\} = \exp \left\{ it^*\mathbb{E}[X] - \frac{t^*\text{Cov}(X)t}{2} \right\}.$$

Réciproquement, si X a pour fonction caractéristique $t \longmapsto \exp \left\{ it^*m - \frac{t^*\Gamma t}{2} \right\}$ alors, pour tout $c \in \mathbf{R}$,

$$\varphi_{t^*X}(c) = \varphi_X(ct) = \exp \left\{ it^*m c - \frac{t^*\Gamma t c^2}{2} \right\}.$$

Pour tout $t \in \mathbf{R}^d$, t^*X est donc une v.a.r. gaussienne de moyenne t^*m et de variance $t^*\Gamma t$. X est un vecteur gaussien et, pour tout $t \in \mathbf{R}^d$, $t^*m = t^*\mathbb{E}[X]$, $t^*\Gamma t = t^*\text{Cov}(X)t$. Par suite, $m = \mathbb{E}[X]$; prenant, $t = e_i + \alpha e_j$, on obtient, les matrices Γ et $\text{Cov}(X)$ étant symétriques, $\Gamma_{i,i} + 2\alpha\Gamma_{i,j} + \alpha^2\Gamma_{j,j} = \text{Cov}(X)_{i,i} + 2\alpha\text{Cov}(X)_{i,j} + \alpha^2\text{Cov}(X)_{j,j}$ ce qui montre que les matrices Γ et $\text{Cov}(X)$ sont égales. \square

Remarque (Importante). Il faut bien retenir du résultat précédent que la loi d'un vecteur gaussien est entièrement déterminée par sa moyenne et sa matrice de covariance.

En particulier, on peut lire l'indépendance des composantes d'un vecteur gaussien sur sa matrice de covariance. En effet si $X = (X_1, \dots, X_d)^*$ est un vecteur gaussien, les variables X_1, \dots, X_d sont indépendantes si et seulement si la matrice de covariance de X , Γ , est diagonale. Il suffit de regarder la fonction caractéristique de X ; en effet, comme $\varphi_{X_i}(t_i) = \varphi_X(t_i e_i)$, nous avons

$$\forall t \in \mathbf{R}^d, \quad \varphi_X(t) = \varphi_{X_1}(t_1) \dots \varphi_{X_d}(t_d) \exp \left\{ - \sum_{1 \leq k < l \leq d} \Gamma_{k,l} t_k t_l \right\}.$$

X_1, \dots, X_d sont donc indépendantes si et seulement si $\Gamma_{k,l} = \Gamma_{l,k} = 0$ pour $k < l$ c'est à dire si et seulement si Γ est diagonale.

Si $X_i, i = 1, \dots, d$ sont des variables i.i.d. suivant la loi $\mathcal{N}(0, 1)$, $X = (X_1, \dots, X_d)^*$ est un vecteur gaussien (voir la remarque précédente ou sa fonction caractéristique); sa moyenne est nulle et $\text{Cov}(X) = I_d$: X suit la loi $\mathcal{N}(0, I_d)$. D'après les résultats sur l'indépendance, X a pour densité la fonction

$$x \mapsto \frac{1}{\sqrt{(2\pi)^d}} \exp \left\{ - \frac{|x|^2}{2} \right\}.$$

Proposition 6. Soient X un vecteur gaussien de \mathbf{R}^d , $b \in \mathbf{R}^q$ et A une matrice réelle $q \times d$. Alors $Y = AX + b$ est un vecteur gaussien de \mathbf{R}^q de moyenne $A\mathbb{E}[X] + b$ et de matrice de covariance $ACov(X)A^*$.

Démonstration. Soit $u \in \mathbf{R}^q$. $u^*(Y - b) = u^*AX = (A^*u)^*X$ est une v.a.r. gaussienne puisque X est un vecteur gaussien; il en va de même de $u^*Y = u^*(Y - b) + u^*b$. Y est donc un vecteur gaussien de \mathbf{R}^q . La moyenne et la covariance résultent des formules rappelées au début du paragraphe. \square

Proposition 7. Soient X un vecteur gaussien, A et B deux matrices réelles de tailles respectives $q \times d$ et $r \times d$. AX et BX sont indépendants si et seulement si $ACov(X)B^* = 0$.

