



FaST

Probabilité III

Edoh KATCHEKPELE

Mousson 2020-2021

Table des matières

1	Convergence en loi-Théorème Central Limite	3
1.1	Convergence étroite	3
1.1.1	Rappels, notations et définitions	3
1.1.2	Convergence étroite	3
1.2	Convergence en loi	9
1.3	Le Théorème Centrale Limite	10
1.4	Exercices	13
2	Suites et séries de variables aléatoires réelles - Loi des grands nombres	14
2.1	Lemme de Borel-Cantelli	14
2.2	Divers modes de convergences	15
2.3	Convergence p.s. de séries de variables aléatoires indépendantes-Théorèmes li- mites forts	18
2.4	Loi des grand nombres (LGN)	20
2.4.1	Loi faible des grands nombres (LfGN)	20
2.4.2	Loi forte des grands nombres (LFGN)	21
2.5	Exercices	24
3	Vecteurs Gaussiens	26
3.1	Définition	26
3.2	Quelques propriétés	27
3.3	Représentation de lois normales sur $\mathbb{R}^n$	28
3.4	Singularité et absolue continuité des lois gaussiennes	28
3.5	Exercices	29
4	Introduction aux processus aléatoires	32
4.1	Processus de Poisson	32
4.1.1	Définition et premières propriétés	32
4.1.2	La seconde définition du processus de Poisson	33
4.2	Marche aléatoire sur $\mathbb{Z}$	34
4.3	Introduction aux chaînes de Markov	35
4.3.1	Définitions et exemples	35
4.3.2	Distribution stationnaire	38
4.4	Exercices	38

Ce polycopié est très récent. Veuillez m'excuser pour toutes les coquilles qu'il doit contenir. Je remercie d'avance tous ceux qui voudront bien me les signaler à l'adresse edohkatchek-pele@gmail.com.

Certaines parties du présent polycopié sont, en partie, inspirées des notes de cours du Prof. GNEYOU Kossi de l'Université de Lomé à qui j'exprime ici toute ma gratitude.

Chapitre 1

Convergence en loi-Théorème Central Limite

Ce chapitre introduit la notion convergence en loi puis traite d'un théorème important en théorie de probabilités, le TCL. Les notions de convergence vague, de convergence faible et de convergence étroite, y sont présentées.

1.1 Convergence étroite

1.1.1 Rappels, notations et définitions

Soit E un espace métrique localement compact et dénombrable à l'infini (i.e E possède un recouvrement au plus dénombrable de parties compactes e.g $E = \mathbb{R}^d$) et $\mathcal{B}$ sa tribu borélienne. Nous commençons par donner quelques notations utilisées tout au long du chapitre : notons par $\mathcal{M}$ (resp. $\mathcal{M}_1$) l'ensemble des mesures positives bornées (resp. l'ensemble des mesures de probabilités) définies sur l'espace mesurable $(E, \mathcal{B})$.

On note également

$C_k(E)$ l'espace vectoriel des fonctions sur E continues à support compact.

$C_0(E)$ l'espace vectoriel des fonctions continues sur E et nulles à l'infini (i.e f tend vers 0 à l'infini si, pour tout $\varepsilon > 0$, il existe un compact $\mathbf{K}$ tel que $|f(x)| < \varepsilon$ sur $\mathbf{K}^c$).

$C_b(E)$ l'espace vectoriel des fonctions continues et bornées sur E .

On a

$C_k(E) \subset C_0(E) \subset C_b(E)$ et on a l'égalité si E est compact.

Soit $(\mu_n)_n$ une suite d'éléments de $\mathcal{M}_1$ et $\mu \in \mathcal{M}_1$. On veut donner un sens à “ $(\mu_n)_n$ converge vers μ ”. Il semble naturel de demander que, pour tout $A \in \mathcal{B}$, $\mu_n(A) \rightarrow \mu(A)$ mais ceci est très contraignant. Par exemple, sur $\mathbb{R}$, si $\mu_n = \delta_1$ et $\mu = \delta_0$, on a $\mu_n([0, 1]) = 1$ et $\mu([0, 1]) = 0$ et donc, en ce sens, μ_n ne converge pas vers μ . C'est pourquoi on introduit la notion de convergence étroite.

1.1.2 Convergence étroite

Définition 1.1.1

Soit $(\mu_n)_n$ une suite d'éléments de $\mathcal{M}$ et $\mu \in \mathcal{M}$. On dit que :

1. $(\mu_n)_n$ converge vaguement vers μ si $\forall f \in C_k(E)$, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.

2. $(\mu_n)_n$ converge faiblement vers μ si $\forall f \in C_0(E)$, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.
3. $(\mu_n)_n$ converge étroitement vers μ si $\forall f \in C_b(E)$, $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$.

Remarque 1.1.1

3) $\implies$ 2) $\implies$ 1), mais, il n'y a pas réciprocity en général (par exemple $n\delta_n$ tend vers 0 vaguement mais non faiblement, et δ_n tend vers 0 faiblement mais non étroitement).

Dans toute la suite, on notera par $\|f\| = \sup_{x \in E} |f(x)|$ et éventuellement par $\mu(f) = \int_E f d\mu$.

On rappelle qu'un espace vectoriel $\mathcal{H}$ est total si l'ensemble des combinaisons linéaires des éléments de $\mathcal{H}$ forme un sous-espace vectoriel dense.

Proposition 1.1.1

Soit $\mathcal{H}$ un sous ensemble total de $C_0(E)$, $(\mu_n, n \geq 0)$ une suite d'éléments de $\mathcal{M}_1$ et $\mu \in \mathcal{M}_1$.

Si $\forall f \in \mathcal{H}$ et $\int_E f d\mu_n \rightarrow \int_E f d\mu$,
alors μ_n converge vers μ étroitement.

Preuve

Montrons d'abord que, dans ces conditions, μ_n converge vers μ faiblement.

Soit V le s.e.v engendré par $\mathcal{H}$. On a $\bar{V} = C_0(E)$ et pour tout $g \in V$, $\int_E g d\mu_n \rightarrow \int_E g d\mu$. Soient $f \in C_0(E)$ et $g \in V$,

$$\begin{aligned} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq \left| \int_E f d\mu_n - \int_E g d\mu_n \right| + \left| \int_E g d\mu_n - \int_E g d\mu \right| + \left| \int_E g d\mu - \int_E f d\mu \right| \\ &\leq \left| \int_E g d\mu_n - \int_E g d\mu \right| + 2\|f - g\| \end{aligned}$$

On a donc $\limsup_n \left| \int_E f d\mu_n - \int_E f d\mu \right| \leq 2\|f - g\|$.

On donc peut choisir g assez voisin de f pour le terme à droite de la dernière inégalité soit petit (puisque cette dernière quantité est arbitrairement petite).

Ceci fait, on a, pour $f \in C_b(E)$ et $g \in C_k(E)$, $0 \leq g \leq 1$,

$$\begin{aligned} \left| \int_E f d\mu_n - \int_E f d\mu \right| &\leq \left| \int_E f d\mu_n - \int_E f g d\mu_n \right| + \left| \int_E f g d\mu_n - \int_E f g d\mu \right| + \left| \int_E f g d\mu - \int_E f d\mu \right| \\ &\leq \|f\| \left(1 - \int_E g d\mu_n \right) + \left| \int_E f g d\mu_n - \int_E f g d\mu \right| + \|f\| \left(1 - \int_E g d\mu \right). \end{aligned}$$

On a donc $\limsup_n \left| \int_E f d\mu_n - \int_E f d\mu \right| \leq 2\|f\| \left(1 - \int_E g d\mu \right)$.

Vu qu'il existe $g_n \in C_k(E)$, $0 \leq g_n \leq 1$, tels que $g_n \uparrow 1$ et qu'alors $\int_E g d\mu_n \uparrow \int_E 1 d\mu = 1$, $1 - \int_E g d\mu$ est arbitrairement petit et donc $\int_E f d\mu_n \rightarrow \int_E f d\mu$. $\square$

Les résultats suivants concernant les éléments de $\mathcal{M}$ sont à rappeler avant d'établir des critères de convergence étroite.

Théorème 1.1.1

Toute mesure finie sur $(E, \mathcal{B})$ est régulière c'est à dire, $\forall A \in \mathcal{B}$ et $\forall \varepsilon > 0$, il existe un fermé F et un ouvert G de E tels que $F \subset A \subset G$ et $\mu(G - F) < \varepsilon$.

Preuve

C'est le Théorème 1.1 Billingsley (1968) page 7. □

Ce Théorème implique alors que la mesure μ est déterminée par les valeurs $\mu(F)$, F fermé parcourant E .

Théorème 1.1.2

Soit F un fermé de E et soit $\varepsilon > 0$. Alors il existe une fonction $f \in C_b(E)$ telle que $f(x) = 1$ si $x \in F$, $f(x) = 0$ si $d(x, F) > \varepsilon$ et $0 \leq f(x) \leq 1 \ \forall x \in E$. (f pourra être prise uniformément continue.)

Preuve

Il suffit de prendre f telle que $f(x) = \varphi\left(\frac{1}{\varepsilon}d(x, F)\right)$, où φ est une fonction de $\mathbb{R}$ dans lui-même définie par

$$\varphi(t) = \begin{cases} 1 & \text{si } t \leq 0 \\ 1 - t & \text{si } 0 \leq t \leq 1 \\ 0 & \text{si } t \geq 1 \end{cases} . \quad \square$$

(Représenter le graphe de f sur $F = [a; b]$)

Le Théorème suivant fournit des critères de convergence étroite.

Théorème 1.1.3 (Portmanteau)

Soit $(\mu_n)_{n \geq 0}$ une suite d'éléments $\mathcal{M}$ et $\mu \in \mathcal{M}$. Les propositions suivantes sont équivalentes :

- (i) $(\mu_n)_{n \geq 0}$ converge étroitement vers μ .
- (ii) $\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$, pour toute fonction f uniformément continue et bornée sur E .
- (iii) $\limsup_n \mu_n(F) \leq \mu(F)$, pour tout fermé F de E .
- (iv) $\liminf_n \mu_n(G) \geq \mu(G)$, pour tout ouvert G de E .
- (v) $\lim_n \mu_n(A) = \mu(A)$, pour toute partie A de E telle que $\mu(\partial A) = 0$, ∂A étant la frontière de A (on dit que A est un ensemble de μ -continuité).

Preuve

Pour la preuve, nous allons établir les implications $(i) \implies (ii) \implies (iii) \implies (i)$ puis les équivalences $(iii) \iff (iv)$ et $(iii) \iff (v)$.

- $(i) \implies (ii)$ est évident (voir la définition même de la convergence étroite)

- (ii) $\implies$ (iii) : Soit F un fermé de E . Les ouverts $G = \{x \mid d(x, F) < \varepsilon\}$ (ε étant arbitrairement petit), décroissent vers F quand $\varepsilon \searrow 0$. On a alors $\mu(G) < \mu(F) + \delta$, d'après le Théorème 1.1.1.

D'autre part, en choisissant f comme dans la preuve du Théorème 1.1.2, on a f qui est uniformément continue sur E , $f(x) = 1$ sur F , $f(x) = 0$ sur G^c et $0 \leq f(x) \leq 1, \forall x \in E$. Par hypothèse, on a

$$\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu$$

mais,

$$\mu_n(F) = \int_F f d\mu_n \leq \int_E f d\mu_n$$

et

$$\int_E f d\mu = \int_G f d\mu \leq \mu(G) < \mu(F) + \delta.$$

D'où

$$\limsup_n \mu_n(F) \leq \lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu < \mu(F) + \delta, \forall \delta > 0$$

et comme δ est arbitraire, on a (iii).

- (iii) $\implies$ (i)

Soit $f \in C_b(E)$. Par homothétie on peut se ramener au cas où $0 < f < 1$ (on fait $f \rightarrow \frac{b-f}{2b}$, avec b tel que de norme $\|f\| < b$). Nous allons d'abord montrer que

$$\limsup_n \int_E f d\mu_n \leq \int_E f d\mu. \quad (1.1)$$

Supposons d'abord f positive et pour tout $k \in \mathbb{N}^*$ considérons les parties $F_i = f^{-1} \left(\left[\frac{i}{k}, +\infty \right) \right) = \left\{ x \mid \frac{i}{k} \leq f(x) \right\}$ de E , $i = 0, 1, \dots, k$.

Les parties F_i sont bien sùre fermées et recouvrent E (puisque les intervalles $\left[\frac{i-1}{k}, \frac{i}{k} \right]$, $i = 1, \dots, k$, recouvrent l'intervalle $[0; 1]$); on alors pour toute mesure $\mu \in \mathcal{M}$:

$$\sum_{i=1}^k \frac{i-1}{k} \mu \left\{ x \mid \frac{i-1}{k} \leq f(x) < \frac{i}{k} \right\} \leq \int_E f d\mu < \sum_{i=1}^k \frac{i}{k} \mu \left\{ x \mid \frac{i-1}{k} \leq f(x) < \frac{i}{k} \right\}.$$

La somme à droite de la dernière inégalité est égale à

$$\sum_{i=1}^k \frac{i}{k} (\mu(F_{i-1}) - \mu(F_i)) = \frac{1}{k} + \frac{1}{k} \sum_{i=1}^k \mu(F_i).$$

Ceci et une transformation similaire de la somme à gauche de la première inégalité donne

$$\frac{1}{k} \sum_{i=1}^k \mu(F_i) \leq \int_E f d\mu < \frac{1}{k} + \frac{1}{k} \sum_{i=1}^k \mu(F_i). \quad (1.2)$$

Comme par hypothèse on a $\limsup_n \mu_n(F_i) \leq \mu(F_i)$, pour tout $k \in \mathbb{N}^*$ et $i = 0, \dots, k$, alors en appliquant l'inégalité à droite de (1.2) à μ_n et celle de gauche à μ , on a

$$\limsup_n \int_E f d\mu_n < \frac{1}{k} + \frac{1}{k} \sum_{i=1}^k \limsup_n \mu_n(F_i) \leq \frac{1}{k} + \frac{1}{k} \sum_{i=1}^k \mu(F_i) \leq \frac{1}{k} + \int_E f d\mu.$$

En faisant tendre ensuite k vers $+\infty$, on obtient (1.1).

En appliquant (1.1) à $-f$ on obtient

$$\liminf_n \int_E f d\mu_n \geq \int_E f d\mu,$$

soit pour tout $f \in C_b(E)$, on a

$$\lim_{n \rightarrow +\infty} \int_E f d\mu_n = \int_E f d\mu.$$

- (iii) $\implies$ (iv)

On a pour tout ouvert G de E :

$$\begin{aligned} \liminf_n \mu_n(G) &= \liminf_n (\mu_n(E) - \mu_n(G^c)) \\ &= \liminf_n \mu_n(E) - \limsup_n \mu_n(G^c) \\ &\geq \mu(E) - \mu(G^c). \end{aligned}$$

L'implication réciproque (iv) $\implies$ (iii) se démontre de la même manière.

- (iii) $\implies$ (v)

Soit A une partie de E telle que $\mu(\partial(A)) = 0$. Comme $\overset{\circ}{A} \subset A \subset \bar{A}$, on a

$$\mu(\bar{A}) \geq \limsup_n \mu_n(\bar{A}) \geq \limsup_n \mu_n(A) \geq \liminf_n \mu_n(A) \geq \liminf_n \mu_n(\overset{\circ}{A}) \geq \mu(\overset{\circ}{A}).$$

Comme $\mu(\bar{A}) - \mu(\overset{\circ}{A}) = \mu(\partial(A)) = 0$, alors $\mu(\overset{\circ}{A}) = \mu(\bar{A})$. En conclusion la suite $(\mu_n(A))_{n \geq 1}$ a une limite supérieure qui coïncide avec sa limite inférieure, donc converge vers $\mu(\bar{A}) = \mu(\overset{\circ}{A})$, c'est à dire vers $\mu(A)$ car $\mu(\overset{\circ}{A}) \leq \mu(A) \leq \mu(\bar{A})$.

