



PROBABILITÉS

Licence de Mathématiques 3ème année

Jean-Christophe BRETON

Université de LA ROCHELLE

Janvier–Mai 2009

version de Juin 2009

Table des matières

1	Variables aléatoires	1
1.1	Rappel de théorie de la mesure	1
1.2	Définition, exemples	2
1.3	Espérance probabiliste	9
1.4	Convergences monotone et dominée	13
1.5	Moments des variables aléatoires	14
2	Vecteurs aléatoires	20
2.1	Rappels d'intégration	20
2.1.1	Théorèmes de Fubini	20
2.1.2	Changement de variable	22
2.2	Vecteurs aléatoires	25
2.3	Indépendance de variables aléatoires	30
2.3.1	Définition	30
2.3.2	Critères et exemples	31
2.3.3	Covariance et indépendance	34
3	Somme de deux variables aléatoires indépendantes et convolution	35
3.1	Convolution de mesures	35
3.2	Convolution de deux fonctions	37
3.3	Loi d'une somme de variables aléatoires à densité indépendantes	39
3.4	Cas de variables aléatoires discrètes indépendantes	41
4	Loi du 0/1 de Kolmogorov	43
4.1	π -système et d -système	43
4.1.1	En théorie de la mesure	43
4.1.2	En probabilité	45
4.2	Tribus du futur et tribu asymptotique	47
4.3	Liminf et limsup d'ensembles	48
4.4	Lemmes de Borel-Cantelli	51
5	Convergences de variables aléatoires	53
5.1	Convergence presque sûre	53
5.2	Convergence en norme p	53

5.3	Convergence en loi	54
5.4	Convergence en probabilité	57
5.5	Lois des grands nombres (LGN)	61
5.5.1	Version faible de la LGN	61
5.5.2	Application : estimation d'une proportion inconnue	62
5.5.3	Version forte de la LGN	64
6	Séries de variables aléatoires indépendantes et loi des grands nombres	69
6.1	Sommes de variables aléatoires indépendantes	69
6.2	Convergence des sommes S_n	70
6.2.1	Les convergences ps et L^1	70
6.2.2	Convergence ps et L^2	72
6.3	Loi des grands nombres (LGN)	74
6.3.1	LGN L^2	74
6.3.2	LGN L^1	75
6.4	Applications	77
6.4.1	Estimateurs	77
6.4.2	Méthode de Monte Carlo	78
7	Fonction caractéristique	80
7.1	Définition et propriétés	80
7.2	Variables et vecteurs gaussien	84
7.3	Théorème central limite (TCL)	91

Chapitre 1

Variables aléatoires

1.1 Rappel de théorie de la mesure

Définition 1.1.1 (Tribu) $\mathcal{A} \subset \mathcal{P}(X)$ est une tribu (ou une σ -algèbre) si

- $X \in \mathcal{A}$
- Si pour tout $i \in \mathbb{N}$ $A_i \in \mathcal{A}$ alors $\cup_{i \in \mathbb{N}} A_i \in \mathcal{A}$: \mathcal{A} est stable par réunion dénombrable.
- Si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$: \mathcal{A} est stable par complémentaire.

On appelle (ensemble) mesurable tout ensemble A élément d'une tribu \mathcal{A} . Un ensemble muni d'une tribu (X, \mathcal{A}) s'appelle un espace mesurable.

Définition 1.1.2 (Tribu borélienne) Lorsque X est un espace topologique (çàd muni d'une famille d'ouverts), la plus petite tribu contenant tous les ouverts est appelée tribu borélienne. Elle est notée $\mathcal{B}(X)$.

Les mesurables $A \in \mathcal{B}(X)$ s'appellent aussi les boréliens. Par définition, les boréliens typiques sont les ouverts, les fermés.

Définition 1.1.3 Une application $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ est dite mesurable si

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

Exemple. Lorsqu'on travaille avec les tribus boréliennes, les fonctions $f : X \rightarrow Y$ continues sont mesurables.