Démonstration. Notons $m = \mathbb{E}[X]$ et $\Gamma = \text{Cov}(X)$. Soit C la matrice réelle de taille $(q + r) \times d$ $C = \begin{pmatrix} A \\ B \end{pmatrix}$ et Y le vecteur aléatoire de \mathbf{R}^{q+r} , $Y = CX = \begin{pmatrix} AX \\ BX \end{pmatrix}$. D'après la Proposition 6, Y est un vecteur gaussien et on a $\mathbb{E}[Y] = Cm$, $\text{Cov}(Y) = C\Gamma C^*$. On a, via le Théorème 5,

$$\forall u \in \mathbf{R}^{q+r}, \quad \varphi_Y(u) = \exp \left\{ iu^*Cm - \frac{1}{2}u^*C\Gamma C^*u \right\}.$$

Si on note s les q premières composantes de u et t les r dernières c'est à dire si on écrit $u = \begin{pmatrix} s \\ t \end{pmatrix}$, on a $u^*C = s^*A + t^*B$, et par suite

$$u^*C\Gamma C^*u = s^*A\Gamma A^*s + t^*B\Gamma B^*t + s^*A\Gamma B^*t + t^*B\Gamma A^*s = s^*A\Gamma A^*s + t^*B\Gamma B^*t + 2s^*A\Gamma B^*t.$$

Finalement, nous avons

$$\varphi_Y(u) = \exp \left\{ is^*Am - \frac{1}{2}s^*A\Gamma A^*s \right\} \exp \left\{ it^*Bm - \frac{1}{2}t^*B\Gamma B^*t \right\} \exp \{-s^*A\Gamma B^*t\}.$$

Or AX et BX sont également des vecteurs gaussiens et d'après le Théorème 5, l'égalité précédente se réécrit

$$\varphi_Y(u) = \varphi_{AX}(s) \varphi_{BX}(t) \exp \{-s^*A\Gamma B^*t\}.$$

Par conséquent, AX et BX sont indépendants si et seulement si, pour tous $s \in \mathbf{R}^q$, $t \in \mathbf{R}^r$, $s^*A\Gamma B^*t = 0$ c'est à dire si et seulement si $A\Gamma B^* = 0$. \square

Remarque. Si X et Y sont deux vecteurs gaussiens indépendants à valeurs dans \mathbf{R}^q et \mathbf{R}^r , alors le vecteur $Z = (X^*, Y^*)^*$ est un vecteur gaussien de \mathbf{R}^{q+r} . En effet, conservant les notations précédentes, si $u \in \mathbf{R}^{q+r}$, $u^*Z = s^*X + t^*Y$ est une variable réelle gaussienne comme somme de deux v.a.r. gaussiennes indépendantes.

Nous finissons en montrant que si $m \in \mathbf{R}^d$ et si Γ est une matrice symétrique semi-définie positive, alors il existe un vecteur gaussien X de moyenne m et de matrice de covariance Γ .

Théorème 8. Soient $m \in \mathbf{R}^d$ et Γ une matrice réelle $d \times d$ symétrique et semi-définie positive.

1. Il existe un vecteur gaussien X à valeurs dans \mathbf{R}^d de loi $\mathcal{N}(m, \Gamma)$ c'est à dire tel que $\mathbb{E}[X] = m$ et $\text{Cov}(X) = \Gamma$.
2. X possède une densité si et seulement si Γ est inversible. Dans ce cas, X a pour densité

$$\forall x \in \mathbf{R}^d, \quad p(x) = \frac{1}{\sqrt{(2\pi)^d \det \Gamma}} \exp \left\{ -\frac{(x-m)^* \Gamma^{-1} (x-m)}{2} \right\}.$$

Si non, X est porté par $m + (\ker \Gamma)^\perp$.

3. Il existe $\alpha > 0$ tel que $\mathbb{E} \left[e^{\alpha |X|^2} \right] < +\infty$.

Démonstration. Soient Y_1, \dots, Y_d des variables indépendantes de loi $\mathcal{N}(0, 1)$; notons Y le vecteur $Y = (Y_1, \dots, Y_d)^*$. Y est un vecteur gaussien tel que $\mathbb{E}[Y] = 0$ et $\text{Cov}(Y) = I_d$.