- (v) $\implies$ (iii)

Soit F un fermé de E . Pour tout $\delta > 0$, posons

$$F_\delta = \{x \mid d(x, F) \leq \delta\}.$$

On a $\partial(F_\delta) \subset \{x \mid d(x, F) = \delta\}$. Donc ces frontières sont disjointes pour des δ distincts. Ils sont alors de mesures nulles sauf pour un ensemble dénombrable d'entre eux.

On peut donc trouver une suite décroissante $(\delta_k)_{k \geq 1}$ de nombres réels strictement positifs, suite tendant vers 0, telle que pour tout k , $\mu(\partial(F_k)) = 0$, où $F_k = \{x \mid d(x, F) \leq \delta_k\}$.

Alors $\forall k \geq 1$

$$\limsup_n \mu_n(F) \leq \limsup_n \mu_n(F_k)$$

or d'après (v)

$$\limsup_n \mu_n(F_k) = \lim_n \mu_n(F_k) = \mu(F_k).$$

Pour tout $k \geq 1$,

$$\limsup_n \mu_n(F) \leq \mu(F_k).$$

Finalement F_k décroît lorsque $k \rightarrow +\infty$ et $\bigcap_{k \geq 1} F_k = \bar{F} = F$, donc

$$\inf_k \mu(F_k) = \lim_k \mu(F_k) = \mu(F),$$

d'où (iii). Ce qui achève la démonstration du Théorème 1.1.3. □

Remarque 1.1.2

Si la partie A n'est pas un ensemble de μ -continuité, la condition $\lim_{n \rightarrow +\infty} \mu_n(A) = \mu(A)$ ne saurait suffire pour garantir la convergence étroite de μ_n . Le cas suivant en donne une illustration. Soit $\mu_n = \delta_{\frac{1}{n}}$ et $\mu = \delta_0$. on a μ_n converge étroitement vers μ mais si $A =]0; 1[$, $\mu_n(A) = 1$ qui ne converge donc pas vers $\mu(A) = 0$

On a les résultats suivants dans le cas $E = \mathbb{R}^d$:

Théorème 1.1.4 (Un critère de convergence étroite)

Pour que $\mu_n \rightarrow \mu$ étroitement, il faut et il suffit que pour tout $t \in \mathbb{R}^d$, $\varphi_{\mu_n}(t)$ tende vers $\varphi_\mu(t)$ où $\varphi_v(t)$ désigne la transformée de Fourier de v au point t .

Pour tout $\sigma > 0$, posons

$$g_\sigma(u) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left(-\frac{\|u\|^2}{2\sigma^2}\right), \quad \forall u \in \mathbb{R}^d. \quad (1.3)$$

On peut vérifier immédiatement que

$$e^{-\frac{\|t\|^2}{2}} = \int_{\mathbb{R}^d} e^{i\langle t, u \rangle} g_1(u) du$$

d'où,

$$e^{-\frac{\|x-y\|^2}{2\sigma^2}} = \int_{\mathbb{R}^d} e^{i\langle x-y, \frac{u}{\sigma} \rangle} g_1(u) du = \sigma^d \int_{\mathbb{R}^d} g_1(\sigma u) e^{i\langle x-y, u \rangle} du \quad (1.4)$$

Alors la démonstration de Théorème 1.1.4 est basée sur les lemmes suivants :

Lemme 1.1.1

Soit μ une mesure bornée sur $\mathbb{R}^d$, alors

$$\int_{\mathbb{R}^d} g_\sigma(x-y) d\mu(x) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \varphi_\mu(u) g_1(\sigma u) e^{-i\langle u, y \rangle} du.$$

Preuve

Puisque $|g_1(\sigma u) e^{i\langle u, x-y \rangle}| \leq c e^{-\frac{\sigma^2}{2}\|u\|^2}$ et que cette dernière expression est $d_\mu(x).du$ intégrable, alors, par le Théorème de Fubini, on a

$$\int_{\mathbb{R}^d} g_\sigma(x-y) d\mu(x) = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} d\mu(x) \int_{\mathbb{R}^d} g_1(\sigma u) e^{i(x-y)'u} du \quad (1.5)$$

$$= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{ix'u} d\mu(x) \right) g_1(\sigma u) e^{-iy'u} du, \quad (1.6)$$

où x' désigne ici et dans la suite la transposée du vecteur x . □

Pour montrer qu'une partie de C_0 est dense, le Théorème de Stone-Weierstrass est un outil précieux. Rappelons qu'une sous-algèbre V de C_0 est un sous-espace vectoriel tel que $f, g \in V$ implique $fg \in V$. Alors :

Lemme 1.1.2 (Théorème de Stone-Weierstrass)

Soit A une sous-algèbre de C_0 vérifiant

1. pour tous $x, y \in \mathbb{R}^d$, $x \neq y$, il existe $f \in A$ telle que $f(x) \neq f(y)$,

2. pour tout $x \in \mathbb{R}^d$, il existe $f \in A$ telle que $f(x) \neq 0$,
alors $\bar{A} = C_0$.

Lemme 1.1.3

L'espace vectoriel engendré par l'application $x \mapsto g_\sigma(x - y)$ lorsque $\sigma > 0$ et $y \in \mathbb{R}^d$ est dense dans $C_0(\mathbb{R}^d)$

Preuve

Soit V l'espace vectoriel engendré par les fonctions $x \mapsto g_\sigma(x - y)$, $\sigma > 0$ et $y \in \mathbb{R}^d$. V est une sous algèbre de $C_0(\mathbb{R}^d)$. En effet, on montre que

$$g_\sigma(x - a)g_\rho(x - b) = Cg_\tau(x - c) \text{ avec}$$

$$\tau = \frac{\rho^2\sigma^2}{\rho^2 + \sigma^2}, \quad c = \frac{\rho^2a + \sigma^2b}{\rho^2 + \sigma^2}.$$

De plus cet espace vérifie les points 1. et 2. du Théorème 1.1.2 et donc $\bar{V} = C_0$. $\square$

Preuve du Théorème 1.1.4

La condition est évidemment nécessaire puisque la fonction $f_x(t) = e^{i\langle t, x \rangle} \in C_b$ et de la caractérisation ii) du Théorème 1.1.3.

Réciproquement, d'après le Lemme 1.1.1,

$$\int g_\sigma(x - y)d\mu_n(x) = (2\pi)^{-\frac{d}{2}} \int \varphi_{\mu_n}(u)g_1(\sigma u)e^{-i\langle u, y \rangle} du$$

et d'après le Théorème de Lebesgue, cette dernière inégalité tend vers

$$(2\pi)^{-\frac{d}{2}} \int \varphi_\mu(u)g_1(\sigma u)e^{-i\langle u, y \rangle} du = \int g_\sigma(x - y)d\mu(x).$$

On a donc $\int f d\mu_n \rightarrow \int f d\mu$ pour tout $f(x) = g_\sigma(x - y)$, $\sigma > 0$ et $y \in \mathbb{R}^d$. L'espace vectoriel engendré par ces fonction étant dense (voir Lemme 1.1.3), on conclut grâce à la Proposition 1.1.1. $\square$

On admet le Théorème de Paul Lévy suivant.

Théorème 1.1.5 (Paul Lévy)

Soit (μ_n) une suite de mesures positives finies sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Alors

1. Si $\mu_n \rightarrow \mu$ étroitement alors $\varphi_{\mu_n} \rightarrow \varphi_\mu$ simplement
2. Si $\varphi_{\mu_n} \rightarrow \varphi_\mu$ simplement et si φ est continue en 0 alors il existe $\mu \in \mathcal{M}(\mathbb{R}^d)$ tel que $\mu_n \rightarrow \mu$ étroitement et on a $\varphi = \varphi_\mu$.
En particulier $\mu_n \rightarrow \mu$ étroitement si et seulement si $\varphi_{\mu_n} \rightarrow \varphi_\mu$ simplement.

1.2 Convergence en loi

Définition 1.2.1 (Convergence en loi)

On dit que la suite de v.a. $(X_n)_{n \geq 0}$ converge en loi vers une v.a. X et on note $X_n \xrightarrow{\mathcal{L}} X$ si la loi de Probabilité μ_n de X_n converge étroitement vers la loi de probabilité μ de X . En d'autres termes

$$X_n \xrightarrow{\mathcal{L}} X \iff \forall f \in C_b(E), \quad \mathbb{E}(f(X_n)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(f(X))$$

ou encore si

$$X_n \xrightarrow{\mathcal{L}} X \iff \forall x \mid \mathbb{P}(X = x) = 0, \quad F_n(x) = \mathbb{P}(X_n < x) \xrightarrow{n \rightarrow +\infty} F(x) = \mathbb{P}(X < x)$$

.

Il est à noter que c'est par abus de langage qu'on dit que " X_n converge en loi vers X ". Ce n'est pas la suite de v.a. X_n qui converge mais c'est plutôt la suite de leurs loi.

1.3 Le Théorème Centrale Limite

Théorème 1.3.1 (TCL)

Soit $X_1, X_2, \dots, X_n$ une suite de v.a. indépendantes et identiquement distribuées (i.i.d.) à valeurs dans $\mathbb{R}^d$ telle que $\mathbb{E}||X_1||^2 < +\infty$. Alors en posant $m = \mathbb{E}(X_1)$, $D = \text{Var}(X)$ (vecteur espérance et matrice de dispersion) et $S_n = \sum_{i=1}^n X_i$,

La loi de $\frac{S_n - nm}{\sqrt{n}}$ converge étroitement vers la loi $\mathcal{N}_d(0, D)$.

En particulier, si $d = 1$, en posant $\sigma^2 = \text{Var}(X_1) = \mathbb{E}(X_1 - m)^2$, on a

$$\mathbb{P} \left[a \leq \frac{S_n - nm}{\sigma\sqrt{n}} \leq b \right] \xrightarrow{n \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Avant la preuve, rappelons la proposition suivante où φ_X désigne la fonction caractéristique de la v.a. $X \in \mathbb{R}^d$.

Proposition 1.3.1

Soit X une v.a. à valeurs dans $\mathbb{R}^d$.

i) Si $X \in L_d^1$, φ_X est dérivable et $\frac{\partial \varphi_X}{\partial t_k}(t) = \mathbb{E}(iX_k e^{i\langle t, X \rangle})$. En particulier $\frac{\partial \varphi_X}{\partial t_k}(0) = i\mathbb{E}(X_k)$.

ii) Si $X \in L_d^2$, φ_X est deux fois dérivable et $\frac{\partial^2 \varphi_X}{\partial t_j \partial t_k}(t) = -\mathbb{E}(X_j X_k e^{i\langle t, X \rangle})$. En particulier $\frac{\partial^2 \varphi_X}{\partial t_j \partial t_k}(0) = -\mathbb{E}(X_j X_k)$.

Preuve du Théorème 1.3.1

Soit $\varphi(t)$ la fonction caractéristique de $X_1 - m$. D'après la proposition 1.3.1 φ est de classe C^2 et on a $\frac{\partial \varphi}{\partial t_j}(0) = 0$, $j = 1, \dots, d$ puis $\left[\frac{\partial^2 \varphi}{\partial t_j \partial t_k}(0), j, k = 1, \dots, d \right] = -D$. D'où d'après la formule de Taylor-Young,

$$\varphi(t) = 1 - \frac{1}{2} t' D t + ||t||^2 \varepsilon(t) \text{ avec } \lim_{||t|| \rightarrow 0} \varepsilon(t) = 0.$$

Or la fonction caractéristique $\varphi_n(t)$ de $\frac{S_n - nm}{\sqrt{n}}$ est $\left(\varphi \left(\frac{t}{\sqrt{n}} \right) \right)^n$. En passant par la fonction log on a

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \lim_{n \rightarrow +\infty} \left[1 - \frac{1}{2n} \left[t' D t + 2||t||^2 \varepsilon \left(\frac{t}{\sqrt{n}} \right) \right] \right]^n = \exp \left[-\frac{1}{2} t' D t \right].$$

Ce qui achève la démonstration du Théorème car la fonction caractéristique de la loi $\mathcal{N}_d(0, D)$ est $\exp \left[-\frac{1}{2} t' D t \right]$. □

Remarque 1.3.1

L'hypothèse $\mathbb{E}\|X_1\|^2 < +\infty$ est essentielle. Sans elle il peut y avoir convergence vers d'autres lois que la loi normale.

Exemple 1.3.1

$X \rightsquigarrow \mathcal{B}(n, p)$ et on se propose d'évaluer $\mathbb{P}\left[\left|\frac{X}{n} - p\right| > \alpha\right]$. Comme on peut récrire $X = \sum_{i=1}^n X_i$ avec $X_1, \dots, X_n$ i.i.d. $\mathcal{B}(1, p)$, $\mathbb{E}(X_1) = p$ et $\text{Var}(X_1) = p(1-p)$. On a d'après le TCL $\frac{X - np}{\sqrt{np(1-p)}} \rightarrow \mathcal{N}(0, 1)$ pour n assez grand. D'où

$$\mathbb{P}\left[\left|\frac{X}{n} - p\right| > \alpha\right] = \mathbb{P}\left[\frac{|X - np|}{\sqrt{np(1-p)}} > \frac{\alpha\sqrt{n}}{\sqrt{p(1-p)}}\right] \simeq \frac{2}{\sqrt{2\pi}} \int_{\alpha_n}^{+\infty} e^{-\frac{t^2}{2}} dt$$

avec $\alpha_n = \frac{\alpha\sqrt{n}}{\sqrt{p(1-p)}}$ (ceci, en utilisant la symétrie par rapport à 0 de la loi $\mathcal{N}(0, 1)$).

Une table de la loi $\mathcal{N}(0, 1)$ montre que $\frac{1}{\sqrt{2\pi}} \int_u^{+\infty} e^{-\frac{t^2}{2}} dt \neq 0.025$ si $u = 1.96$. D'où pour $n = 400$ et $p = \frac{1}{2}$, $\mathbb{P}\left[\left|\frac{X}{400} - \frac{1}{2}\right| > 0.005\right] \simeq 0.005$.

Lorsque les v.a. X_n ont des densités régulières, on peut renforcer le Théorème 1.3.1.

Théorème 1.3.2 (Théorème de la limite locale)

Soit $X_1, X_2, \dots, X_n$ une suite de v.a. i.i.d. à valeurs dans $\mathbb{R}^d$ de fonction caractéristique φ_{X_1} telles que $\mathbb{E}\|X_1\|^2 < +\infty$ et $\varphi_{X_1} \in L^1$. Alors D est inversible, les v.a. $\frac{S_n - nm}{\sqrt{n}}$ ont des densités continues et ces densités convergent uniformément sur $\mathbb{R}^d$ vers la densité de la loi $\mathcal{N}_d(0, D)$.

La preuve du Théorème 1.3.2 est basée sur les lemmes suivants :

Lemme 1.3.1

Soit X une v.a. à valeurs dans $\mathbb{R}^d$ tel que $\mathbb{E}\|X\|^2 < +\infty$ et $\varphi = \varphi_X \in L^2$. Alors X a une densité continue, $D(X)$ est inversible et

- i) Il existe $\rho > 0$, $a > 0$ tel que $|\varphi(t)| \leq e^{-at^2}$ pour $|t| \leq \rho$,
- ii) Pour tout $\rho > 0$, $\sup_{|t| \geq \rho} |\varphi(t)| < 1$.