Définition 1.1.4 (Mesure) Une mesure μ sur (X, \mathcal{A}) est une application de $\mathcal{A} \rightarrow [0, +\infty]$ telle que

- $\mu(\emptyset) = 0$
- si $(A_n)_{n \in \mathbb{N}}$ est une suite dénombrable d'ensembles de \mathcal{A} deux à deux disjoints alors

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n) \quad \sigma\text{-additivité.}$$

Le triplet (X, \mathcal{A}, μ) est appelé un espace mesuré (espace mesurable + mesure).

Exemples de mesure.

— Mesure de Dirac sur $(X, \mathcal{P}(X))$: soit $a \in X$,

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A. \end{cases}$$

— Mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$: c'est la mesure qui généralise la notion de longueur des intervalles.

$$\lambda([a, b]) = b - a, \quad \lambda(A + x) = \lambda(A).$$

— Mesure image : Soit $f : (X, \mathcal{A}, \mu) \rightarrow (Y, \mathcal{B})$ une fonction mesurable. On définit sur (Y, \mathcal{B}) la mesure image de f notée μ_f par :

$$\mu_f(B) = \mu(f^{-1}(B)).$$

1.2 Définition, exemples

Grâce à la théorie de la mesure, on unifie la présentation du cadre discret et du cadre continu des variables aléatoires. Par définition, il s'agit maintenant tout simplement des fonctions mesurables sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 1.2.1 *Un espace de probabilité est un espace mesurable (Ω, \mathcal{F}) muni d'une mesure de probabilité \mathbb{P} , c'est à dire une mesure de masse totale 1 : $\mathbb{P}(\Omega) = 1$.*

Les ensembles mesurables $A \in \mathcal{F}$ sont appelés les évènements (ou les observables).

Définition 1.2.2 (variable aléatoire) *On appelle variable aléatoire (va) toute application mesurable X d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ dans \mathbb{R} muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$.*

Il s'agit donc tout simplement d'une application mesurable sur un espace de probabilité.

Définition 1.2.3 (Loi) *Soit $X : \Omega \rightarrow \mathbb{R}$, on appelle loi de X la mesure \mathbb{P}_X , mesure image sur \mathbb{R} de \mathbb{P} par X :*

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\omega \in \Omega : X(\omega) \in A), \quad A \in \mathcal{B}(\mathbb{R}).$$

La loi \mathbb{P}_X d'une variable aléatoire X définit alors une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Il est facile en effet de vérifier que $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$ et que \mathbb{P}_X est σ -additive.

Définition 1.2.4 *Soit X une va. On appelle atome de X (ou de sa loi) tout $x \in \mathbb{R}$ telle que $\mathbb{P}(X = x) \neq 0$.*

Fonction de répartition

Définition 1.2.5 (Répartition) On appelle fonction de répartition d'une v.a. $X : \Omega \rightarrow \mathbb{R}$ la fonction F_X définie sur \mathbb{R} par

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]).$$

Proposition 1.2.1 (Propriétés de la fonction de répartition)

a) F_X est croissante.

b) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

c) F_X est continue à droite et admet une limite à gauche :

$$\lim_{x \rightarrow x_0, x \leq x_0} F_X(x) = \mathbb{P}(X < x_0).$$

On dit que la fonction F_X est cadlag (continue à droite avec une limite à gauche).

d) En fait si x n'est pas un atome de X , alors F_X est continue à gauche (donc continue) en x .

Démonstration : • a) est du à la croissance de la mesure de probabilité \mathbb{P} .

b) et c) s'obtiennent en appliquant les propriétés de monotonie séquentielle des mesures.

• Pour b), prendre d'abord $A_n =]-\infty, -n]$, on a $\bigcap_n A_n = \emptyset$, ensemble de mesure \mathbb{P}_X nulle, si bien que

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq -n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X\left(\bigcap_n A_n\right) = 0.$$

Puis prendre $B_n =]-\infty, n]$ de réunion $\bigcup_n B_n =]-\infty, +\infty[= \mathbb{R}$ de mesure $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = 1$ si bien que

$$\lim_{x \rightarrow +\infty} F_X(x) = \lim_{n \rightarrow +\infty} F_X(n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X\left(\bigcup_n B_n\right) = \mathbb{P}_X(\mathbb{R}) = 1.$$