La matrice Γ étant symétrique, elle est diagonalisable et comme elle est semi-définie positive toutes les valeurs propres sont positives. Notons σ_i^2 $i = 1, \dots, d$ – $\sigma_i \geq 0$ – les valeurs propres comptées avec leur multiplicité : les σ_i ne sont pas nécessairement distincts. Il existe une matrice orthogonale (les colonnes sont des vecteurs propres) A telle que $\Gamma = A \Sigma^2 A^*$ où Σ est la matrice diagonale formée des σ_i .

- Soit $X = m + A \Sigma Y$. X est un vecteur gaussien d'après la Proposition 6. De plus, on a $\mathbb{E}[X] = A \Sigma \mathbb{E}[Y] + m = m$, $\text{Cov}(X) = A \Sigma \text{Cov}(Y) \Sigma^* A^* = A \Sigma^2 A^* = \Gamma$.

- Supposons que Γ est inversible ce qui signifie que toutes les valeurs propres sont strictement positives. Y a pour densité sur \mathbf{R}^d la fonction $y \mapsto (2\pi)^{-\frac{d}{2}} e^{-\frac{|y|^2}{2}}$; si $f : \mathbf{R}^d \rightarrow \mathbf{R}_+$ est borélienne, nous avons

$$\mathbb{E}[f(X)] = \mathbb{E}[f(m + A \Sigma Y)] = \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbf{R}^d} f(m + A \Sigma y) e^{-\frac{|y|^2}{2}} dy.$$

Effectuons, le changement de variable $x = m + A \Sigma y$, $y = \Sigma^{-1} A^{-1} (x - m)$. On a alors,

$$\mathbb{E}[f(X)] = \frac{1}{\sqrt{(2\pi)^d}} \int_{\mathbf{R}^d} f(x) \exp \left\{ -\frac{|\Sigma^{-1} A^{-1} (x - m)|^2}{2} \right\} |\det \Sigma^{-1} A^{-1}| dx.$$

Remarquons pour conclure que, A étant orthogonale, $|\det \Sigma^{-1} A^{-1}| = (\det \Gamma)^{-\frac{1}{2}}$ et que, puisque $A^* = A^{-1}$,

$$\left| \Sigma^{-1} A^{-1} (x - m) \right|^2 = (x - m)^* A \Sigma^{-1} \Sigma^{-1} A^{-1} (x - m) = (x - m)^* \Gamma^{-1} (x - m).$$

Par conséquent, nous avons

$$\mathbb{E}[f(X)] = \frac{1}{\sqrt{(2\pi)^d \det \Gamma}} \int_{\mathbf{R}^d} f(x) \exp \left\{ -\frac{(x-m)^* \Gamma^{-1} (x-m)}{2} \right\} dx,$$

ce qui montre que X a la densité requise.

Si Γ n'est pas inversible, considérons u_1, \dots, u_r ($1 \leq r \leq d$) une base orthonormale de $\ker \Gamma$. On a donc

$$\{X \in m + (\ker \Gamma)^\perp\} = \bigcap_{i=1}^r \{u_i^*(X - m) = 0\}.$$

Or, pour tout $u \in \ker \Gamma$, on a $\mathbb{V}(u^*(X - m)) = u^* \text{Cov}(X)u = 0$. Par conséquent, la variable réelle $u^*(X - m)$ est presque sûrement égale à sa moyenne qui est nulle. Par suite, $\mathbb{P}(u_i^*(X - m) = 0) = 1$ pour tout $i = 1, \dots, r$, et donc $\mathbb{P}(u_i^*(X - m) = 0, \forall i = 1, \dots, r) = 1$. Dans ce cas, X ne possède pas de densité puisqu'il est porté par un espace affine de dimension strictement inférieure à d donc négligeable pour la mesure de Lebesgue λ_d .