Preuve

X a une densité continue d'après le Théorème d'inversion. Si $D = D(X)$ n'était pas inversible, il existerait $t \neq 0$ tel que $t'Dt = 0$, d'où $\mathbb{E}(t'X)^2 = 0$ i.e. $t'X = 0$ p.s. et X ne peut avoir de densité. Remarquons que ceci entraîne l'existence de $k > 0$ et $K > 0$ tels que

$$k\|t\|^2 \leq t'Dt \leq K\|t\|^2 \text{ pour tout } t \text{ (e.g. } k = \inf S_p(D) \text{ et } K = \rho(D)). \quad (1.7)$$

- i) Posons $\psi(t) = e^{-a\|t\|^2} \varphi(t)$ où a sera précisé ultérieurement. Un calcul facile montre que

$$\psi(0) = 1, \quad \frac{\partial \psi}{\partial t_j}(0) = 0, \quad \left[\frac{\partial^2 \psi}{\partial t_j \partial t_k}(0) \right]_{j,k} = aI - D$$

et de (1.7) on en déduit l'existence de $a > 0$, $m > 0$ et $M > 0$ telles que

$$m\|t\|^2 \leq t'(aI - D)t \leq M\|t\|^2, \quad \forall t \in \mathbb{R}^d. \quad (1.8)$$

Alors $\psi(t) = 1 - \frac{1}{2}t'(aI - D)t + ||t||^2\lambda(t)$ avec $|\lambda(t)| \mapsto 0$ si $||t|| \mapsto 0$ et pour $M||t||^2 < 1$,

$$|\psi(t)| = \left| 1 - \frac{1}{2}t'(aI - D)t + ||t||^2\lambda(t) \right| \leq 1 - ||t||^2 \left(\frac{m}{2} - \lambda(t) \right)$$

et donc $|\psi(t)| < 1$ pour $||t|| \leq \rho$.

- ii) Remarquons d'abord que pour $t \neq 0$, $|\varphi(t)| < 1$; sinon il existerait $s \in \mathbb{R}$ tel que $\mathbb{E}(e^{it'X}) = e^{is}$ donc $\mathbb{E}(e^{i(t'X-s)} - 1) = 0$ et en prenant la partie réelle, $\mathbb{E}(\cos(t'X - s)) = 1$ c'est à dire $\cos(t'X - s) = 1$ p.s, soit $X \in \{x | t'x = s + 2k\pi\}$. Mais cet ensemble est de mesure nulle, ce qui contredit le fait que X ait une densité. On a donc pour tout $R > 0$, $\rho > 0$, $\sup\{|\varphi(t)|, \rho \leq ||t|| \leq R\} < 1$.

On conclut grâce au lemme suivant qui est admis.

Lemme 1.3.2

$|\varphi(t)| \longrightarrow 0$, quand $t \rightarrow +\infty$.

Preuve du Théorème 1.3.2

1. Ceci étant, soit $\varphi(t)$ la f.c. de $X_1 - \mathbb{E}(X_1)$. Comme $\varphi(t) = e^{-itm}\varphi_{X_1}(t) \in L^1$, on déduit du Lemme 1.3.1 que $X_1 - \mathbb{E}(X_1)$ a une densité continue $f(x)$, donc $S_n - nm$ a pour densité $f * f * \dots * f$ et donc $\frac{S_n - nm}{\sqrt{n}}$ a une densité continue et que D est inversible.
2. D'abord montrons que la f.c. de $\frac{S_n - nm}{\sqrt{n}}$, à savoir $\psi_n(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right) \right)^n$ tend vers $\exp\left(-\frac{1}{2}t'Dt\right)$ dans L^1 . On écrit pour cela

$$\psi_n(t) = \psi_n(t)\mathbf{1}_{\{||t|| \leq \rho\sqrt{n}\}} + \psi_n(t)\mathbf{1}_{\{||t|| > \rho\sqrt{n}\}},$$

le ρ étant fourni par le i) du Lemme 1.3.1. Alors en chaque t ,

$$\psi_n(t)\mathbf{1}_{\{||t|| \leq \rho\sqrt{n}\}} \longrightarrow \exp\left(-\frac{1}{2}t'Dt\right) \text{ (d'après le T.C.L.) et}$$

$$\left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right| \mathbf{1}_{\{||t|| \leq \rho\sqrt{n}\}} \leq \left\{ \exp\left(-a\left\|\frac{t}{\sqrt{n}}\right\|^2\right) \right\}^n = \exp(-a||t||^2) \in L^1 \text{ (Lemme 1.3.1).}$$

Donc

$$\psi_n(t)\mathbf{1}_{\{||t|| \leq \rho\sqrt{n}\}} \longrightarrow \exp\left(-\frac{1}{2}t'Dt\right) \text{ dans } L^1.$$

D'autre part, si $c = \sup\{|\varphi(t)|; ||t|| \geq \rho\} < 1$ (Lemme 1.3.1, ii)), on a

$$\int_{\{||t|| \leq \rho\sqrt{n}\}} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt \leq c^{n-1} \int_{\mathbb{R}^d} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt = c^{n-1} n^{\frac{d}{2}} \int_{\mathbb{R}^d} |\varphi(t)| dt \rightarrow 0, \quad n \rightarrow +\infty$$

donc $\psi_n(t)\mathbf{1}_{\{||t|| > \rho\sqrt{n}\}} \rightarrow 0, n \rightarrow +\infty$ dans L^1 et on a le résultat cherché.

3. Mais alors en notant par $f_n(x)$ la densité de $\frac{1}{\sqrt{n}}(S_n - nm)$ et par $g(x)$ celle de $\mathcal{N}_d(0, D)$, on a grâce au Théorème d'inversion

$$f_n(x) - g(x) = (2\pi)^{-d} \int e^{-it'x} \left[\psi_n(t) - \exp\left(-\frac{1}{2}t'Dt\right) \right] dt$$

d'où

$$||f_n - g||_\infty \leq (2\pi)^{-d} \int \left| \psi_n(t) - \exp\left(-\frac{1}{2}t'Dt\right) \right| dt \xrightarrow{n \rightarrow +\infty} 0. \quad \square$$

1.4 Exercices

Exercice 1.1

Montrer que l'ensemble des lois normales sur $\mathbb{R}$ est fermé pour la convergence étroite.

Exercice 1.2

Étudier la convergence vague de la suite de mesures :

- a) $d\mu_n(x) = \sin(nx)dx$,
- b) $d\mu_n(x) = \sin^2(nx)dx$,
- c) $\mu_n = \delta_{a_n}$ où (a_n) est une suite dans $\mathbb{R}$.

Exercice 1.3

Soit (X_n) une suite de v.a. a.c. telle que pour tout $n > 1$, $X_n \rightsquigarrow \mathcal{E}\left(\frac{1}{n}\right)$. On pose $X_n^0 = X_n - [X_n]$. Montrer que la suite (X_n^0) converge en loi vers une v.a. a.c. que l'on précisera.

Exercice 1.4

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue bornée. Montrer qu'on a

$$\lim_{n \rightarrow +\infty} e^{-n} \sum_{k=0}^{+\infty} f\left(\frac{k-n}{\sqrt{n}}\right) \frac{n^k}{k!} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{-\frac{t^2}{2}} dt.$$

Exercice 1.5

Soit $(X_n)_{n \geq 1}$ une suite variables aléatoires indépendantes de même loi de Poisson de paramètre

1. Soit $S_n = \sum_{k=1}^n X_k$. Rappeler la loi de S_n et calculer la limite de la suite

$$\left(e^{-n} \sum_{k=0}^n \frac{n^k}{k!} \right)_{n \geq 1}.$$

Exercice 1.6

Soit $t \in \mathbb{R}$ et $f \in C_b(\mathbb{R})$. Étudier la limite lorsque $n \rightarrow +\infty$ de la suite

$$Y_n = e^{-nt} \sum_{k \in \mathbb{N}} \frac{(nt)^k}{k!} f\left(\frac{k}{n}\right).$$

Exercice 1.7

Soient $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ et f une application continue de $\mathbb{R}$ dans $\mathbb{R}$. On suppose que $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une variable aléatoire X . Montrer que la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$.

Exercice 1.8

Soit (X_n) une suite de v.a. suivant la loi $\mathcal{P}(\lambda_n)$. On suppose que $\lambda_n \rightarrow +\infty$. Étudier la convergence en loi de (X_n) et de $\left(\frac{X_n}{\lambda_n}\right)$.

Exercice 1.9

Soit (X_n) une suite de v.a. absolument continue de densité f_n et X une de v.a. absolument continue. Montrer que si $f_n \rightarrow f$ p.p. alors $X_n \xrightarrow{\mathcal{L}} X$.

Étudier la réciproque. (On pourra considérer par exemple la suite $f_n(x) = (1 + \sin(2\pi nx)) \mathbf{1}_{[0;1]}(x)$).

Exercice 1.10

Soit (X_n) une suite de v.a.r. indépendantes et de même loi

$$P[X_n = 1] = P[X_n = -1] = \frac{1}{2}.$$

Étudier la convergence en loi de la suite (Y_n) définie par $Y_n = \sum_{k=1}^n \frac{X_k}{2^k}$.

Chapitre 2

Suites et séries de variables aléatoires réelles - Loi des grands nombres

On considère une suite $X_1, \dots, X_n$, de variables aléatoires (v.a.) à valeurs réelles définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et on s'intéresse au comportement asymptotique de cette suite aléatoire. L'étude repose en partie sur un résultat élémentaire très utile.

2.1 Lemme de Borel-Cantelli

On considère une suite d'événements $\{\Lambda_n, n \in \mathbb{N}\}$ dans $\mathcal{A}$. La suite des événements $\left(\bigcup_{k \geq n} \Lambda_k, n \in \mathbb{N}\right)$ est décroissante, son intersection

$$\Lambda = \limsup_n \Lambda_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} \Lambda_k$$

est encore un événement dans $\mathcal{A}$. On remarquera que Λ représente l'ensemble des aléas ω qui appartiennent à une infinité d'événements Λ_n , autrement dit $\mathbf{1}_\Lambda = \limsup_{n \rightarrow +\infty} \mathbf{1}_{\Lambda_n}$; ce qui justifie la notation.

Lemme 2.1.1 (Borel-Cantelli)

Soit $(\Lambda_n)_{n \in \mathbb{N}^*}$ une suite d'événements. Deux cas sont possibles en ce qui concerne l'événement Λ :

- a) Si $\sum_{n \geq 1} \mathbb{P}(\Lambda_n) < +\infty$ alors $\mathbb{P}[\Lambda] = 0$. En d'autres termes, avec probabilité 1, au plus un nombre fini des Λ_n se produisent si la série précédente converge.
- b) Si les (Λ_n) sont indépendants et $\sum_{n \geq 1} \mathbb{P}(\Lambda_n) = +\infty$ alors $\mathbb{P}[\Lambda] = 1$. En d'autres termes, avec probabilité 1, une infinité des Λ_n se produisent si $\sum_{n \geq 1} \mathbb{P}(\Lambda_n) = +\infty$.

Preuve

- i) Fixons $\varepsilon > 0$. Il existe un entier N tel que $\sum_{n \geq N} \mathbb{P}(\Lambda_n) < \varepsilon$ et en conséquence

$$\mathbb{P}\left(\bigcup_{n \geq N} \Lambda_n\right) < \varepsilon \text{ car } \mathbb{P}\left(\bigcup_{n \geq N} \Lambda_n\right) \leq \sum_{n \geq N} \mathbb{P}(\Lambda_n). \text{ De la définition de } \Lambda, \text{ on tire a fortiori } \mathbb{P}(\Lambda) < \varepsilon.$$

ii) Comme les Λ_n sont indépendants, il en est de même pour leurs complémentaires, et on a

$$\mathbb{P}\left(\bigcup_{j=k}^m \Lambda_j\right) = 1 - \mathbb{P}\left(\bigcap_{j=k}^m \Lambda_j^c\right) = 1 - \prod_{j=k}^m \mathbb{P}(\Lambda_j^c) \geq 1 - \exp\left\{-\sum_{j=k}^m \mathbb{P}(\Lambda_j)\right\},$$

car $\log(1 - \mathbb{P}(\Lambda_j)) \leq -\mathbb{P}(\Lambda_j)$, $\forall j$.

On fait tendre m vers $+\infty$, et on tire

$$\mathbb{P}\left(\bigcup_{j=k}^{+\infty} \Lambda_j\right) \geq 1 - \exp\left\{-\sum_{j=k}^{+\infty} \mathbb{P}(\Lambda_j)\right\} = 1.$$

Comme les événements $\bigcup_{j=k}^{+\infty} \Lambda_j$ décroissent vers Λ quand k tend vers $+\infty$ on a bien $\mathbb{P}(\Lambda) \geq 1$ et donc $\mathbb{P}(\Lambda) = 1$. $\square$

2.2 Divers modes de convergences

Définition 2.2.1 (Convergence p.s.)

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers une v.a. X et on note $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$, s'il existe un événement Λ avec $\mathbb{P}(\Lambda) = 1$, tel que

$$\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega), \text{ pour tout } \omega \in \Lambda.$$

En d'autres termes, si $\forall \varepsilon > 0$ $\mathbb{P}\left[\limsup_{n \rightarrow +\infty} |X_n - X| > \varepsilon\right] = 0$.

Définition 2.2.2 (Convergence en probabilité)

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers une v.a. X et on note $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$, si $\forall \varepsilon > 0$

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

Proposition 2.2.1

Si $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$, alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$.

Preuve

Soit Λ l'événement de probabilité 1 qui apparaît dans la définition. On fixe $\varepsilon > 0$, et pour chaque entier n , on considère l'événement

$$\Gamma_n = \{\omega \in \Lambda : \sup_{k \geq n} |X_k(\omega) - X(\omega)| > \varepsilon\}.$$

La suite des Γ_n est décroissante et $\bigcap \Gamma_n = \emptyset$ (puisque X_n converge vers X sur Λ). On a donc $\lim_{n \rightarrow +\infty} \mathbb{P}(\Gamma_n) = 0$. Comme

$$\{|X_n - X| > \varepsilon\} \subset \Gamma_n \cup \Lambda^c$$

,

on a a fortiori $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. $\square$

La réciproque est fautive; voici un contre exemple : On prend $\Omega = [1, 2]$ qu'on munit de la mesure de Lebesgue. Pour chaque n , on note k l'unique entier tel que $2^k \leq n < 2^{k+1}$, et on

prend $X_n(t) = \mathbb{1}_{[n2^{-k}, (n+1)2^{-k}]}$. Il est clair que pour tout $\varepsilon \in]0, 1[$, $\mathbb{P}(|X_n| > \varepsilon) = 2^{-k}$, de sorte que $X_n \rightarrow 0$ en probabilité. Néanmoins, pour tout $t \in [0, 1]$, il existe une infinité d'entiers n pour lesquels $X_n(t) = 1$, et $X_n(t)$ ne converge pas vers 0.

En revanche on a une réciproque partielle.

Proposition 2.2.2

Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$, alors il existe une suite extraite $X_{N(n)}$ qui converge vers X p.s.

Preuve

Comme $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$ pour tout $\varepsilon > 0$, on peut trouver pour tout $n \geq 1$ un entier $N(n)$ tel que $\mathbb{P}\left(|X_{N(n)} - X| > \frac{1}{n}\right) \leq 2^{-n}$

La série $\sum \mathbb{P}\left(|X_{N(n)} - X| > \frac{1}{n}\right)$ converge donc. D'après le lemme de Borel-Cantelli, la probabilité de l'événement

$$\bigcap_{n \geq 1} \bigcup_{k \geq n} \left\{ |X_{N(k)} - X| > \frac{1}{k} \right\}$$

est nulle. Si Λ désigne l'événement complémentaire, on a donc $\mathbb{P}(\Lambda) = 1$, et (par définition de Λ) pour tout $\omega \in \Lambda$, il existe un entier $k(\omega)$ tel que

$$|X_{N(n)}(\omega) - X(\omega)| < \frac{1}{n}, \text{ pour tout } n > k(\omega).$$

Ceci montre que

$$X_{N(n)}(\omega) \rightarrow X(\omega). \quad \square$$

Définition 2.2.3 (convergence dans L^p)

Pour tout $p \geq 1$, on dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de v.a. converge dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$ vers une v.a. $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ et on note $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$ si $\mathbb{E}(|X_n|^p) < +\infty$ pour tout n et si

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^p) = 0$$

Comparons cette notion avec les précédentes.

Proposition 2.2.3

Si $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$, alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$.

Réciproquement, si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et s'il existe une v.a. réelle $Y \in L^p$ telle que $|X_n| \leq Y$ pour tout n , alors $X_n \xrightarrow[n \rightarrow +\infty]{L^p} X$.