• Pour le c), prendre d'abord $A_n =]-\infty, x + 1/n]$ d'intersection $\bigcap_n A_n =]-\infty, x]$, ensemble de mesure $\mathbb{P}_X(\bigcap_n A_n) = F_X(x)$ si bien que

$$\lim_{y \rightarrow x^+} F_X(y) = \lim_{n \rightarrow +\infty} F_X(x + 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x + 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X\left(\bigcap_n A_n\right) = F_X(x).$$

Ensuite, prendre $B_n =]-\infty, x - 1/n]$ de réunion $\bigcup_n B_n =]-\infty, x[$, ensemble de mesure $\mathbb{P}_X(\bigcup_n B_n) = \mathbb{P}(X < x)$.

Attention : $\mathbb{P}(X < x)$ peut être distinct de $\mathbb{P}(X \leq x)$ car

$$\mathbb{P}(X \leq x) - \mathbb{P}(X < x) = \mathbb{P}(\{X \leq x\} \setminus \{X < x\}) = \mathbb{P}(X = x)$$

qui peut être non nul si la loi de X a un atome en x . On a alors

$$\lim_{y \rightarrow x^-} F_X(y) = \lim_{n \rightarrow +\infty} F_X(x - 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x - 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X\left(\bigcup_n B_n\right) = \mathbb{P}(X < x).$$

• Pour le d), on constate que si $\mathbb{P}(X = x) = 0$ alors $\mathbb{P}(X < x) = \mathbb{P}(X \leq x)$ et la continuité à gauche manquante vient. \square

Remarque (culturelle). Toute fonction $F : \mathbb{R} \rightarrow [0, 1]$ qui est croissante continue à droite et avec une limite à gauche en tout point et telle que

$$\lim_{t \rightarrow -\infty} F(t) = 0, \quad \lim_{t \rightarrow +\infty} F(t) = 1$$

est la fonction de répartition d'une certaine variable aléatoire X . De plus l'ensemble des points où la fonction F a un saut est l'ensemble des atomes de X .

Variable aléatoire discrète

Définition 1.2.6 (v.a. discrète) Une variable aléatoire X est discrète si elle est à valeur dans un ensemble au plus dénombrable (en bijection avec une partie de \mathbb{N}). Autrement dit : $X(\Omega)$ est fini ou dénombrable (on peut compter ses éléments).

Le support d'une v.a. X discrète est l'ensemble de ses atomes.

Rappel (mesure de Dirac). La mesure de Dirac δ_a en a est définie par

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

Si X est une v.a. discrète alors sa loi est une somme de mesures de Dirac en ses atomes :

$$\mathbb{P}_X = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \delta_x$$

ce qui s'écrit

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{a_i}.$$

Le cas générique est une v.a. X qui prend les valeurs $\{a_i, i \in I\}$, $i \in I$, avec les probabilités respectives p_i , $i \in I$, avec $\sum_{i \in I} p_i = 1$ (où I est fini ou dénombrable). On a

$$X(\Omega) = \{a_i, i \in I\}, \quad \mathbb{P}_X(a_i) = \mathbb{P}(X = a_i) = p_i.$$

Exemples : (v.a. discrètes usuelles)

• Si $X = c$ est une v.a. constante, alors sa loi est $\mathbb{P}_X = \delta_c$ et $X(\Omega) = \{c\}$. En effet

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(c \in A) = \delta_c(A).$$

• Soit X une v.a. qui prend les valeurs a_1, \dots, a_n avec les probabilités respectives p_1, \dots, p_n et $p_1 + \dots + p_n = 1$. Alors son support est $X(\Omega) = \{a_1, \dots, a_n\}$ et sa loi est donnée par :

$$\mathbb{P}_X = p_1 \delta_{a_1} + \dots + p_n \delta_{a_n}.$$

Ici $p_i = \mathbb{P}(X = a_i)$. C'est le cas général d'une loi discrète à support fini.