• Rappelons que si G est une v.a.r. de loi $\mathcal{N}(0, 1)$, on a, pour $s < 1/2$, via le changement de variable $z = x\sqrt{1 - 2s}$,

$$\beta(s) := \mathbb{E}[e^{sG^2}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{sx^2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2(1-2s)}{2}} dx = \frac{1}{\sqrt{1-2s}},$$

et pour $s \geq 1/2$, $\beta(s) = +\infty$. On a alors, puisque A est orthogonale,

$$\mathbb{E}[e^{\alpha|X-m|^2}] = \mathbb{E}[e^{\alpha|A\Sigma Y|^2}] = \mathbb{E}[e^{\alpha|\Sigma Y|^2}] = \mathbb{E}[e^{\alpha(\sigma_1^2 Y_1^2 + \dots + \sigma_d^2 Y_d^2)}],$$

et comme les variables Y_i sont i.i.d. suivant la loi $\mathcal{N}(0, 1)$,

$$\mathbb{E}[e^{\alpha|X-m|^2}] = \prod_{i=1}^d \mathbb{E}[e^{\alpha\sigma_i^2 Y_i^2}] = \prod_{i=1}^d \beta(\alpha\sigma_i^2),$$

qui est finie dès que $\alpha \max_{i \leq d} \sigma_i^2 < \frac{1}{2}$. Pour finir, notons que, $|X|^2 \leq 2|X - m|^2 + 2|m|^2$ et donc que

$$\mathbb{E}[e^{\alpha|X|^2}] = e^{2\alpha|m|^2} \mathbb{E}[e^{2\alpha|X-m|^2}].$$

Ceci montre l'existence d'un $\alpha > 0$ répondant à la question. \square

Définition. Soit $X = (X_1, \dots, X_n)$ un n -échantillon de loi $\mathcal{N}(0, 1)$: les variables X_i sont i.i.d. suivant la loi $\mathcal{N}(0, 1)$. La loi de $|X|^2 = X_1^2 + \dots + X_n^2$ s'appelle la loi du chi-deux à n degrés de liberté ; on la note χ_n^2 .

Remarque. On appelle loi gamma de paramètres $\alpha > 0$ et $s > 0$, notée $\Gamma_{\alpha, s}$, la probabilité sur \mathbf{R} de densité

$$\gamma_{\alpha, s}(x) = \frac{\alpha^s}{\Gamma(s)} x^{s-1} e^{-\alpha x} \mathbf{1}_{\mathbf{R}_+^*}(x), \quad x \in \mathbf{R}, \quad \text{avec} \quad \Gamma(s) = \int_0^{+\infty} x^{s-1} e^{-x} dx.$$

On montre facilement que si X et Y sont indépendantes de lois respectives $\Gamma_{\alpha, s}$ et $\Gamma_{\alpha, t}$, $X + Y$ suit la loi $\Gamma_{\alpha, s+t}$.

Lemme 9. La loi du chi-deux à n degrés de liberté est la loi $\Gamma_{\frac{1}{2}, \frac{n}{2}}$ donc de densité

$$x \longmapsto \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{\mathbf{R}_+^*}(x).$$

Démonstration. Si G a pour loi $\mathcal{N}(0, 1)$, la variable G^2 suit la loi $\Gamma_{\frac{1}{2}, \frac{1}{2}}$. En effet, si $f : \mathbf{R} \rightarrow \mathbf{R}^+$ est borélienne, nous avons, via $y = x^2$,

$$\mathbb{E} \left[f \left(G^2 \right) \right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x^2) e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} f(x^2) e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} f(y) e^{-\frac{y}{2}} \frac{dy}{2\sqrt{y}}.$$

G^2 suit bien la loi $\Gamma_{\frac{1}{2}, \frac{1}{2}}$ puisque $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Si X est un n -échantillon de loi $\mathcal{N}(0, 1)$, $|X|^2$ suit donc la loi $\Gamma_{\frac{1}{2}, \frac{n}{2}}$ d'après la remarque précédente. \square

Proposition 10. Soient X un n -échantillon de loi $\mathcal{N}(0, 1)$ et $u \in \mathbf{R}^n$ tel que $|u| = 1$. Alors $Y = X - u^* X u$ et $u^* X$ sont indépendants. En outre, $|Y|^2$ suit la loi χ_{n-1}^2 .