Preuve

Fixons $\varepsilon > 0$ et $\eta > 0$. On sait qu'on peut trouver un entier n_0 tel que $\mathbb{E}(|X_n - X|^p) \leq \eta \varepsilon^p$ dès que $n \geq n_0$. En appliquant l'inégalité de Markov, on obtient dans ce cas

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

Réciproquement, si la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité, c'est encore le cas pour toute suite extraite $(X_{N(n)})_{n \in \mathbb{N}}$. On sait qu'on peut extraire de cette dernière une sous-suite $(X_{N(n)})_{n \in \mathbb{N}}$ qui converge p.s. vers X . Par hypothèse on peut appliquer le Théorème de convergence dominée, c'est-à-dire $X_{N(n)} \rightarrow X$ dans $L^p(\Omega, \mathbb{P})$. Ainsi, de toute suite extraite de $(X_n)_{n \in \mathbb{N}}$, on a su extraire une sous-sous-suite qui converge vers X dans L^p . Donc $X_n \rightarrow X$ dans L^p . $\square$

Un des résultats utile en probabilité et statistique est celui qui dit, dans un cas particulier, que si (X_n) et (Y_n) sont deux suites de v.a. telles que $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y$ et si f est une fonction continue sur $\mathcal{X} \times \mathcal{X}$ alors $f(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} f(X, Y)$. Pour établir ce résultat on a :

Proposition 2.2.4

Une condition nécessaire et suffisante pour que $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ est que toute sous-suite de (X_n) contienne une sous-suite qui converge presque sûrement vers X .

Preuve

Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ alors d'après la définition même de la convergence en probabilité, toute sous-suite (X_{n_k}) de (X_n) converge en probabilité vers X . La condition devient alors nécessaire d'après la Proposition 2.2.2.

Pour montrer qu'elle est suffisante, supposons que (X_n) ne tende pas vers X en probabilité. Alors il existe $\varepsilon_0 > 0$ et $\delta_0 > 0$ tel que $\limsup \mathbb{P}[|X_n - X| > \varepsilon_0] = \delta_0 > 0$ et aussi il existe une sous-suite (X_{n_k}) de (X_n) tel que $\mathbb{P}[|X_{n_k} - X| > \varepsilon_0] \xrightarrow[k \rightarrow +\infty]{} \delta_0 > 0$.

Par conséquent, ni (X_{n_k}) ni aucune de ses sous-suites ne converge en probabilité vers X . Donc aucune sous-suite de (X_{n_k}) ne peut converger p.s. vers X .

Proposition 2.2.5

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^q$ une fonction continue sur un borélien B de $\mathbb{R}^2$ tel que $\mathbb{P}[X \in B] = 1$. Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ alors $f(X_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} f(X)$.

Preuve

Soit $(f(X_{n_j}))$ une sous-suite de la suite $(f(X_n))$. D'après la Proposition 2.2.4, on a seulement besoin de montrer qu'il existe une sous-suite $(f(X_{k_j}))$ de la sous-suite $(f(X_{n_j}))$ qui converge p.s.

Comme $X_{n_j} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$, il existe une sous-suite (X_{k_j}) de la sous-suite (X_{n_j}) tel que $X_{k_j} \xrightarrow[k \rightarrow +\infty]{p.s.} X$.

Soit alors $A = \left\{ X_{k_j} \xrightarrow[k \rightarrow +\infty]{p.s.} X \right\} \cap \{X \in B\}$. Il est clair que $\mathbb{P}(A) = 1$ et $\omega \in A \implies X_{k_j}(\omega) \xrightarrow[k \rightarrow +\infty]{p.s.} X(\omega) \in B$. f étant continue sur B , pour tout $\omega \in A$, $f(X_{k_j}(\omega)) \xrightarrow[k \rightarrow +\infty]{} f(X(\omega))$.

En d'autres termes $f(X_{k_j}) \xrightarrow[k \rightarrow +\infty]{p.s.} f(X)$ et donc $f(X_n) \xrightarrow[n \rightarrow +\infty]{p.s.} f(X)$.

Corollaire 2.2.1

Soit (X_n) et (Y_n) deux suites de v.a. telles que $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y$. Si f est une fonction mesurable sur $\mathbb{R}^2$ tel que (X, Y) appartienne p.s. à l'ensemble de continuité de f alors $f(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} f(X, Y)$.

La démonstration du corollaire est basée sur le Lemme suivant :

Lemme 2.2.1

Soit X et Y deux v.a. Alors $\forall \varepsilon > 0$, $\mathbb{P}[|X + Y| \geq \varepsilon] \leq \mathbb{P}\left[|X| \geq \frac{\varepsilon}{2}\right] + \mathbb{P}\left[|Y| \geq \frac{\varepsilon}{2}\right]$.

Preuve

Vérifions d'abord que $\{|X + Y| \geq \varepsilon\} \subset \left\{|X| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|Y| \geq \frac{\varepsilon}{2}\right\}$. Si $|X(w) + Y(w)| \geq \varepsilon$ alors d'après l'inégalité triangulaire $|X(w)| + |Y(w)| \geq \varepsilon$. Or si la somme de deux nombres est supérieure ou égale à ε alors l'un au moins des deux nombres est supérieur ou égale à $\frac{\varepsilon}{2}$. Donc $|X|(w) \geq \frac{\varepsilon}{2}$ ou $|Y|(w) \geq \frac{\varepsilon}{2}$. d'où le résultat. Le lemme en résulte en prenant les probabilités

dans chaque membre de l'inclusion.

Preuve du Corollaire 2.2.1

Montrons d'abord que si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y$ alors $(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} (X, Y)$.

Notons par 0 la v.a. égale p.s. à 0. On a alors $|(X_n, Y_n) - (X, Y)| \leq |(X_n, 0) - (X, 0)| + |(0, Y_n) - (0, Y)|$. Comme $|(X_n, 0) - (X, 0)| = |X_n - X|$ et $|(0, Y_n) - (0, Y)| = |Y_n - Y|$, alors $(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} (X, Y)$ d'après le Lemme 2.2.1, et le Corollaire 2.2.1 résulte immédiatement de la Proposition 2.2.5.

Corollaire 2.2.2

Si $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} Y$, alors

1. $aX_n + bY_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} aX + bY, \forall a, b \in \mathbb{R}.$
2. $\frac{1}{X_n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{1}{X}$ si $\mathbb{P}[X_n \neq 0] = \mathbb{P}[X \neq 0] = 1$
3. $X_n Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} XY$

2.3 Convergence p.s. de séries de variables aléatoires indépendantes-Théorèmes limites forts

On se donne une suite v.a. $(X_n)_{n \in \mathbb{N}^*}$ que l'on suppose indépendante c'est-à-dire que pour tout n , $X_1, \dots, X_n$ sont des v.a. indépendantes. On note $S_n = \sum_{i=1}^n X_i$ la somme partielle jusqu'à l'ordre n , et on s'intéresse à la convergence de la suite des v.a. $(S_n)_{n \in \mathbb{N}^*}$.

On dira que la série $\sum X_n$ converge en loi s'il existe une distribution de probabilité F tel que $F_{S_n} \xrightarrow[n \rightarrow +\infty]{e} F$.

De même $\sum X_n$ converge en probabilité ou en moyenne d'ordre r s'il existe une v.a. X tel que $S_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ ou $S_n \xrightarrow[n \rightarrow +\infty]{L^r} X$.

L'outil fondamental pour les théorèmes limites forts est

Théorème 2.3.1 (Inégalité de Kolmogorov)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite v.a. indépendantes de moment d'ordre deux fini. Alors

$$\mathbb{P} \left[\max_{1 \leq k \leq n} |S_k - \mathbb{E}(S_k)| \geq \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{j=1}^n \text{Var}(X_j).$$

Preuve

Sans perte de généralité, on peut supposer $\mathbb{E}(X_k) = 0$ pour tout k . Posons $A_0 = \Omega$ et $\forall k \geq 1$,

$$A_k = \left\{ \max_{1 \leq j \leq k} |S_j| < \varepsilon \right\} = \bigcap_{j=1}^k \{|S_j| < \varepsilon\} \quad (2.1)$$

$$B_k = A_{k-1} \cap A_k^c = \left(\bigcap_{j=1}^k \{|S_j| < \varepsilon\} \right) \cap \{|S_k| \geq \varepsilon\} \quad (2.2)$$

B_k est l'évènement " S_k est la 1^{ère} somme partielle dont la valeur absolue est supérieure ou égale à ε " et il est facile de vérifier que

$$A_n^c = \left\{ \max_{1 \leq j \leq n} |S_j| \geq \varepsilon \right\} = \bigcup_{k=1}^n B_k.$$

En outre $B_1, B_2, \dots, B_n$ sont des évènements disjoints et pour tout $1 \leq k \leq n$, on a

$$\begin{aligned} \int_{B_k} S_n^2 d\mathbb{P} &= \mathbb{E} (S_n \mathbf{1}_{B_k})^2 = \mathbb{E} ((S_n - S_k) \mathbf{1}_{B_k} + S_k \mathbf{1}_{B_k})^2 \\ &= \mathbb{E} ((S_n - S_k) \mathbf{1}_{B_k})^2 + 2\mathbb{E} ((S_n - S_k) S_k \mathbf{1}_{B_k}) + \mathbb{E} (S_k \mathbf{1}_{B_k})^2. \end{aligned}$$

Comme les v.a. $S_n - S_k$ et $S_k \mathbf{1}_{B_k}$ sont indépendantes et $\mathbb{E}(X_j) = 0$ pour $1 \leq j \leq n$, on a

$$\mathbb{E} ((S_n - S_k) S_k \mathbf{1}_{B_k}) = \mathbb{E}((S_n - S_k)) \mathbb{E} (S_k \mathbf{1}_{B_k}) = 0$$

d'où

$$\int_{B_k} S_n^2 d\mathbb{P} \geq \mathbb{E} (S_k \mathbf{1}_{B_k})^2 \geq \varepsilon^2 \mathbb{P}(B_k).$$

Il en résulte que

$$\sum_{k=1}^n \text{Var}(X_k) = \text{Var}(S_n) \geq \int_{A_n^c} S_n^2 d\mathbb{P} = \sum_{k=1}^n \int_{B_k} S_n^2 d\mathbb{P} = \varepsilon^2 \mathbb{P} \left[\bigcup_{k=1}^n B_k \right]. \quad \square$$

Théorème 2.3.2

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite v.a. indépendantes de moment d'ordre deux fini. Si $\sum_{k=1}^{+\infty} \text{Var}(X_k) < +\infty$

alors $\sum_{n=1}^{+\infty} (X_n - \mathbb{E}(X_n))$ converge p.s.

Preuve

Posons $S_n = \sum_{n=1}^n (X_n - \mathbb{E}(X_n))$. Alors $\limsup_n S_n$ et $\liminf_n S_n$ sont finies p.s. En effet, pour tout $m \geq 1$, $\limsup_{n \rightarrow +\infty} |S_n| \leq \limsup_{n \rightarrow +\infty} |S_n - S_m| + |S_m| \leq \sup_{n \geq m} |S_n - S_m| + |S_m|$ et d'après le Théorème 2.3.1, on a

$$\begin{aligned} \mathbb{P} \left[\sup_{n \geq m} |S_n - S_m| \geq \varepsilon \right] &\leq \mathbb{P} \left[\bigcup_{n=m}^{+\infty} \{|S_n - S_m| \geq \varepsilon\} \right] \\ &= \lim_{M \rightarrow \infty} \mathbb{P} \left[\bigcup_{n=m}^M \{|S_n - S_m| \geq \varepsilon\} \right] = \lim_{M \rightarrow \infty} \mathbb{P} \left[\max_{m \leq n \leq M} \{|S_n - S_m| \geq \varepsilon\} \right] \\ &\leq \frac{1}{\varepsilon^2} \sum_{n=m}^{+\infty} \text{Var}(X_n) < +\infty \text{ d'après l'hypothèse.} \end{aligned}$$

Donc en faisant tendre ε vers $+\infty$, la 1^{ère} probabilité tend vers 0 et on a montré que $\limsup_n S_n$ et de même $\liminf_n S_n$ sont finies p.s.

Pour montrer que S_n converge p.s. nous avons seulement besoin de montrer que $\forall \varepsilon > 0$, $\mathbb{P}[\limsup_n S_n - \liminf_n S_n \geq \varepsilon] = 0$.

Pour le faire, observons d'abord que si (a_n) est une suite de nombres réels alors $\sup |a_n| = \max\{|\sup a_n|, |\inf a_n|\}$ et $|\sup a_n| > \varepsilon$ ou $|\inf a_n| > \varepsilon$ entraîne $\sup |a_n| > \varepsilon$. Donc $\forall m \in \mathbb{N}^*$, $\forall \varepsilon > 0$

$$\begin{aligned} \{\limsup_n S_n - \liminf_n S_n \geq 2\varepsilon\} &\subset \left\{ \sup_{n \geq m} S_n - \inf_{n \geq m} S_n \geq 2\varepsilon \right\} \\ &\subset \left\{ \left| \sup_{n \geq m} S_n - S_m \right| \geq \varepsilon \right\} \cup \left\{ \left| S_m - \inf_{n \geq m} S_n \right| \geq \varepsilon \right\} \\ &\subset \left\{ \sup_{n \geq m} |S_n - S_m| \geq \varepsilon \right\}. \end{aligned}$$

Alors d'après l'inégalité de Kolmogorov

$$\mathbb{P}[\limsup_n S_n - \liminf_n S_n \geq \varepsilon] \geq \frac{1}{\varepsilon^2} \sum_{j=m}^{+\infty} \text{Var}(X_j) \xrightarrow{m \rightarrow +\infty} 0. \quad \square$$

Corollaire 2.3.1

Si (X_n) est une suite de v.a. indépendantes telles que $\sum_n \text{Var}(X_n) < +\infty$ et $\sum_n \mathbb{E}(X_n) < +\infty$ alors $\sum_n X_n$ converge p.s.

Preuve

C'est une conséquence immédiate du Théorème 2.3.2 et du fait que la somme de deux séries convergentes est convergente.

2.4 Loi des grand nombres (LGN)

La loi des grands nombres est la formulation rigoureuse des faits intuitifs suivants : si on lance un « grand » nombre de fois une pièce en l'air, il y aura en moyenne 50% de piles. De même, si on lance un « grand » nombre de fois un dé à 6 faces en l'air, il y aura en moyenne 1/6-ème des faces qui seront, par exemple, des 4 (si la pièce et le dé sont équilibrés).

Il existe une première version de la LGN pour la convergence en probabilité.

2.4.1 Loi faible des grands nombres (LfGN)

Théorème 2.4.1 (LfGN)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. deux à deux indépendantes, de même loi ayant un moment d'ordre 2. Alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}[X_1].$$

Preuve

Ici, la variable aléatoire limite est la constante $\mathbb{E}[X_1]$ ($= \mathbb{E}[X_i]$ pour tout i car les v.a X_i sont de même loi, donc de même espérance)

Il nous revient, pour la preuve, de vérifier que

$$\forall \varepsilon \lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1] \right| \geq \varepsilon \right) = 0$$

Posons $M_n = \frac{1}{n} \sum_{i=1}^n X_i$, par linéarité, on a

$$\mathbb{E}[M_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_1].$$

D'autre part, par indépendance des X_i on a

$$\text{Var}(M_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\text{Var}(X_1)}{n}.$$

En appliquant l'inégalité de Tchebychev à M_n , on a, pour tout $\varepsilon > 0$ et pour tout $n \in \mathbb{N}^*$

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1]\right| \geq \varepsilon\right) = \mathbb{P}(|M_n - \mathbb{E}[M_n]| \geq \varepsilon) \leq \frac{\text{Var}(M_n)}{\varepsilon^2} = \frac{\text{Var}(X_1)}{n\varepsilon^2}.$$

On conclut donc en faisant tendre $n \rightarrow +\infty$. $\square$

Souvent, on se trouve dans le cas particulier où les v.a. considérées sont de loi de Bernoulli, la LfGN se réécrit alors :

Corollaire 2.4.1

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. indépendante de Bernoulli de même paramètre p . Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} p.$$

Preuve

$\mathbb{E}(X_i) = p$ et comme $\text{Var}(X_i) = p(1-p) < +\infty$, la LfGN s'applique. $\square$

Avec la LfGN, on a donc la convergence de la moyenne arithmétique (dite aussi moyenne empirique) vers la moyenne probabiliste (l'espérance probabiliste). La LGN faible est encore vraie si on ne suppose que l'existence du moment d'ordre 1 : $\mathbb{E}[|X_1|] < +\infty$. Cependant, l'hypothèse $\mathbb{E}[|X_1|^2] < +\infty$ permet une preuve facile. Le cas avec seulement l'existence du moment d'ordre 1 sera obtenu comme conséquence de la LGN forte.