- Soit X la variable qui indique la face $1, \dots, 6$ obtenue par le lancer d'un dé. La variable X est discrète.

$$\mathbb{P}_X = \frac{1}{6}\delta_1 + \dots + \frac{1}{6}\delta_6.$$

Il s'agit de la loi équirépartie sur $\{1, \dots, 6\}$.

De façon générale, une v.a. X suit la loi équirépartie sur un ensemble fini $\{x_1, \dots, x_n\}$ si $X(\Omega) = \{x_1, \dots, x_n\}$ et $\mathbb{P}_X = \sum_{i=1}^n \frac{1}{n}\delta_{x_i}$.

- Si $A \in \mathcal{F}$ est un évènement, alors $X = \mathbf{1}_A$ est une variable aléatoire. Elle vaut 1 si l'évènement A est réalisé 0 sinon, son support est donc $X(\Omega) = \{0, 1\}$ et sa loi :

$$\mathbb{P}_X = p\delta_1 + (1 - p)\delta_0.$$

Il s'agit de la loi de Bernoulli $b(p)$ si $p = \mathbb{P}(A)$.

De façon générale, une v.a. X suit la loi de Bernoulli de paramètre $p \in [0, 1]$ si elle ne prend que deux valeurs, la plupart du temps 0 et 1 avec :

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p := q.$$

Exemple : pile ou face avec $p = 1/2$ si la pièce est équilibrée, $p \neq 1/2$ si elle est truquée.

- La variable X qui indique le nombre de succès dans une suite de n épreuves chacune ayant une probabilité p de succès est discrète $X(\Omega) = \{0, \dots, n\}$.

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

Il s'agit de la loi binomiale $\mathcal{B}(n, p)$ où p est la probabilité de succès à chaque épreuve.

On rappelle que $C_n^k = \frac{n!}{k!(n-k)!}$ est le coefficient binomial. Il s'agit bien d'une loi de probabilité car la formule du binôme de Newton (d'où le nom de la loi) donne :

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

- Soit X la variable qui indique le numéro du premier lancer où on obtient un 6 lors d'une suite infinie de lancer de dé. La variable X est discrète (l'ensemble des valeurs possibles est $X(\Omega) = \mathbb{N}^*$).

$$\mathbb{P}_X = \sum_{n=1}^{+\infty} \left(1 - \frac{1}{6}\right)^{n-1} \frac{1}{6} \delta_n.$$

Il s'agit de la loi géométrique $\mathcal{G}(1/6)$.

De façon générale, une v.a. X suit la loi géométrique de paramètre $p \in]0, 1[$ notée $\mathcal{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et

$$\mathbb{P}_X = \sum_{n=1}^{+\infty} (1-p)^{n-1} p \delta_n.$$

Exemple : dans une suite infinie d'épreuves indépendantes avec probabilité p de succès à chacune, elle modélise le rang du premier succès.

- Une v.a. X suit la loi de Poisson $\mathcal{P}(\lambda)$ donnée par $X(\Omega) = \mathbb{N}$ et

$$\mathbb{P}_X = \sum_{n=0}^{+\infty} \frac{e^{-\lambda} \lambda^n}{n!} \delta_n.$$

Il s'agit donc d'une v.a. discrète puisque $X(\Omega) = \mathbb{N}$. Cette loi sert à modéliser le temps d'attente dans les files d'attente.

Variables aléatoires à densité

Définition 1.2.7 Soit (X, \mathcal{A}) un espace mesurable muni de deux mesures μ et ν . On dit que μ est absolument continue par rapport à ν si pour tout $A \in \mathcal{A}$, on a

$$\nu(A) = 0 \implies \mu(A) = 0.$$

On le note $\mu \ll \nu$.

Théorème 1.2.1 (de Radon Nikodym) Si $\mu \ll \nu$, alors il existe $f : (X, \mathcal{A}) \rightarrow \mathbb{R}$ mesurable telle que pour tout $A \in \mathcal{A}$

$$\mu(A) = \int_A f d\nu.$$

La fonction f s'appelle la densité de μ par rapport à ν . De plus

- si μ est une mesure finie alors $f \in L^1(\nu)$.
- On a le lien suivant entre les intégrales par rapport à ν et celles par rapport à μ :

$$\int g d\mu = \int g f d\nu.$$

De façon formelle, on écrit « $d\mu = f d\nu$ ».