En particulier, la moyenne et la variance empiriques de X

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2,$$

sont indépendants et $(n-1)V_n$ suit la loi du chi-deux à $n-1$ degrés de liberté.

Démonstration. X est un vecteur gaussien de \mathbf{R}^n de loi $\mathcal{N}(0, I_n)$ et $Y = (I_n - uu^*)X$. Pour vérifier l'indépendance des variables Y et $u^* X$, il suffit de montrer, d'après la Proposition 7, que $(I_n - uu^*)\text{Cov}(X)u = 0$. Or $(I_n - uu^*)\text{Cov}(X)u = u(1 - |u|^2) = 0$.

D'après la Proposition 6, Y est un vecteur gaussien centré dont la matrice de covariance est $(I_n - uu^*)^2 = I_n - uu^*$. Considérons u_1, \dots, u_{n-1} une base orthonormale de $\text{vect}(u)^\perp$. La matrice $P = (u_1, \dots, u_{n-1}, u)$ est alors orthogonale et donc $|Y| = |P^* Y|$. Toujours, via la Proposition 6, $P^* Y$ est un vecteur gaussien centré de covariance $P^*(I_n - uu^*)P = I_n - P^* uu^* P$. Or on a $u^* P = (u^* u_1, \dots, u^* u_n, u^* u) = (0, \dots, 0, 1)$, de sorte que

$$P^*(I_n - uu^*)P = \begin{pmatrix} I_{n-1} & 0 \\ 0 & 0 \end{pmatrix};$$

$|Y|^2$ suit la loi χ_{n-1}^2 .

On obtient le résultat sur la moyenne et la variance empiriques en prenant pour u le vecteur dont toutes les coordonnées sont $1/\sqrt{n}$; en effet, nous avons

$$\sqrt{n}M_n = u^* X, \quad \text{et} \quad (n-1)V_n = |Y_n|^2, \quad \text{où} \quad Y_n = X - u^* X u,$$

ce qui termine la démonstration. \square

Remarque. De la même manière, si X un vecteur gaussien centré de \mathbf{R}^n de matrice de covariance $\Gamma = I_n - uu^*$ où $u \in \mathbf{R}^n$ avec $|u| = 1$. Alors $|X|^2$ suit la loi χ_{n-1}^2 .

3. TCL multidimensionnel.

Nous allons, pour clore ce chapitre sur le théorème limite central, en donner une version valable pour les variables à valeurs dans \mathbf{R}^d .

Commençons par remarquer que si $(X_n)_{n \in \mathbf{N}^*}$ est une suite i.i.d. de variables à valeurs dans \mathbf{R}^d telle que X_1 soit intégrable alors nous avons convergence presque sûre de la suite $(M_n)_{n \in \mathbf{N}^*}$ vers

le vecteur $\mathbb{E}[X_1]$. Il suffit d'appliquer la loi forte des grands nombres à chacune des composantes de la suite $(X_n)_{n \in \mathbf{N}^*}$.

Nous allons voir que le théorème limite central se généralise également au cas de variables à valeurs dans \mathbf{R}^d .

Théorème 11. *Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables à valeurs dans \mathbf{R}^d , i.i.d. avec $X_1 \in L^2$. On note $m = \mathbb{E}[X_1]$ et Γ la matrice de covariance de X_1 .*

Alors la suite de variables $T_n = \sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right)$ converge en loi, lorsque $n \rightarrow +\infty$, vers un vecteur gaussien G de loi $\mathcal{N}(0, \Gamma)$.