2.4.2 Loi forte des grands nombres (LFGN)

La loi des grands nombres admet une version pour la convergence presque sûre : c'est la loi forte des grands nombres. Avant d'énoncer le théorème le plus important, commençons par la série de lemmes suivants.

Lemme 2.4.1 (Césaro)

Soit (a_n) est une suite de nombres réels tendant vers a quand $n \rightarrow +\infty$, alors

$$\frac{1}{n} \sum_{i=1}^n a_i \xrightarrow[n \rightarrow +\infty]{} a.$$

Preuve Facile (confère cours d'analyse L_2). $\square$

Lemme 2.4.2 (Kronecker)

Soit $\sum_{n \geq 1} b_n$ une série convergente de nombres réels alors

$$\frac{1}{n} \sum_{i=1}^{+\infty} i b_i \xrightarrow[n \rightarrow +\infty]{} 0.$$

Preuve

Posons $b = \sum_{n \geq 1} b_n$ et $a_i = \sum_{j=1}^i b_j$. Alors, d'après le Lemme 2.4.1,

$$\frac{1}{n} \sum_{i=1}^n a_i = \frac{b_1 + (b_1 + b_2) + \cdots + (b_1 + b_2 + \cdots + b_n)}{n} \xrightarrow{n \rightarrow +\infty} b.$$

Or

$$\frac{nb_1 + (n-1)b_2 + \cdots + b_n}{n} = \frac{n+1}{n} \sum_{i=1}^n b_i - \frac{1}{n} \sum_{i=1}^n i b_i \xrightarrow{n \rightarrow +\infty} b. \quad \square$$

Lemme 2.4.3

Si (X_n) est une suite de v.a. indépendante telle que $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s.} X$ alors X est une constante p.s.

Preuve

Pour chaque i fixé, $\frac{X_i}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. Nous déduisons donc de l'hypothèse que

$$\frac{X_i + X_{i+1} + \cdots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} X, \quad \forall i$$

et donc X est mesurable par rapport à la tribu asymptotique $\mathcal{A}^\infty$ engendrée par les X_n , $\mathcal{A}^\infty = \bigcap_{n=1}^{+\infty} \sigma(X_n, X_{n+1}, \dots)$, d'où le résultat.

Lemme 2.4.4

Si X est une v.a. alors $\mathbb{E}(X)$ existe si, et seulement si $\sum_{n \geq 0} \mathbb{P}[|X| \geq n] < +\infty$.

Preuve

La conclusion du lemme découle des deux inégalités suivantes :

$$\int |X| d\mathbb{P} = \int \left(\sum_{n=1}^{+\infty} |X| \mathbf{1}_{\{n-1 \leq |X| < n\}} \right) d\mathbb{P} \leq \sum_{n=1}^{+\infty} n \mathbb{P}[n-1 \leq |X| < n] = \sum_{n=0}^{+\infty} \mathbb{P}[|X| \geq n]$$

et

$$\int |X| d\mathbb{P} \geq \sum_{n=0}^{+\infty} n \mathbb{P}[|X| \geq n] - 1.$$

Théorème 2.4.2 (Kolmogorov)

Si (X_n) est une suite de v.a. indépendantes tel que $\mathbb{E}(X_n^2) < +\infty$ pour tout n et

$$\sum_{k=1}^{+\infty} \frac{\text{Var}(X_k)}{k^2} < +\infty, \text{ alors } \frac{1}{n} \sum_{k=1}^{+\infty} \frac{X_k - \mathbb{E}(X_k)}{k} \xrightarrow[n \rightarrow +\infty]{p.s.} 0.$$

Preuve

Comme $\sum_{k=1}^{+\infty} \frac{\text{Var}(X_k)}{k^2} < +\infty$ alors d'après le Théorème 2.3.2 $\sum_{k=1}^{+\infty} \frac{X_k - \mathbb{E}(X_k)}{k}$ converge presque sûrement et Lemme 2.4.2 de Kronecker, on peut conclure que $\frac{1}{n} \sum_{k=1}^{+\infty} \frac{X_k - \mathbb{E}(X_k)}{k} \xrightarrow[n \rightarrow +\infty]{p.s.} 0. \quad \square$

Théorème 2.4.3 (LFGN de Kolmogorov)

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. indépendantes et de même loi. Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X_1].$$

si et seulement si, $\mathbb{E}(X_1) < +\infty$.

Preuve

Supposons que $\frac{1}{n} \sum_{i=1}^n X_i$ converge p.s. Posons $S_n = \sum_{i=1}^n X_i$. Alors $\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_{n-1}}{n-1} \frac{n-1}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$, soit $\frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. Donc $\mathbb{P} \left[\frac{|X_n|}{n} \geq 1 \text{ i.o.} \right] = 0$ ou encore $\mathbb{P}[|X_n| \geq n \text{ i.o.}] = 0$. Alors d'après le lemme de Borel-Cantelli, comme les événements $\{|X_n| \geq n\}$ sont indépendants, $\sum_{n=1}^{+\infty} \mathbb{P}[|X_n| \geq n] < +\infty$.

Mais $\mathbb{P}[|X_n| \geq n] = \mathbb{P}[|X_1| \geq n]$. Donc $\sum_{n=1}^{+\infty} \mathbb{P}[|X_1| \geq n] < +\infty$ et $\mathbb{E}(X_1)$ existe d'après le Lemme 2.4.4

Réciproquement supposons que $\mathbb{E}(X_1)$ existe. Alors si on note $Y_n = X_n \mathbb{1}_{\{|X_n| < n\}}$, on a :

$\mathbb{P}[Y_n \neq X_n] = \mathbb{P}[|X_n| \geq n] = \mathbb{P}[|X_1| \geq n]$ d'où d'après le Lemme 2.4.4, $\sum_{n=1}^{+\infty} \mathbb{P}[Y_n \neq X_n] < +\infty$ ce qui implique $\mathbb{P}[Y_n \neq X_n \text{ i.o.}] = 0$ d'après le lemme de Borel-Cantelli. Il suffira donc de montrer que $\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}(X_1)$ p.s.

Comme $\mathbb{E}(Y_n) = \mathbb{E}(X_n \mathbb{1}_{\{|X_n| < n\}}) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1)$ (convergence dominée)

alors $\frac{1}{n} \sum_{i=1}^n \mathbb{E}(Y_i) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}(X_1)$ d'après le Lemme 2.4.1.

Il suffit alors de montrer que $\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)) \xrightarrow[n \rightarrow +\infty]{p.s.} 0$. D'après le Théorème 2.4.2, il suffirait

de montrer que $\sum_{i=1}^{\infty} \frac{\text{Var}(Y_i)}{i^2} < \infty$.

Comme $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2 \leq \mathbb{E}(Y_i^2)$, il suffirait de montrer tout simplement que $\sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} < \infty$. Or $\mathbb{E}(Y_i^2) = \mathbb{E}(X_1^2 \mathbb{1}_{\{|X_1| < i\}})$ donc on va montrer que $\sum_{i=1}^{\infty} \frac{\mathbb{E}(X_1^2 \mathbb{1}_{\{|X_1| < i\}})}{i^2} < \infty$.

D'après le théorème de Beppo-Levi on pourra montrer seulement que $\int \left(\sum_{i=1}^{\infty} \frac{X_1^2}{i^2} \mathbb{1}_{\{|X_1| < i\}} \right) d\mathbb{P} < +\infty$. Mais alors pour $m = 1, 2, \dots$, on a

$$\begin{aligned} \int \left(\sum_{i=1}^{\infty} \frac{X_1^2}{i^2} \mathbb{1}_{\{|X_1| < i\}} \right) \mathbb{1}_{\{m-1 \leq |X_1| < m\}} d\mathbb{P} &= \int \left(\sum_{i=m}^{\infty} \frac{X_1^2}{i^2} \right) \mathbb{1}_{\{m-1 \leq |X_1| < m\}} d\mathbb{P} \\ &\leq \int m^2 \left(\sum_{i=m}^{\infty} i^{-2} \right) \mathbb{1}_{\{m-1 \leq |X_1| < m\}} d\mathbb{P} \\ &\leq m^2 \frac{1}{m-1} \mathbb{P}[m-1 \leq |X_1| < m] \\ &< 2m \mathbb{P}[m-1 \leq |X_1| < m]. \end{aligned}$$

Donc $\int \left(\sum_{i=1}^{\infty} \frac{X_1^2}{i^2} \mathbb{1}_{\{|X_1| < i\}} \right) d\mathbb{P} \leq 2 \sum_{m=1}^{\infty} m \mathbb{P}[m-1 \leq |X_1| < m] \leq 2(\mathbb{E}|X_1| + 1) < +\infty.$ $\square$

2.5 Exercices

Exercice 2.1

Soit (ξ_n) une suite de v.a. absolument continue (a.c.) ayant pour densité $f_n(t) = \frac{n}{\pi} \frac{1}{1+n^2 t^2}$, $t \in \mathbb{R}$. Montrer que (ξ_n) converge en probabilité. Est-elle convergente en moyenne d'ordre $r \geq 1$?

Exercice 2.2

Soit $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R})$ l'ensemble des v.a. définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}$. Pour tout couple (X, Y) éléments de $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R})$, on pose $d(X, Y) = \int_{\Omega} \frac{|X - Y|}{1 + |X - Y|} d\mathbb{P}$

1. Montrer que d ainsi définie est une semi-distance sur $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R})$.
2. Soit $X \in \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R})$ et (X_n) est une suite d'éléments de $\mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P}, \mathbb{R})$.
Démontrer que $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$ si et seulement si $\lim_{n \rightarrow +\infty} d(X_n, X) = 0$.

Exercice 2.3

1. Montrer que si $X_n \xrightarrow[n \rightarrow +\infty]{L^r} X$, $r \geq 1$ alors $\mathbb{E}|X_n|^r \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}|X|^r$.
2. Montrer que si $X_n \xrightarrow[n \rightarrow +\infty]{L^1} X$, alors $\mathbb{E}X_n \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}X$. La réciproque est-elle vraie ?
3. Montrer que si $X_n \xrightarrow[n \rightarrow +\infty]{L^2} X$, alors $\text{Var } X_n \xrightarrow[n \rightarrow +\infty]{} \text{Var } X$.
4. Montrer que si $|X_n| \leq Z$, $\forall n$ avec $\mathbb{E}(Z) < \infty$, $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X \implies X_n \xrightarrow[n \rightarrow +\infty]{L^1} X$.

Exercice 2.4

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilités. Déterminer pour chacune des convergences suivantes à quelle condition sur la suite $(A_n)_{n \geq 1}$ elle a lieu.

1. La suite $(\mathbb{1}_{A_n})_{n \geq 1}$ converge en probabilité vers 0.
2. La suite $(\mathbb{1}_{A_n})_{n \geq 1}$ converge dans L^2 vers 0.
3. La suite $(\mathbb{1}_{A_n})_{n \geq 1}$ converge en presque sûrement vers 0.

Exercice 2.5 (Inégalité de Chernoff)

Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes et de même loi $\mathcal{N}(m, \sigma)$. On pose $\varphi(u) = \mathbb{E}(e^{u(X_1 - m)})$ et $S_n = X_1 + \dots + X_n$.

1. Calculer $\varphi(u)$.
2. Montrer que $\mathbb{P}[S_n - nm \geq a] \leq e^{-ua}(\varphi(u))^n$, $\forall a \in \mathbb{R}$, $\forall u \in \mathbb{R}^+$.
3. Montrer que $\mathbb{P}[S_n - nm \leq a] \leq e^{-ua}(\varphi(u))^n$, $\forall a \in \mathbb{R}$, $\forall u \in \mathbb{R}^-$.
4. En déduire que $\forall \varepsilon > 0$, $\mathbb{P}\left[\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right] \leq 2e^{-n \frac{\varepsilon^2}{2\sigma^2}}$.

Exercice 2.6

Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires de loi exponentielle de paramètre 1.

1. Montrer que

$$\mathbb{P} \left(\limsup_n \frac{X_n}{\log n} > 1 \right) = 0$$

On suppose que désormais que les v.a. $X_1, \dots, X_n$ sont indépendantes.

2. Montrer que

$$\mathbb{P} \left(\limsup_n \frac{X_n}{\log n} < 1 \right) = 0.$$

Montrer que ce résultat peut être faux sans l'hypothèse d'indépendance.

3. Montrer que $\limsup_n \frac{X_n}{\log n}$ est presque sûrement égale à une variable constante que l'on déterminera.

4. Montrer que $\liminf_n \frac{X_n}{\log n}$ est presque sûrement égale 0.

Exercice 2.7

Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. uniformément bornée $\left(\sup_{n, \omega} |X_n(\omega)| < \infty \right)$ et $S_n = X_1 + \dots + X_n$.

1. Montrer que si $\frac{S_n}{n^2} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ alors $\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$.

2. On suppose de plus que $X_1, X_2, \dots$ sont indépendantes et centrées. Montrer sans utiliser la LFGN que $\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$.

Exercice 2.8

Montrer que si les variables aleatoires réelles $(X_n)_{n \geq 0}$ sont indépendantes, la série $\sum_{n \geq 0} X_n$ converge ou diverge presque sûrement

Exercice 2.9

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite v.a. indépendantes de moment d'ordre deux fini

1. Montrer que $\mathbb{P} \left[\max_{1 \leq k \leq n} |S_k - \mathbb{E}(S_k)| \geq \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{j=1}^n \text{Var}(X_j)$.

2. On suppose de plus, à présent que $\sum_{k=1}^{+\infty} \text{Var}(X_k) < +\infty$

Montrer alors que $\sum_{n=1}^{+\infty} (X_n - \mathbb{E}(X_n))$ converge p.s.

3. En déduire alors que si $\sum_n \text{Var}(X_n) < +\infty$ et $\sum_n \mathbb{E}(X_n) < +\infty$, $\sum_n X_n$ converge p.s.

Chapitre 3

Vecteurs Gaussiens

3.1 Définition

Définition 3.1.1

Soit X un v.a. dans $\mathbb{R}^n$. On dit que X est normal ou gaussien s'il existe $m \in \mathbb{R}^n$ et $\Gamma \in \mathcal{M}_n(\mathbb{R})$ symétrique positive telle que

$$\Phi_X(t) = \exp(i \langle t, m \rangle) \exp\left(-\frac{1}{2} \langle \Gamma t, t \rangle\right), \quad t \in \mathbb{R}^n$$

où $\langle t, m \rangle = \sum_{i=1}^n t_i m_i$, $\langle \Gamma t, t \rangle = t' \Gamma t = \sum_{1 \leq i, j \leq n} \gamma_{ij} t_i t_j$, γ_{ij} étant les éléments de la matrice.

On note $X \rightsquigarrow \mathcal{N}(m, \Gamma)$.

Théorème 3.1.1

Un vecteur aléatoire $X = (X_1, \dots, X_n)'$ de $\mathbb{R}^n$ est gaussien si et seulement si pour tout $(a_1, \dots, a_n)' \in \mathbb{R}^n$, la v.a. $Y = \sum_{i=1}^n a_i X_i = \langle a, X \rangle$ est une variable aléatoire réelle gaussienne.