Les lois des v.a. sont des mesures sur l'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Cet espace a pour mesure de référence la mesure de Lebesgue λ . On peut donc se demander s'il y a une relation d'absolue continuité entre la loi \mathbb{P}_X d'une v.a. X et la mesure de Lebesgue λ sur \mathbb{R} .

Ce n'est évidemment pas toujours vrai. Par exemple la loi de Poisson $\mathcal{P}(\alpha)$ n'est pas absolument continue par rapport à λ puisque si $X \simeq \mathcal{P}(\alpha)$:

$$\mathbb{P}_X(\{n\}) = \frac{\alpha^n e^{-\alpha}}{n!}, \quad \text{alors que} \quad \lambda(\{n\}) = 0.$$

Plus généralement, aucune loi discrète n'est absolument continue par rapport à λ puisque qu'une telle loi \mathbb{P}_X a des atomes :

$$\mathbb{P}_X(x) = \mathbb{P}(X = x) > 0, \quad \text{alors que} \quad \lambda(\{x\}) = 0.$$

Par définition, les lois qui sont absolument continues par rapport à la mesure de Lebesgue sont les lois à densité :

Définition 1.2.8 Une v.a. X est une variable aléatoire de densité f si $\mathbb{P}_X \ll \lambda$ et

$$\mathbb{P}(X \in A) = \int_A f d\lambda, \quad \mathbb{P}(X \in [a, b]) = \int_a^b f(x) dx.$$

Dans les calculs d'intégration, on a alors l'écriture symbolique $d\mathbb{P}_X = f(x) dx$.

Remarque 1.2.1 On observe que la densité f doit vérifier $f(x) \geq 0$ et $\int_{\mathbb{R}} f(x) dx = 1$.

Par exemple,

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad \frac{1}{b-a} \mathbf{1}_{[a,b]}, \quad \alpha e^{-\alpha x} \mathbf{1}_{\mathbb{R}^+}(x), \quad \frac{1}{\pi(1+x^2)}$$

sont les densités respectivement des lois normale standard $\mathcal{N}(0, 1)$, uniforme $\mathcal{U}([a, b])$, exponentielle $\mathcal{E}(\alpha)$ et de Cauchy $\mathcal{C}(1)$. Plus généralement, la loi normale (ou de Gauss) $\mathcal{N}(m, \sigma^2)$ est de densité

$$f_{m, \sigma^2}(x) = \frac{e^{-(x-m)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

Dans le cas à densité, la densité f est reliée à la fonction de répartition F_X de la façon suivante.

Proposition 1.2.2 Si X est une v.a. de densité f , sa fonction de répartition F_X vérifie :

- 1) $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt.$
- 2) F_X est continue sur \mathbb{R} .
- 3) Si f est continue au point x_0 , alors F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$.

D'après 2), la fonction de répartition est toujours continue. De là, vient le nom qu'on donne parfois aux variables aléatoires à densité : variables aléatoires continues.

Démonstration : Puisque X a pour densité f , et comme

$$F_X(b) = \mathbb{P}(X \in]-\infty, b]) = \mathbb{P}(X \in]-\infty, a] \cup]a, b]) = F_X(a) + \mathbb{P}(X \in]a, b]),$$

on a pour tous réels $a < b$:

$$F_X(b) - F_X(a) = \mathbb{P}(X \in]a, b]) = \int_a^b f(t) dt. \quad (1.1)$$

1) Il suffit d'appliquer la monotonie séquentielle des probabilités avec $b = x$ fixé et $a = -n$ pour chaque $n \in \mathbb{N}$ tel que $n > -x$. La suite d'évènements

$$A_n = \{\omega, X(\omega) \in]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et de réunion $A = \{\omega, X(\omega) \in]-\infty, x]\} = \{X \leq x\}$. Par la propriété de continuité monotone séquentielle (ou par convergence dominée), on a $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$, d'où

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \int_{-n}^x f(t) dt = \int_{-\infty}^x f(t) dt$$

en notant que l'intégrale généralisée de la densité f converge en $-\infty$.