Démonstration. Comme dans le cas réel, nous allons appliquer le théorème de Paul Lévy. Soit $t \in \mathbf{R}^d$. On a

$$\varphi_{T_n}(t) = \mathbb{E} \left[e^{it^*T_n} \right] = \varphi_{t^*T_n}(1), \quad \text{et} \quad t^*T_n = \sqrt{n} \left(\frac{t^*X_1 + \dots + t^*X_n}{n} - t^*m \right).$$

Les variables réelles $(t^*X_n)_{n \in \mathbf{N}^*}$ sont i.i.d., t^*X_1 est de carré intégrable et $\mathbb{E}[t^*X_1] = t^*m$, $\mathbb{V}(t^*X_1) = t^*\Gamma t$. D'après le Théorème 2, t^*T_n converge en loi vers une variable réelle de loi $\mathcal{N}(0, t^*\Gamma t)$. Par conséquent,

$$\lim_{n \rightarrow +\infty} \varphi_{T_n}(t) = \lim_{n \rightarrow +\infty} \varphi_{t^*T_n}(1) = \exp \left\{ -\frac{t^*\Gamma t}{2} \right\}.$$

Ceci termine la démonstration puisque nous avons convergence simple sur \mathbf{R}^d de φ_{T_n} vers la fonction caractéristique de la loi $\mathcal{N}(0, \Gamma)$. \square

Donnons un exemple d'utilisation de ce résultat.

Exemple (Loi multinomiale). Soient μ une probabilité et B_1, \dots, B_r des boréliens deux à deux disjoints de \mathbf{R}^d tels que $\sum_{k=1}^r \mu(B_k) = 1$. Si (X_1, \dots, X_n) est un n -échantillon de loi μ on pose, pour $k = 1, \dots, r$, $N_n(B_k) = \sum_{i=1}^n \mathbf{1}_{B_k}(X_i)$ et on note N_n le vecteur de \mathbf{R}^r de coordonnées $N_n(B_k)$.

N_n suit la loi multinomiale de paramètres n et $\mu(B_k)$, $k = 1, \dots, r$:

$$\mathbb{P}(N_n(B_1) = n_1, \dots, N_n(B_r) = n_r) = \frac{n!}{n_1! \dots n_r!} \mu(B_1)^{n_1} \dots \mu(B_r)^{n_r}$$

dans le cas où $n = n_1 + \dots + n_r$ avec les $n_k \in \mathbf{N}$ et 0 dans le cas contraire et avec les conventions $0! = 1$, $0^0 = 1$.

Une démonstration possible consiste en une récurrence sur la taille de l'échantillon. Commençons par remarquer que, $0 \leq N_n(B_k) \leq n$ pour tout $k = 1, \dots, r$, et que, comme $\mu(B_1) + \dots + \mu(B_r) = 1$, $N_n(B_1) + \dots + N_n(B_r) = n$ presque sûrement. Pour $n = 1$, le résultat est immédiat. Supposons la formule vraie pour n et démontrons-la à l'ordre $n + 1$. Soient donc $n_1 + \dots + n_r = n + 1$ avec les $n_k \in \mathbf{N}$. On a, comme $\mathbb{P}(X_{n+1} \in \cup_{k \leq r} B_k) = 1$ et comme les B_k sont disjoints,

$$\begin{aligned} & \mathbb{P}[N_{n+1}(B_1) = n_1, \dots, N_{n+1}(B_r) = n_r] \\ &= \sum_{k=1}^r \mathbb{P}[N_{n+1}(B_1) = n_1, \dots, N_{n+1}(B_r) = n_r, X_{n+1} \in B_k] \\ &= \sum_{k=1}^r \mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k), X_{n+1} \in B_k], \end{aligned}$$

avec $n_l(k) = n_l$ si $k \neq l$ et $n_k(k) = n_k - 1$. D'autre part, pour $k = 1, \dots, r$, nous avons, vu l'indépendance des variables X_i ,

$$\begin{aligned} \mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k), X_{n+1} \in B_k] \\ &= \mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k)] \mathbb{P}(X_{n+1} \in B_k) \\ &= \mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k)] \mu(B_k). \end{aligned}$$

D'après l'hypothèse de récurrence,

$$\mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k)] = \frac{n!}{n_1(k)! \dots n_r(k)!} \mu(B_1)^{n_1(k)} \dots \mu(B_r)^{n_r(k)}$$

si $n_k(k) \geq 0$ et 0 si $n_k = 0$. Dans tous les cas, on a

$$\mathbb{P}[N_n(B_1) = n_1(k), \dots, N_n(B_r) = n_r(k)] \mu(B_k) = n_k \frac{n!}{n_1! \dots n_r!} \mu(B_1)^{n_1} \dots \mu(B_r)^{n_r},$$

et par conséquent,

$$\mathbb{P}[N_{n+1}(B_1) = n_1, \dots, N_{n+1}(B_r) = n_r] = \frac{n!}{n_1! \dots n_r!} \mu(B_1)^{n_1} \dots \mu(B_r)^{n_r} \sum_{k=1}^r n_k$$

ce qui donne le résultat puisque $n_1 + \dots + n_r = n + 1$.