Preuve

Supposons que X soit un vecteur gaussien et soit $t \in \mathbb{R}$ et posons $a = (a_1, \dots, a_n)$.

$$\begin{aligned} \Phi_Y(t) &= \mathbb{E}(e^{itY}) = \mathbb{E}\left(e^{it \sum_{i=1}^n a_i X_i}\right) = \mathbb{E}(e^{i \langle ta, X \rangle}) \\ &= \Phi_X(ta) \\ &= e^{i \langle m, ta \rangle} e^{-\frac{1}{2} \langle \Gamma(ta), ta \rangle} \\ &= e^{it \langle m, a \rangle} e^{-\frac{1}{2} t^2 \langle \Gamma a, a \rangle}. \end{aligned}$$

$\Phi_Y(t)$ est donc la f.c. d'une loi normale sur $\mathbb{R}$ de moyenne $\langle m, a \rangle = \sum_{i=1}^n a_i m_i$ et de variance

$$\sigma^2 = \langle \Gamma a, a \rangle = \sum_{i,j=1}^n \gamma_{ij} a_i a_j, \quad \gamma_{ij}, \quad 1 \leq i, j \leq n \text{ étant les éléments de la matrice } \Gamma.$$

En effet si $a_i = 1$ et $a_j = 0$, $\forall j \neq i$, $Y = X_i \rightsquigarrow \mathcal{N}(m_i, \gamma_{ii})$ avec $m_i = \mathbb{E}(X_i)$ et $\gamma_{ii} = \text{Var}(X_i)$.

Si $a_i = a_j = 1$ et $a_k = 0$, $\forall k \neq i, j$, $Y = X_i + X_j \rightsquigarrow \mathcal{N}(m_i + m_j, \sigma_{ij})$ avec

$$m_i + m_j = \mathbb{E}(X_i) + \mathbb{E}(X_j)$$

et

$$\sigma_{ij} = \text{Var}(X_i + X_j) = \text{Var}(X_i) + \text{Var}(X_j) + 2 \text{Cov}(X_i, X_j) = \gamma_{ii} + \gamma_{jj} + 2\gamma_{ij}.$$

Il en résulte que $[\gamma_{ij}]_{1 \leq i, j \leq n} = \Gamma$.

Réciproquement, soit X un vecteur aléatoire $X = (X_1, \dots, X_n)'$ de $\mathbb{R}^n$ tel que toute combinaison linéaire soit gaussienne sur $\mathbb{R}$. Posons $m_i = \mathbb{E}(X_i)$, $\gamma_{ij} = \text{Cov}(X_i, X_j)$, $m = (m_i)_{1 \leq i \leq n}$ et $\Gamma = [\gamma_{ij}]_{1 \leq i, j \leq n}$. Soit $t = (t_1, \dots, t_n) \in \mathbb{R}^n$.

Par hypothèse $\langle t, X \rangle = \sum_{i=1}^n t_i X_i$ est une v.a. réelle gaussienne de moyenne

$$\sum_{i=1}^n t_i \mathbb{E}(X_i) = \sum_{i=1}^n t_i m_i = \langle m, t \rangle$$

et de variance

$$\sigma^2 = \sum_{i,j=1}^n \text{Cov}(t_i X_i, t_j X_j) = \sum_{i,j=1}^n t_i t_j \text{Cov}(X_i, X_j) = \sum_{i,j=1}^n t_i t_j \gamma_{ij} = \langle \Gamma t, t \rangle$$

De plus $\langle \Gamma t, t \rangle = \sigma^2 \geq 0$, c'est-à-dire Γ est du type positif. Il est aussi symétrique et on a $\mathbb{E}(e^{i\langle t, X \rangle}) = \Phi_{\langle t, X \rangle}(1) = e^{i\langle m, t \rangle} e^{-\frac{1}{2}\langle \Gamma t, t \rangle}$, $\forall t \in \mathbb{R}^n$. Donc X est gaussien avec $X \rightsquigarrow \mathcal{N}(m, \Gamma)$

Corollaire 3.1.1

Si $X = (X_1, \dots, X_n) \rightsquigarrow \mathcal{N}(m, \Gamma)$ alors $m = [\mathbb{E}(X_i)]_{1 \leq i \leq n}$ et $\Gamma = [\text{Cov}(X_i, X_j)]_{1 \leq i, j \leq n}$

3.2 Quelques propriétés

Théorème 3.2.1

Soit X un vecteur gaussien de $\mathbb{R}^n$: $X = (X_1, \dots, X_n) \rightsquigarrow \mathcal{N}(m, \Gamma)$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ une fonction affine telle que $g(x) = \varphi(x) + b$ avec $\varphi \in L(\mathbb{R}^n)$ et $b \in \mathbb{R}^p$. Alors $g(X)$ est gaussien de loi $\mathcal{N}(g(m), \varphi \Gamma \varphi')$.

Preuve

$$\begin{aligned} \forall s \in \mathbb{R}^p, \Phi_{g(X)} &= \mathbb{E}(e^{i\langle g(X), s \rangle}) = \mathbb{E}(e^{i\langle \varphi(X) + b, s \rangle}) = e^{i\langle b, s \rangle} \mathbb{E}(e^{i\langle \varphi(X), s \rangle}) \\ &= e^{i\langle b, s \rangle} \mathbb{E}(e^{i\langle X, \varphi' s \rangle}) = e^{i\langle b, s \rangle} \Phi_X(\varphi' s) \\ &= e^{i\langle \varphi(m) + b, s \rangle} e^{-\frac{1}{2}\langle \varphi \Gamma \varphi' s, s \rangle} = e^{i\langle g(m), s \rangle} e^{-\frac{1}{2}\langle \varphi \Gamma \varphi' s, s \rangle} \\ &= \phi_{\mathcal{N}(g(m), \varphi \Gamma \varphi')}(s) \end{aligned}$$

d'où, $g(X) \rightsquigarrow \mathcal{N}(g(m), \varphi \Gamma \varphi')$ d'après le théorème d'unicité.

Remarque 3.2.1

Pour se souvenir de l'emplacement de la transposée dans $\Gamma_Y = A \Gamma_X A'$, il suffit de regarder les dimensions. En effet, le nombre de lignes de A donne la dimension de Y et par conséquent la dimension de la matrice carrée Γ_Y . Ainsi, A et Γ_Y ont le même nombre de lignes.

Exemple 3.2.1

Soit $X = (X_1, \dots, X_4)' \in \mathbb{R}^4$ un vecteur Gaussien tel que

$$m_X = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix} \quad \text{et} \quad \Gamma_X = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{pmatrix}.$$

Montrer que $Y = (X_1 + 2X_2 + 1, X_2 - X_3)' \in \mathbb{R}^2$ est un vecteur gaussien et donner la loi en fonction de m_X et Γ_X .

3.3 Représentation de lois normales sur $\mathbb{R}^n$

Soit $m \in \mathbb{R}^n$, Γ une matrice carrée d'ordre n et $X \rightsquigarrow \mathcal{N}(m, \Gamma)$. Si Γ est du type positif de rang $p < n$, Γ est semblable à une matrice diagonale c'est-à-dire, il existe une base orthonormée de $\mathbb{R}^n$ dans laquelle Γ est semblable à $\tilde{\Gamma} = \text{diag}(\lambda_1, \dots, \lambda_p, 0, \dots, 0)$ avec $\lambda_i > 0$, $\forall i = 1, 2, \dots, p$. Soit P la matrice de passage. P est orthogonale c'est-à-dire $P^{-1} = P'$ et on a $\Gamma = P\tilde{\Gamma}P'$. Posons $Y = P'(X - m)$. Y est gaussienne comme fonction linéaire d'un vecteur gaussien. On a $\mathbb{E}(Y) = 0$ et

$$\begin{aligned}\Phi_Y(s) &= e^{-\frac{1}{2}\langle P'\Gamma P s, s \rangle} = e^{-\frac{1}{2}\langle \tilde{\Gamma} s, s \rangle} = e^{-\frac{1}{2}\sum_{i=1}^p \lambda_i s_i^2} = \prod_{i=1}^p e^{-\frac{1}{2}\lambda_i s_i^2} \\ &= \Phi_{\mathcal{N}(0, \lambda_1)}(s_1) \times \Phi_{\mathcal{N}(0, \lambda_2)}(s_2) \times \dots \times \Phi_{\mathcal{N}(0, \lambda_p)}(s_p) \times \Phi_{\delta_0}(s_{p+1}) \times \dots \times \Phi_{\delta_0}(s_n) \\ &= \Phi_{\mathcal{N}(0, \lambda_1) \otimes \mathcal{N}(0, \lambda_2) \otimes \dots \otimes \mathcal{N}(0, \lambda_p) \otimes \delta_0 \otimes \dots \otimes \delta_0}(s), \quad \forall s \in \mathbb{R}^n.\end{aligned}$$

Donc $Y = (Y_1, \dots, Y_p, 0, \dots, 0)$ avec $Y_1, \dots, Y_p$ indépendantes telles que pour tout $i = 1, 2, \dots, p$ $Y_i \rightsquigarrow \mathcal{N}(0, \sigma_i^2)$, où $\sigma_i^2 = \lambda_i$ est une valeur propre de la matrice carrée symétrique Γ . On a donc $X = PY + m$.

En conclusion, si X est un v.a. gaussien dans $\mathbb{R}^n$, c'est-à-dire $X \rightsquigarrow \mathcal{N}(m, \Gamma_X)$, $m \in \mathbb{R}^n$, $\Gamma_X \in \mathcal{M}_n(\mathbb{R})$ symétrique positive et de rang p , il existe une suite de p v.a.r. $Y_1, Y_2, \dots, Y_p$ indépendantes, de loi $\mathcal{N}(0, \lambda_i)$, $i = 1, 2, \dots, p$ respectivement où les λ_i sont les valeurs propres non nulles de Γ_X et une matrice orthogonale P telles que $Y = (Y_1, \dots, Y_p, 0, \dots, 0)$ et $X = PY + m$. Réciproquement, on a

Théorème 3.3.1

$\forall m \in \mathbb{R}^n$ et $\forall \Gamma \in \mathcal{M}_n(\mathbb{R})$ symétrique positive, il existe un vecteur aléatoire X dans $\mathbb{R}^n$ tel que $X \rightsquigarrow \mathcal{N}(m, \Gamma)$.

Exemple 3.3.1

Montrer qu'il existe un vecteur gaussien de matrice de covariance et de vecteur moyenne

$$\Gamma_Y = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{et} \quad m_X = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

3.4 Singularité et absolue continuité des lois gaussiennes

Rappel

Une mesure ν est dite étrangère à une mesure μ (on dit aussi ν est μ -singulière) s'il existe N tel que $\mu(N) = 0$ et $\nu(N) = 1$. En général toute mesure discrète sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ est étrangère à la mesure de Lebesgue sur $\mathbb{R}^n$, $n > 1$. D'autre part on dit que ν est absolument continue par rapport à μ s'il existe une fonction mesurable positive f telle que $\frac{d\nu}{d\mu} = f$ (Théorème de Radon-Nikodym) et f est appelée densité de ν par rapport à μ .

Théorème 3.4.1

Soit X un vecteur gaussien dans $\mathbb{R}^n$ de moyenne m et de variance Γ_X

1. Si $\text{rg}(\Gamma_X) < n$, alors X reste presque sûrement dans un sous-espace affine de $\mathbb{R}^n$ de dimension $\text{rg}(\Gamma_X)$ et P_X est étrangère à la mesure de Lebesgue sur $\mathbb{R}^n$.

2. Si $rg(\Gamma_X) = n$ alors X admet une densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ donnée par

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} < \Gamma_X^{-1}(x - m), (x - m) > \right\}, \quad \forall x \in \mathbb{R}^n$$

Preuve

1. Soit $p = rg(\Gamma_X) < n$ et $Y = (Y_1, \dots, Y_p, 0, \dots, 0)' \in E_p = \mathbb{R}^p \times \{0\}^{n-p}$ telle que $X = PY + m$ comme dans la représentation précédente. Alors X reste presque sûrement dans $V_X = PE_p + m$ qui est une variété de dimension p . La loi de X est donc portée par V_X et donc $\lambda(V_X) = 0$ où λ est la mesure de Lebesgue sur $\mathbb{R}^n$. Par suite P_X est étrangère à λ .
2. Si $p = n$, on a $Y = (Y_1, \dots, Y_n)'$ avec $Y_i \in \mathcal{N}(0, \sigma_i^2)$, $\sigma_i^2 = \lambda_i$, $i = 1, \dots, n$ et $(Y_1, \dots, Y_n)$ indépendante. Donc Y admet pour densité

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} < \tilde{\Gamma}_X^{-1} y, y > \right\}, \quad \forall y \in \mathbb{R}^n.$$

Mais $X = PY + m = \Theta(Y) \implies Y = \Theta^{-1}(X) = P'(X - m)$ et $\Gamma_X = P\tilde{\Gamma}_X P'$.

Comme P est orthogonale donc $\det P = 1$, P est un isomorphisme. D'après la formule de changement de variable, $\forall x \in \mathbb{R}^n$,

$$\begin{aligned} f_X(x) &= f_Y(\Theta^{-1}(x)) |J_{\Theta^{-1}}| = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} < \tilde{\Gamma}_X^{-1} P'(x - m), P'(x - m) > \right\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} < P\tilde{\Gamma}_X^{-1} P'(x - m), (x - m) > \right\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} < \Gamma_X^{-1}(x - m), (x - m) > \right\}. \quad \square \end{aligned}$$

Remarque 3.4.1

Pour $n = 1$, X gaussienne sur $\mathbb{R}$ de moyenne $m \in \mathbb{R}$ et de variance $\Gamma = \sigma^2$ si et seulement si,

$$\Phi_X(t) = e^{itm} e^{-\frac{1}{2}\sigma^2 t^2}$$

Pour $\sigma = 0$, $\Phi_X(t) = e^{itm} = \Phi_{\delta_m}(t)$ et $X \rightsquigarrow \delta_m$ où δ_m est la mesure de Dirac en m , par conséquent $X = m$ p.s. On dit que X ou la loi de X est dégénérée.

Proposition 3.4.1 (Indépendance dans un vecteur gaussien)

Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur gaussien.

- i) Pour tout $i \neq j \in \{1, \dots, n\}$, X_i et X_j sont indépendantes si et seulement $\text{Cov}(X_i, X_j) = 0$
- ii) Soit $I = \{i_1 < \dots < i_k\} \subset \{1, \dots, n\}$. Les variables aléatoires sont mutuellement indépendantes si et seulement si $\text{Cov}(X_i, X_j) = 0$, $\forall i \neq j \in I$

3.5 Exercices

Exercice 3.1

Soit (X, Y) un vecteur gaussien centré, avec $\mathbb{E}(X^2) = 4$ et $\mathbb{E}(Y^2) = 1$ et tel que les variables $2X + Y$ et $X - 3Y$ sont indépendantes.

1. Déterminer la matrice de covariance de (X, Y) .

- Montrer que le vecteur $(X+Y, 2X-Y)$ est également gaussien, puis déterminer sa matrice de covariance.

Exercice 3.2

Soit $(X, Y, Z) \in \mathbb{R}^3$ un vecteur gaussien. On pose $U = X + Y + Z$ et $V = X - Y$.

- Montrer que $(U, V) \in \mathbb{R}^2$ est gaussien.
- A quelle condition sur la matrice de covariance de (X, Y, Z) les variables U et V sont-elles indépendantes ?

Exercice 3.3

Soit $X = (X_1, X_2)$ un vecteur gaussien centré. On donne $\mathbb{E}(X_1^2) = a$, $\mathbb{E}(X_1, X_2) = b$ et $\mathbb{E}(X_2^2) = c$.

- Calculer la matrice de covariance et la fonction caractéristique de X .
- Donner une condition nécessaire et suffisante sur (a, b, c) pour que cette loi possède une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ et déterminer cette densité.

Exercice 3.4

Soit $(X, Y, Z)' \in \mathbb{R}^3$ un vecteur gaussien d'espérance $(1, 1, 1)'$ et de matrice de covariance $2I_3$. Le vecteur $(X + 2Y + Z, 2X - Y + Z + 2)'$ est-il gaussien ? Déterminer sa loi.