2) On fixe $x_0 \in \mathbb{R}$ quelconque. D'abord F_X est continue à droite en tout point car c'est une fonction de répartition.

Il reste à voir la continuité à gauche. Soit $x_n < x_0$ une suite croissante qui converge vers x_0 . Il faut vérifier

$$\lim_{n \rightarrow +\infty} F_X(x_n) = F_X(x_0).$$

On a

$$F_X(x_0) - F_X(x_n) = \int_{x_n}^{x_0} f(t) dt = \int f(t) \mathbf{1}_{[x_n, x_0]}(t) dt.$$

Or $|f(t) \mathbf{1}_{[x_n, x_0]}(t)| \leq f(t)$, intégrable, puisque f est une densité, puis pour presque chaque $t \in \mathbb{R}$, $f(t) \mathbf{1}_{[x_n, x_0]}(t) \rightarrow 0$ puisque $\lim_{n \rightarrow +\infty} \mathbf{1}_{[x_n, x_0]}(t) = \mathbf{1}_{[t_0, t_0]}(t)$.

Le théorème de convergence dominée de Lebesgue s'applique et donne

$$\lim_{n \rightarrow +\infty} F_X(x_0) - F_X(x_n) = \int 0 dt = 0,$$

ce qui est le résultat souhaité.

3) Comme par hypothèse f est continue en x_0 , elle est définie sur tout un voisinage de x_0 et donc sur un intervalle $[a, b]$ qui contient x_0 . La continuité de f en x_0 s'écrit : $\forall \varepsilon > 0, \exists \delta > 0$ tel que $]x_0 - \delta, x_0 + \delta[\subset]a, b[$ et

$$\forall t \in]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon.$$

Pour tout h tel que $0 < |h| < \delta$, on a alors $F_X(x_0 + h) - F_X(x_0) = \int_{x_0}^{x_0+h} f(t) dt$. D'où

$$|F_X(x_0+h) - F_X(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} (f(t) - f(x_0)) dt \right| \leq \int_{x_0}^{x_0+h} |(f(t) - f(x_0))| dt \leq |h|\varepsilon.$$

En divisant par h puis en faisant $h \rightarrow 0$, on constate que F_X est dérivable en x_0 , de dérivée $F_X'(x_0) = f(x_0)$. \square

1.3 Espérance probabiliste

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^+$ une variable aléatoire **positive**.

L'intégrale de X par rapport à la mesure \mathbb{P} est appelée son espérance :

Définition 1.3.1 $\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int X d\mathbb{P}.$

Une variable $X \geq 0$ est dite intégrable si son espérance est finie.

Un exemple de variable aléatoire positive est $X = \mathbf{1}_A$ où $A \in \mathcal{F}$ est un évènement. On a alors

$$\mathbb{E}[X] = \mathbb{E}[\mathbf{1}_A] = \int_{\Omega} \mathbf{1}_A d\mathbb{P} = \mathbb{P}(A).$$

La variable $\mathbf{1}_A$ qui indique si l'évènement A se réalise ou non a pour espérance $\mathbb{P}(A)$.

Définition 1.3.2 Soit X une variable de signe quelconque. Elle est dite intégrable si la va positive $|X|$ est d'espérance (forcément définie) finie. On note alors

$$\mathbb{E}[X] = \int X d\mathbb{P}.$$

La quantité $\mathbb{E}[|X|]$ s'appelle aussi le moment d'ordre 1.

On peut formuler de la façon suivante : X est intégrable si son moment d'ordre 1 est fini.

Définition 1.3.3 Une va X intégrable est dite centrée si $\mathbb{E}[X] = 0$.

Conséquence : des propriétés de l'intégration, on déduit pour des variables aléatoires intégrables X, Y et des réels a, b :

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (linéarité de E).
- Inégalité de Markov : Si X est une v.a. positive $\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$.
- L'espérance n'est rien d'autre que l'intégrale (au sens de Lebesgue) de la fonction mesurable par rapport à la mesure de probabilité \mathbb{P} .