Désignons par μ_n la mesure empirique de l'échantillon (X_1, \dots, X_n) :

$$\mu_n = \frac{1}{n} (\delta_{X_1} + \dots + \delta_{X_n}).$$

Pour tout borélien B de \mathbf{R}^d , $\mu_n(B) = N_n(B)/n$.

La loi forte des grands nombres nous apprend que $\mu_n(B_k)$ converge presque sûrement vers $\mu(B_k)$ pour tout $k = 1, \dots, r$.

Si on applique le TCL multidimensionnel, on obtient la convergence en loi, lorsque n tend vers $+\infty$ du vecteur T_n de coordonnées

$$\sqrt{n} [\mu_n(B_k) - \mu(B_k)], \quad k = 1, \dots, r,$$

vers un vecteur gaussien centré G de matrice de covariance Γ donnée par : pour $1 \leq k, l \leq r$,

$$\begin{aligned} \Gamma_{k,l} &= \text{Cov}(\mathbf{1}_{B_k}(X_1), \mathbf{1}_{B_l}(X_1)) \\ &= \mathbb{E}[\mathbf{1}_{B_k}(X_1)\mathbf{1}_{B_l}(X_1)] - \mathbb{E}[\mathbf{1}_{B_k}(X_1)]\mathbb{E}[\mathbf{1}_{B_l}(X_1)] \\ &= \mu(B_k \cap B_l) - \mu(B_k)\mu(B_l) \end{aligned}$$

soit, $\Gamma_{k,l} = \mu(B_k) \mathbf{1}_{\{0\}}(k-l) - \mu(B_k)\mu(B_l)$.

Supposons à présent que, $\mu(B_k) > 0$ pour tout $k = 1, \dots, r$ et notons Λ la matrice diagonale formée des réels $\mu(B_1)^{-\frac{1}{2}}, \dots, \mu(B_r)^{-\frac{1}{2}}$. L'application $x \mapsto \Lambda x$, de \mathbf{R}^r dans lui-même, est continue et par conséquent, ΛT_n converge en loi vers ΛG . D'après la Proposition 6, ΛG est un vecteur gaussien de covariance $\Lambda \Gamma \Lambda$. Or $\Lambda \Gamma \Lambda = I_r - uu^*$ où u est le vecteur de coordonnées $\sqrt{\mu(B_k)}$, $k = 1, \dots, r$. Comme $|u| = \sum_{k=1}^r \mu(B_k) = 1$, $|\Lambda G|^2$ suit la loi χ_{r-1}^2 . L'application

$x \mapsto |x|^2$ est continue sur \mathbf{R}^r ; par conséquent $|\Lambda T_n|^2$ converge en loi vers $|\Lambda G|^2$ qui suit la loi χ_{r-1}^2 : la suite de variables aléatoires

$$\sum_{k=1}^r \frac{n}{\mu(B_k)} [\mu_n(B_k) - \mu(B_k)]^2 = \sum_{k=1}^r \frac{[N_n(B_k) - n\mu(B_k)]^2}{n\mu(B_k)}$$

converge en loi, lorsque $n \rightarrow +\infty$, vers la loi du chi-deux à $r - 1$ degrés de liberté, χ_{r-1}^2 .

Nous venons d'établir le résultat suivant :

Proposition 12. *Soient p_1, \dots, p_r des réels strictement positifs tels que $p_1 + \dots + p_r = 1$. Si N_n est une variable aléatoire de loi multinomiale de paramètres n et (p_1, \dots, p_r) alors*

$$\frac{(N_n^{(1)} - np_1)^2}{np_1} + \dots + \frac{(N_n^{(r)} - np_r)^2}{np_r}$$

converge en loi vers χ_{r-1}^2 .