Exercice 3.5

Soit $X \in \mathbb{R}^3$ un vecteur gaussien centré de matrice de covariance

$$Q = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

- X possède-t-il une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^3$? Si oui donner son expression.
- Trouver un opérateur linéaire $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ tel que les composantes du vecteur $A.X$ soient des variables indépendantes.
- Déterminer la loi de $X_1 + 2X_2 - X_3$ où $X = (X_1, X_2, X_3)$.

Exercice 3.6

Soit $(X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur gaussien de loi $\mathcal{N}(0, I_n)$.

- Déterminer la loi de (X_1^2) .
- Déterminer les lois de $(X_1^2 + X_2^2 + \dots + X_n^2)$ et de $\sqrt{X_1^2 + X_2^2 + \dots + X_n^2}$.
- Soit Y une v.a. telle que $(X_1, \dots, X_n, Y) \in \mathbb{R}^{n+1}$ soit un vecteur gaussien de loi $\mathcal{N}(0, I_{n+1})$. Déterminer la loi de la variable

$$\frac{Y}{\sqrt{X_1^2 + X_2^2 + \dots + X_n^2}}.$$

Exercice 3.7

- Soit $a \in \left[-\frac{\pi}{4}; \frac{\pi}{4}\right]$. Déterminer le carré de la matrice

$$\begin{pmatrix} \cos a & \sin a \\ \sin a & \cos a \end{pmatrix}$$

2. Soit Q la loi gaussienne sur $\mathbb{R}^2$, centrée, dont la matrice de covariance est

$$\begin{pmatrix} 1 & \sin 2a \\ \sin 2a & 1a \end{pmatrix}$$

3. Montrer que Q est la loi image de la loi gaussienne centrée réduite $\mathcal{N}(0, I_2)$ sur $\mathbb{R}^2$ par une application linéaire A dont on donnera la matrice.
4. Soit $Y = (Y_1, Y_2)$ un vecteur aléatoire de $\mathbb{R}^2$ dont la loi est Q . Déterminer l'espérance de Y_1^n , où n est un entier naturel quelconque.

Chapitre 4

Introduction aux processus aléatoires

Dans ce chapitre, nous nous intéressons à des phénomènes dont les observations dépendent du temps. En effet, le temps joue une part essentielle dans la vie au cours de laquelle beaucoup de quantités se développent de façon aléatoire lorsque le temps passe, entraînant des modèles de calculs de probabilités qui peuvent vite se compliquer. De telles quantités aléatoires dépendant du temps sont appelées processus aléatoires ou stochastiques qui sont de types différents. De façon formelle, on appelle processus aléatoire une famille de v.a. $X = \{X(t) : t \in T\}$ à valeurs dans un espace $\mathcal{X}$ où T est un ensemble quelconque. Lorsque $T \subset \mathbb{N}$ on dit que le processus X est à temps discret et lorsque T est un intervalle de $\mathbb{R}^+$ on dit que X est un processus à temps continu. Lorsque $\mathcal{X} = \mathbb{R}$, X est dit processus réel. L'application $T \longrightarrow \mathcal{X}, t \longmapsto X(t)$ est appelée trajectoire du processus et pour tout $t \in T$, $\omega \longmapsto X(t, \omega)$ est appelé état du processus au temps t . Nous étudions ici trois exemples de processus aléatoires courant : le processus de Poisson, la marche aléatoire et les chaînes de Markov.

4.1 Processus de Poisson

Le processus de Poisson sur une droite est un processus à temps continu et à valeurs entières positives. On dit encore que c'est un processus de comptage, que l'on note $\{N_t : t \geq 0\}$. Il s'agit d'étudier le nombre aléatoire N_t de certains événements qui se produisent dans un intervalle de temps $[0, t]$ donné. Sa grande popularité dans les applications vient notamment du fait que beaucoup de calculs le concernant sont explicites.

4.1.1 Définition et premières propriétés

Supposons qu'à partir de l'instant $t = 0$, on observe un phénomène aléatoire tels qu'un appel téléphonique au standard d'un hôtel ou l'arrivée d'un client dans un magasin.

Soit N_t le nombre de phénomènes observés dans l'intervalle de temps $]0, t[$ avec $N_0 = 0$. $(N_t)_{t \geq 0}$ est à valeurs dans $\mathbb{N}$ vérifiant les hypothèses suivantes :

1. $\forall 0 \leq s \leq t$, $N_s \leq N_t$ ($t \longmapsto N_t$ est croissante).
2. $\forall 0 \leq s$, le nombre de phénomènes $N_t - N_s$ observés dans l'intervalle $]s, t]$ est indépendant du nombre de phénomènes N_s observés dans l'intervalle $]0, s]$ (on dit que le processus $(N_t)_{t \geq 0}$ est à accroissements indépendants).
3. $\forall s \geq 0$, la loi de $N_t - N_s$ ne dépend que de la longueur de l'intervalle $]s, t]$: $\mathcal{L}(N_t - N_s) = \mathcal{L}(N_{t-s})$ (on dit que le processus $(N_t)_{t \geq 0}$ est à accroissements stationnaires).
4. On ne peut observer deux phénomènes ou plus simultanément : $\mathbb{P}[N_{t+h} - N_t > 1] = o(h)$ quand $h \longrightarrow 0$ (on dit que le processus $(N_t)_{t \geq 0}$ est localement continu en probabilité).

Le Théorème suivant est admis :

Théorème 4.1.1

Il existe un réel $\lambda > 0$ appelé *taux d'arrivée* ou *intensité* du processus telle que $\forall t > 0$, $N_t \rightsquigarrow \mathcal{P}(\lambda t)$ (loi de Poisson de paramètre λt), c'est à dire

$$\mathbb{P}[N_{s+t} - N_s = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Définition 4.1.1

Le processus $(N_t)_{t \geq 0}$ satisfaisant les hypothèses 1 à 4 est appelé *processus de Poisson d'intensité λ* .

Voici quelques propriétés que vérifie un processus de Poisson.

Propriété 4.1.1

Soit $(N_t)_{t \geq 0}$ un processus de Poisson de d'intensité $\lambda > 0$. Alors, lorsque h tend vers 0,

- i) $P[N_h = 1] = \lambda h + o(h)$.
- ii) $P[N_h = 0] = 1 - \lambda h + o(h)$.

Preuve

- i) $P[N_h = 1] = e^{-\lambda h}(\lambda h) = \lambda h + o(h)$.
- ii) $P[N_h = 0] + P[N_h = 1] + P[N_h \geq 2] = 1$. Or,

$$\begin{aligned} P[N_h \geq 2] &= \sum_{k \geq 2} P[N_h = k] = e^{-\lambda h} \sum_{k \geq 2} \frac{(\lambda h)^k}{k!} = e^{-\lambda h} (\lambda h)^2 \sum_{k \geq 0} \frac{(\lambda h)^k}{(k+2)!} \\ &\leq e^{-\lambda h} (\lambda h)^2 \sum_{k \geq 0} \frac{(\lambda h)^k}{k!} = (\lambda h)^2. \text{ D'où, } P[N_h \geq 2] = o(h) \end{aligned}$$

4.1.2 La seconde définition du processus de Poisson

Soit T_n l'instant d'arrivée du $n^{\text{ième}}$ phénomène et X_n la durée séparant le $(n-1)^{\text{ième}}$ phénomène du $n^{\text{ième}}$, $\forall n \geq 1$.

On a

$$T_n = \sum_{i=1}^n X_i \text{ et } X_n = T_n - T_{n-1}, i = 1, 2, \dots$$

Ainsi la connaissance de la famille $(T_n)_{n \geq 0}$ équivaut à celle de la famille $(X_n)_{n \geq 1}$.

D'autre part, $\{T_n \leq t\}$ signifie que le $n^{\text{ième}}$ phénomène a eu lieu à l'instant t ou avant, c'est-à-dire qu'à l'instant t , au moins n phénomènes ont eu lieu, c'est-à-dire que $\{N_t \geq n\}$. Ainsi

$$F_{T_n}(t) = \mathbb{P}(T_n \leq t) = \mathbb{P}(N_t \geq n) = 1 - \sum_{k=0}^{n-1} \mathbb{P}(N_t = k)$$

et

$$\mathbb{P}(N_t = n) = \mathbb{P}(N_t \geq n) - \mathbb{P}(N_t \geq n+1) = \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t).$$

On a également

$$T_n = \inf\{t \geq 0 \mid N_t = n\} \quad (4.1)$$

Alors $0 = T_0 \leq T_1 \leq \dots \leq T_n \leq \dots$ et

$$N_t = n \Leftrightarrow t \in [T_n, T_{n+1}[\quad (4.2)$$

$(T_n)_{n \geq 0}$ est une suite de v.a. dont les valeurs déterminent complètement le processus $(N_t)_{t \geq 0}$. En effet, si on connaît les T_n alors N_t est défini par (4.2) et (4.2) peut se résumer à

$$N_t = \max\{n \geq 0 \mid T_n \leq t\} \quad (4.3)$$

ou encore

$$N_t = \sum_{k=0}^{\infty} 1_{\{T_k \leq t\}}. \quad (4.4)$$

Les v.a. X_n sont appelées durées inter-arrivées du processus.

Théorème 4.1.2

Le processus $(N_t)_{n \geq 0}$ est un processus de Poisson d'intensité λ si et seulement si les durées inter-arrivées du processus sont indépendantes de même loi exponentielle $\mathcal{E}(\lambda)$.

Preuve

Supposons que $N_t \rightsquigarrow \mathcal{P}(\lambda t)$.

$\mathbb{P}[X_1 > t] = \mathbb{P}[N_t = 0] = e^{-\lambda t}$, $\forall t \geq 0$. On a bien donc X_1 qui suit $\mathcal{E}(\lambda)$.

$$\begin{aligned} \mathbb{P}([X_2 > t] \mid [X_1 = t_1]) &= \mathbb{P}([N_{t_1+t} = 1] \mid [N_{t_1} = 1] \cap [N_s = 0, \text{ pour } s < t_1]) \\ &= \mathbb{P}([N_{t_1+t} - N_{t_1} = 0] \mid [N_{t_1} = 1] \cap [N_s = 0, \text{ pour } s < t_1]) \\ &= \mathbb{P}([N_{t_1+t} - N_{t_1} = 0]) \text{ (l'indépendance des accroissements)} \\ &= \mathbb{P}([N_t = 0]) \text{ (stationnarité)} \\ &= e^{-\lambda t}. \end{aligned}$$

Donc X_2 est bien indépendante de T_1 et de même loi exponentielle $\mathcal{E}(\lambda)$.

De façon plus générale,

$$\begin{aligned} \mathbb{P}([X_k > t] \mid [X_1 = t_1] \cap \dots \cap [X_{k-1} = t_{k-1}]) &= \mathbb{P}([N_{t_{k-1}+t} - N_{t_{k-1}} = 0]) \\ &= \mathbb{P}[N_t = 0] = e^{-\lambda t}. \end{aligned}$$

Donc X_k est indépendante de $X_1, \dots, X_{k-1}$ et de même loi exponentielle $\mathcal{E}(\lambda)$.

La réciproque sera admise.

Conséquence : Les variables aléatoires T_n suivent la loi Gamma $\gamma(\lambda, n)$ de densité définie par

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1} \mathbf{1}_{]0; +\infty[}(t).$$

4.2 Marche aléatoire sur $\mathbb{Z}$

Une particule se trouve à l'instant $t = 0$ à l'origine 0 d'un axe gradué $0, \pm 1, \pm 2, \dots$. A chaque instant $t \in 1, 2, 3, \dots$ elle avance d'un cran à droite avec probabilité p , $0 < p < 1$ ou recule vers la gauche avec probabilité $q = 1 - p$. Soit $S_0 = 0$ et $\forall n \geq 1$, S_n la position de la particule à la date n . On a

$$S_{n+1} = \begin{cases} S_n + 1 & \text{avec probabilité } p \\ S_n - 1 & \text{avec probabilité } q \end{cases}$$

On suppose que les déplacements d'une position à l'autre sont indépendants. On peut donc écrire

$$S_n = S_0 + X_1 + \cdots + X_n, n = 0, 1, 2, \dots$$

où S_0 est la position initiale de la particule sur l'axe et $X_1, X_2, \dots$, sont des v.a. indépendantes prenant chacune la valeur $+1$ avec probabilité p et la valeur -1 avec probabilité q .

Le processus $(S_n)_{n \geq 0}$ est appelé marche aléatoire simple. Elle est dite symétrique si $p = q = \frac{1}{2}$ et antisymétrique sinon.

Cet exemple est un modèle de la marche aléatoire parmi tant d'autres tels un jeu dans un casino où S_0 représente la fortune initiale du joueur et S_n sa fortune à l'issue du même jeu, le développement d'une épidémie (marche aléatoire sur $\mathbb{Z}^n$) ou l'indice de prix dans un marché financier.

À présent, nous allons introduire une classe très importante de processus stochastiques : les chaînes de Markov.

4.3 Introduction aux chaînes de Markov

De manière informelle, une chaîne de Markov décrit un système dont l'évolution aléatoire est telle que la loi du système dans le futur ne dépend que de son état présent et pas de son histoire.

4.3.1 Définitions et exemples

Soit $X_0, X_1, X_2, \dots$ une suite de variables aléatoires prenant valeur dans un ensemble S dénombrable. Nous noterons X le processus stochastique correspondant et $\mathbb{P}$ sa loi.

Définition 4.3.1

Le processus X est une chaîne de Markov s'il possède la propriété de Markov,

$$\mathbb{P}(X = s_n | X_0 = s_0, X_1 = s_1, \dots, X_{n-1} = s_{n-1}) = \mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1}),$$

pour tout $n \geq 1$ et tout $s_0, s_1, \dots, s_n \in S$.

S est appelé espace des états de la chaîne

Les marches aléatoires de la section précédente fournissent un exemple de chaîne de Markov, avec $S = \mathbb{Z}^d$.

Définition 4.3.2

Une chaîne de Markov X est homogène si

$$\mathbb{P}(X_n = j | X_{n-1} = i) = \mathbb{P}(X_1 = j | X_0 = i), \quad \text{pour tout } n, i, j.$$

Dorénavant, par souci de simplicité, nous allons supposer que S est un ensemble fini et que la chaîne de Markov est homogène. Dans ce cas, on voit que l'évolution de la chaîne est caractérisée par la matrice $P = (p(i, j))_{i, j \in S} = (p_{ij})_{i, j \in S}$ définie par

$$p_{ij} = \mathbb{P}(X_1 = j | X_0 = i).$$

Définition 4.3.3

La matrice P est appelée matrice de transition de la chaîne, et les probabilités p_{ij} sont appelées probabilités de transition (de i à j).

Lemme 4.3.1

Une matrice de transition est caractérisée par les deux propriétés suivantes :

1. $p_{ij} \geq 0, \forall i, j \in S$
2. $\sum_{j \in S} p_{ij}, \forall i \in S.$

Une matrice possédant ces deux propriétés est appelée une matrice stochastique.

Preuve

Laissée au lecteur. $\square$

Définition 4.3.4

Soit $\mu = (\mu(i))_{i \in S}$ une mesure de probabilité sur S et P une matrice stochastique. La chaîne de Markov (P, μ) est la chaîne de Markov (homogène dans le temps) de matrice de transition P et de loi initiale μ , c'est-à-dire telle que $\mathbb{P}(X_0 = i) = \mu(i)$, pour tout $i \in S$. On écrira simplement $X \rightsquigarrow (P, \mu)$.

Dans la suite, nous utiliserons les notations suivantes : la loi de la chaîne de Markov (P, μ) sera notée $\mathbb{P}_\mu$, et l'espérance correspondante $\mathbb{E}_\mu$. En particulier, lorsque la loi initiale est concentrée sur un état $i \in S$, c'est-à-dire lorsque $\mu = \delta_i \equiv (\delta_{ij})_{j \in S}$, nous écrirons simplement $\mathbb{P}_i$ et $\mathbb{E}_i$.