D'après le théorème de transfert, $\mathbb{E}[X]$ peut s'écrire comme l'intégrale par rapport à la loi.

Rappelons d'abord, que si on considère (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurables et $\varphi : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ une fonction mesurable puis une mesure μ sur (X, \mathcal{A}) alors la mesure image $\nu = \mu_{\varphi}$ est une mesure sur (Y, \mathcal{B}) . On a un lien entre les intégrales par rapport à μ sur X et celles par rapport à $\nu = \mu_{\varphi}$ sur Y :

Théorème 1.3.1 (Transfert) Soit $h : (Y, \mathcal{B}) \rightarrow K = (\mathbb{R}, \bar{\mathbb{R}}, \mathbb{C})$ mesurable alors h est ν -intégrable ssi $h \circ \varphi$ est μ -intégrable et

$$\int_X h \circ \varphi d\mu = \int_Y h d\nu.$$

Démonstration : cf. cours de Calcul Intégral.

Il s'agit d'une formule de changement de variable abstraite, très générale puisque la seule condition pour le changement de variable $y = \varphi(x)$ est que φ soit mesurable! Cependant, pour les calculs pratiques, ce résultat n'est pas très utile car la mesure image $\nu = \mu_\varphi$ est mal connue.

Corollaire 1.3.1 Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ une variable aléatoire de loi \mathbb{P}_X . Alors X est \mathbb{P} -intégrable ssi

$$\int_{\mathbb{R}} |x| d\mathbb{P}_X < +\infty.$$

et son espérance est alors

$$\mathbb{E}[X] = \int X d\mathbb{P} = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

Démonstration : Appliquer le théorème de transfert avec la fonction mesurable $\varphi = X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$, la mesure image \mathbb{P}_X et la fonction $h(x) = x$:

$$\int_{\Omega} h \circ X d\mathbb{P} = \int_{\mathbb{R}} h d\mathbb{P}_X.$$

□

Plus généralement, si $h : \mathbb{R} \rightarrow \mathbb{R}$ est borélienne (i.e. mesurable) et $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ est une variable aléatoire. Alors $h(X)$ est une variable aléatoire car c'est une application de Ω dans \mathbb{R}

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} \mathbb{R} \xrightarrow{h} \mathbb{R},$$

c'est à dire $h \circ X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$. Puis la variable $h(X)$ est \mathbb{P} -intégrable ssi h est \mathbb{P}_X -intégrable ($\int_{\mathbb{R}} |h(x)| d\mathbb{P}_X < +\infty$) et alors

$$\mathbb{E}[h(X)] = \int_{\Omega} h(X) d\mathbb{P} = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x).$$

Pour cela, appliquer le théorème de transfert avec $(X, \mathcal{A}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$, $(Y, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\varphi = X$.

Espérance d'une variable discrète

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ avec $X(\Omega)$ discret. La loi de X est donnée par la mesure discrète

$$\mathbb{P}_X = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \delta_x.$$

La loi est une somme de mesures de Dirac : en chaque atome $x \in X(\Omega)$, il y a la masse $\mathbb{P}(X = x)$.

Alors X est intégrable ssi

$$\mathbb{E}[|X|] = \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < +\infty$$

et dans ce cas $\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$ où la somme est au plus dénombrable car $X(\Omega)$ est discret (la v.a. X est discrète).

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est mesurable, alors $h(X)$ est une variable discrète, elle est intégrable ssi

$$\sum_{x \in X(\Omega)} |h(x)| \mathbb{P}(X = x) < +\infty.$$

Son espérance est alors

$$\mathbb{E}[h(X)] = \sum_{x \in X(\Omega)} h(x) \mathbb{P}(X = x).$$

Exemples : (espérance des v.a. discrètes classiques)

- Si $X = c$ est une v.a. constante, sa loi est $\mathbb{P}_X = \delta_c$. Son espérance est

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

On le retrouve aussi avec l'expression $\mathbb{E}[X] = \int_{\mathbb{R}} x d\delta_c(x) = c$.