Remarque 4.3.1

Les moments fini-dimensionnels associés à la chaîne de Markov (P, μ) sont donnés par

$$P_\mu(X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = \mu(s_0)p(s_0, s_1) \dots p(s_{n-1}, s_n), \text{ pour toute suite } s_0, \dots, s_n \in S.$$

Exemple 4.3.1

Après une longue collecte de données, Charles a conçu le modèle suivant pour décrire approximativement le temps qu'il fera sur son île.

$$S = \{\text{beau temps, mauvais temps}\} \quad \text{et} \quad P = \begin{pmatrix} 0,8 & 0,2 \\ 0,4 & 0,6 \end{pmatrix}.$$

La matrice P est stochastique et encode donc bien les probabilités de transition d'une chaîne de Markov sur S . Il est usuel de représenter de telles chaînes par un graphe comme sur la Figure. Vendredi, quant à lui, a élaboré un modèle plus complexe, prédisant le temps du lendemain à partir du temps du jour et de celui de la veille. Le processus X qu'il obtient n'est plus une chaîne de Markov sur S , puisque la propriété de Markov n'est plus vérifiée. Il est cependant possible d'en déduire une chaîne de Markov sur un espace d'états étendu, en l'occurrence $S \times S$, en considérant les variables aléatoires $Y_n = (X_n, X_{n-1})$. En effet, la connaissance du couple $Y_n = (X_n, X_{n-1})$ détermine X_n , et donc il ne reste plus qu'à prédire X_{n+1} , dont la probabilité est fonction uniquement de X_n et X_{n-1} .

La matrice P contient toute l'information sur les probabilités de transition d'un état s au temps n vers un état s' au temps $n+1$. On peut facilement l'utiliser pour déterminer également les probabilités de transition d'un état s au temps m vers un état s' en un temps ultérieur $m+n$ quelconque. Notons

$$p_n(i, j) = \mathbb{P}_i(X_n = j).$$

Alors pour tout $n \geq 1$

$$\begin{aligned}
 p_n(i, j) &= \mathbb{P}_i(X_n = j) \\
 &= \sum_{k \in S} \mathbb{P}_i(X_n = j, X_{n-1} = k) \\
 &= \sum_{k \in S} \mathbb{P}_i(X_n = j | X_{n-1} = k) \mathbb{P}_i(X_{n-1} = k) \\
 &= \sum_{k \in S} \mathbb{P}_k(X_1 = j) \mathbb{P}_i(X_{n-1} = k) \\
 &= \sum_{k \in S} p(k, j) p_{n-1}(i, k).
 \end{aligned}$$

Cette relation est connue sous le nom d'équation de Chapman-Kolmogorov. On en déduit facilement le résultat fondamental suivant.

Théorème 4.3.1

La matrice de transition en n pas, $P_n = (p_n(i, j))_{i, j \in S}$, est donnée par la $n^{\text{ième}}$ puissance de la matrice de transition P ,

$$P_n = P^n.$$

Preuve

On peut réécrire l'équation de Chapman-Kolmogorov sous la forme

$$(P_n)_{ij} = \sum_{k \in S} (P_{n-1})_{ik} (P)_{kj} = (P_{n-1}P)_{ij}$$

En particulier $P_n = P_{n-1}P = P_{n-2}P^2 = \dots = P^n$. $\square$

Il suit que l'on peut facilement exprimer la loi de la chaîne au temps n à partir de la loi de la chaîne au temps 0.

Corollaire 4.3.1

Soit $X \rightsquigarrow (P, \mu_0)$. Alors, la loi de la chaîne au temps n , $\mu_n(i) = P_{\mu_0}(X_n = i)$, $i \in S$ est donnée par

$$\mu_n = \mu_0 P^n.$$

Preuve

$$\begin{aligned}
 \mu_n(i) &= \mathbb{P}_{\mu_0}(X_n = i) = \sum_{j \in S} \mathbb{P}_{\mu_0}(X_n = i | X_0 = j) \mathbb{P}_{\mu_0}(X_0 = j) \\
 &= \sum_{j \in S} p_n(j, i) \mu_0(j) = (\mu_0 P^n)_i. \quad \square
 \end{aligned}$$

Définition 4.3.5

Soit P une matrice stochastique sur un ensemble S .

- Un état $j \in S$ est atteignable depuis l'état $i \in S$, noté $i \rightarrow j$, s'il existe $n \geq 0$ tel que $p_n(i, j) > 0$.
- Un état $i \in S$ est absorbant si $p(i, i) = 1$
- P est irréductible si, pour tout $i, j \in S$, on a $i \rightarrow j$.
- P est absorbante si, pour tout $i \in S$, il existe $j \in S$ absorbant avec $i \rightarrow j$.

Si X est une chaîne de Markov de matrice de transition P , on dira que X est irréductible (resp. absorbante), lorsque P est irréductible (resp. absorbante).

Par la suite, on notera $n(i, j) = \inf\{n \geq 1 : p_n(i, j) > 0\}$ le nombre minimal de pas permettant de passer de i à j avec probabilité positive ; en particulier, $n(i, j) < +\infty$ si et seulement si $i \rightarrow j$.

4.3.2 Distribution stationnaire

Pour une chaîne de Markov irréductible X , le processus ne va pas s'arrêter dans un certain état, mais va continuer à évoluer éternellement. Une question fondamentale est alors de déterminer son comportement asymptotique : si l'on observe une telle chaîne après un temps très long, quelle est la probabilité qu'elle se trouve dans un état donné ? Avec quelle fréquence visite-t-elle chaque état ? La réponse à ces questions est étroitement liée à la notion de distribution stationnaire.

Supposons pour un instant qu'une telle convergence ait lieu, c'est-à-dire que, pour un certain $i \in S$, il existe un vecteur π tel que $\lim_{n \rightarrow +\infty} p_n(i, j)\pi(j)$ pour tout $j \in S$. Alors, on devrait nécessairement avoir, d'une part,

$$\sum_{j \in S} \pi(j) = \lim_{n \rightarrow +\infty} \sum_{j \in S} p_n(i, j) = 1$$

et d'autre part pour tout $k \in S$

$$\sum_{j \in S} \pi(j)p(j, k) = \lim_{n \rightarrow +\infty} \sum_{j \in S} p_n(i, j)p(j, k) = \lim_{n \rightarrow +\infty} p_{n+1}(i, k) = \pi(k).$$

Ceci motive la définition suivante.

Définition 4.3.6

Un vecteur $\pi = (\pi(i))_{i \in S}$ est appelé *distribution stationnaire associée à la matrice de transition X* si

1. $\pi(j) \geq 0$, pour tout $j \in S$ et $\sum_{j \in S} \pi(j) = 1$
2. $\pi = \pi P$.

La raison derrière cette terminologie est la suivante : si $X \rightsquigarrow (P, \pi)$, alors il suit du Théorème 4.3.1 que les probabilités d'occupation au temps n sont données par

$$\pi P^n = (\pi P)P^{n-1} = \pi P^{n-1} = \dots = \pi.$$

On voit donc que la distribution π est stationnaire : elle ne change pas lorsque le temps passe.

4.4 Exercices

Exercice 4.1

Calculer la loi du n -ième temps d'arrivée d'un processus de Poisson de deux façons :

1. En utilisant $[S_n \leq t] = [N_t \geq n]$.
2. En utilisant $S_n = T_1 + T_2 + \dots + T_n$

Exercice 4.2

Soit (X_n) une suite de v.a. indépendantes et de même loi $\mathbb{P}[X = 1] = p$ et $\mathbb{P}[X = -1] = q$, $p + q = 1$. On pose $S_0 = 0$ et $S_n = \sum_{i=1}^n X_i$. Calculer

1. $\mathbb{P}[S_n = 0]$.
2. $\mathbb{P}[\overline{\lim} A_n]$ où $A_n = \{S_n = 0\}$, $n \geq 1$.

Exercice 4.3

Soit $X_1, X_2, \dots, X_n$, n v.a. iid de loi $\mathcal{E}(\lambda)$. On pose $S_0 = 0$ et $\forall k \geq 1$, $S_k = X_1 + X_2 + \dots + X_k$.

1. Trouver la loi du n -uplet $(S_1, S_2, \dots, S_n)$. En déduire celle de S_n , $\forall n \geq 1$.
2. On pose $N(t) = \sum_{n=1}^{+\infty} \mathbb{1}_{\{S_n \leq t\}}$, $\forall t \in \mathbb{R}_+$. Quelle est la loi de $N(t)$?

Exercice 4.4

Soit X la durée de vie aléatoire d'un téléphone portable. On suppose que X est continue admettant une densité f et de fonction de répartition F . On suppose qu'après une panne, le téléphone est réparé et sa durée de vie après réparation a même loi que X . Soient $X_1, X_2, \dots, X_n$ les durées de vie jusqu'à la n^{e} panne toutes indépendantes. On définit $S_n = \sum_{i=1}^n X_i$ le moment de la $(n+1)^{\text{e}}$ panne, $N(t)$ le nombre de pannes survenues dans l'intervalle $[0; t]$ avec $N(0) = 0$, f_n et F_n la densité et la f.r. de la v.a. S_n . montrer que

1. $\mathbb{P}[N(t) = n] = F_n(t) - F_{n+1}(t)$, $\forall n = 0, 1, 2, \dots$.
2. $H(t) = \mathbb{E}[N(t)] = \sum_{n=1}^{+\infty} F_n(t)$.
3. Soit $W(t) = S_{N(t)+1} - t$ (le temps écoulé depuis la date t jusqu'à la 1^{ère} panne après t).
Montrer que si f_n et $\sum f_n$ sont continue sur $\mathbb{R}_+$ alors $\mathbb{P}[W(t) \leq s] = F(t+s) - \int_0^t (1 - F(t+s-x))dH(x)$.
4. Quelle est la loi de $N(t)$ et de $W(t)$ si $X \rightsquigarrow \mathcal{E}(\lambda)$, $\lambda > 0$.

Exercice 4.5 (M.A. sur $\mathbb{Z}^2$)

Soit une M.A. en dimension 2 où la particule se déplace sur les points $n(i, j)$, $i, j = 0, +1, +2, \dots$ du plan. Soit p, q, r, s tels que $0 < p, q, r, s < 1$ et $p+q+r+s = 1$. On suppose que si la particule est à la position (i, j) à la date n , sa position à la date $n+1$ est $(i+1, j)$ avec probabilité p , $(i, j+1)$ avec probabilité q , $(i-1, j)$ avec probabilité r et $(i, j-1)$ avec probabilité s et que les déplacements successifs entre les points sont indépendants. Soit $S_0 = (0, 0)$ et S_n la position de la particule à la date n . On a $\forall n > 0$

$$S_{n+1} = \begin{cases} S_n + (1, 0) & \text{avec proba. } p \\ S_n + (0, 1) & \text{avec proba. } q \\ S_n - (1, 0) & \text{avec proba. } r \\ S_n - (0, 1) & \text{avec proba. } s \end{cases}$$

Soit $u_n = P[S_n = S_0]$. Montrer que

$$u_n = \begin{cases} 0 & \text{si } n \text{ impair} \\ \sum_{k=0}^n \frac{(2n!)}{(k!)^2 ((n-k)!)^2} (pr)^k (qs)^{n-k} & \text{si } n = 2m \text{ est pair} \end{cases}$$

Exercice 4.6

On déplace un pion sur les points indiqués ci-contre selon la loi suivante :

- Au début, le pion est en O
- Lorsqu'il est en O , la probabilité qu'il aille à l'un quelconque des point A_i est $\frac{1}{4}$
- Lorsqu'il est en un point A_i , la probabilité qu'il retourne en O est $\frac{1}{3}$ et la probabilité qu'il aille à l'un des B_i voisins de A_i est $\frac{1}{3}$

- Lorsqu'il est en un point B_i , il va à l'un des A_i voisins avec la probabilité $\frac{1}{2}$.

Soit X la v.a. telle que $\{X = n\}$ si et seulement si le pion revient en O pour la première fois à l'issue de son n^{e} déplacement.

Déterminer la loi de X et calculer $\mathbb{E}(X)$.

Exercice 4.7

Soit $(X_n)_{n \geq 0}$ une Chaîne de Markov Homogène (CHM) de loi initiale $\pi = (1, 0, 0)'$ et de matrice de transition

$$P = \begin{pmatrix} \frac{3}{7} & \frac{3}{7} & \frac{1}{7} \\ \frac{1}{11} & \frac{11}{11} & \frac{1}{11} \\ \frac{1}{11} & \frac{1}{11} & \frac{11}{11} \end{pmatrix}.$$

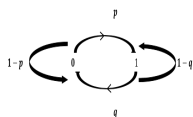
On définit le processus (Y_n) par $Y_n = \begin{cases} 1, & \text{si } X_n = 1 \\ 2, & \text{sinon} \end{cases}, \forall n \geq 0$

Montrer que (Y_n) est une CMH et trouver sa matrice de transition Q .

Exercice 4.8

Soit $(X_n)_{n \geq 0}$ une CHM à espaces d'états $E = \{0; 1\}$ et dont le graphe est

On note P la matrice de transition de la chaîne.



1. Calculer P^n pour tout n et trouver $\lim_{n \rightarrow +\infty} P^n$.

2. Soit π la mesure définie par $\pi_1 = \pi(0) = \frac{q}{p+q}$ et $\pi_2 = \pi(1) = \frac{p}{p+q}$.

Montrer que $\pi P = \pi$ et que $\lim_{n \rightarrow +\infty} P^n(0, 0) = \lim_{n \rightarrow +\infty} P^n(1, 0) = \pi_1$ puis que $\lim_{n \rightarrow +\infty} P^n(0, 1) = \lim_{n \rightarrow +\infty} P^n(1, 1) = \pi_2$.

Exercice 4.9

On dispose de 2 machines identiques fonctionnant indépendamment et pouvant tomber en panne au cours d'une journée avec la probabilité $q = \frac{1}{4}$. On note X_n le nombre de machines en panne au début de la n -ième journée.

1. On suppose que, si une machine est tombée en panne un jour, elle est réparée la nuit suivante et qu'on ne peut réparer qu'une machine dans la nuit. Montrer que l'on peut définir ainsi une chaîne de Markov dont on déterminera le graphe, la matrice de transition et éventuellement les distributions stationnaires.
2. Même question en supposant qu'une machine en panne n'est réparée que le lendemain, le réparateur ne pouvant toujours réparer qu'une machine dans la journée.
3. Le réparateur, de plus en plus paresseux, met maintenant 2 jours pour réparer une seule machine.
 - (a) Montrer que (X_n) n'est plus une chaîne de Markov, mais que l'on peut construire un espace de 5 états permettant de décrire le processus par une chaîne de Markov dont on donnera le graphe des transitions.
 - (b) Calculer la probabilité que les 2 machines fonctionnent après n jours ($n = 1, n = 2$ et $n = 3$) si elles fonctionnent initialement.

Bibliographie

- [1] Alfred Renyi, *Probability Theory*, Dover, 2007.
- [2] Billingsley P., *Convergence of probability measures (2ed.)*, Wiley, 1999.
- [3] Gneyou Kossi E., *Cours de Probabilité (Licence de Mathématiques, semestre 6)*, UL, 2015.
- [4] Dacunha-Castelle Didier et Duflo Marie, *Probabilités et statistiques. Tome 1 Problèmes à temps fixe*, Masson, 1989.
- [5] Dacunha-Castelle Didier et Duflo Marie, *Probabilités et statistiques. Tome 2 Problèmes à temps mobile*, Masson, 2000.
- [6] Marie Cottrell, Valentine Genon-Catalot, Christian Duhamel et Thierry Meyre, *Exercices de probabilités : Licence - Master - Ecoles d'ingénieurs*, Cassini, 2011.
- [7] Geoffrey R. Grimett, David R. Stirzaker, *One thousand exercises in probability*, Oxford, 2001.
- [8] Geoffrey R. Grimett, David R. Stirzaker, *Probability and Random Processes*, Oxford, 2001.