- Soit X de loi de Bernoulli de paramètre p notée $\mathcal{B}(p)$. Son espérance est

$$\mathbb{E}[X] = p \times 1 + (1 - p) \times 0 = p.$$

- Soit X de loi équirépartie sur l'ensemble fini $\{x_1, \dots, x_n\}$. Son espérance est

$$\mathbb{E}[X] = \frac{x_1 + \dots + x_n}{n}.$$

- Soit X de loi binomiale de paramètres n, p notée $\mathcal{B}(n, p)$. Son espérance est $\mathbb{E}[X] = np$. En effet

$$\mathbb{E}[X] = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = np.$$

- Soit X de loi géométrique de paramètre $p \in]0, 1[$, notée $\mathcal{G}(p)$. Son espérance est $\mathbb{E}[X] = 1/p$. En effet

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} k (1-p)^{k-1} p = \frac{1}{p}.$$

- Soit X de loi de Poisson de paramètre λ . Son espérance est $\mathbb{E}[X] = \lambda$. En effet

$$\mathbb{E}[X] = \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda.$$

Espérance d'une variable à densité

Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ une variable de densité f . La loi de X est donnée par la mesure de forme intégrale

$$\mathbb{P}_X(A) = \int_A f d\lambda = \int_A f(x) dx, \quad \forall A \in \mathcal{B}(\mathbb{R})$$

où f est une fonction mesurable positive d'intégrale 1. Alors X est intégrable ssi

$$\mathbb{E}[|X|] = \int_{\mathbb{R}} |x| f(x) dx < +\infty$$

et dans ce cas $\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx$.

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est mesurable, alors $h(X)$ est une variable aléatoire, elle est intégrable ssi $\int_{\mathbb{R}} |h(x)| f(x) dx < +\infty$. Son espérance est alors

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

Notons que l'intégrale de Lebesgue unifie les 2 cas : v.a. discrètes et v.a. à densité. La différence entre les deux cas n'était donc que formelle.

Exemples : (Lois à densité classiques)

- Loi uniforme

La var. X suit une loi uniforme sur l'intervalle $[a, b]$ ($-\infty < a < b < +\infty$) si elle a une densité f constante sur cet intervalle et nulle en dehors. Sa densité est alors

$$f(t) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(t) = \begin{cases} 1/(b-a) & \text{si } t \in [a, b], \\ 0 & \text{si } t \notin [a, b]. \end{cases}$$

Son espérance est $\mathbb{E}[X] = \frac{b+a}{2}$.

En fait on peut définir une loi uniforme sur un ensemble borélien $A \subset \mathbb{R}$ quelconque (pas forcément un intervalle), c'est la loi de densité $\frac{1}{\lambda(A)} \mathbf{1}_A$.

- Loi exponentielle

La v.a. X suit une loi exponentielle de paramètre $\alpha > 0$, notée $\mathcal{E}(\alpha)$ si elle admet pour densité :

$$f(t) = \alpha e^{-\alpha t} \mathbf{1}_{[0, +\infty[}(t).$$

Elle est utilisée pour modéliser un temps d'attente d'un phénomène aléatoire. Le temps d'attente moyen est alors $\mathbb{E}[X] = 1/\alpha$.

- Loi de Cauchy

Une variable aléatoire réelle suit une loi de Cauchy de paramètre $a \in \mathbb{R}_+^*$ si elle admet pour densité :

$$f(t) = \frac{a}{\pi} \frac{1}{a^2 + t^2}.$$

Son espérance n'est pas définie car $\mathbb{E}[|X|] = \int_{\mathbb{R}} \frac{a|x|}{\pi(a^2 + x^2)} dx = +\infty$, la fonction $x/(1+x^2) \simeq 1/x$ n'est pas intégrable en $\pm\infty$.

- Loi normale (standard)

Une variable aléatoire X_0 suit la loi normale standard $\mathcal{N}(0, 1)$ si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Son espérance est $\mathbb{E}[X_0] = 0$.

Une variable aléatoire X de loi normale $\mathcal{N}(m, \sigma^2)$ peut alors se définir comme une translatée et dilatée de X_0 par

$$X = m + \sigma X_0.$$

Son espérance est $\mathbb{E}[X] = m + \sigma \mathbb{E}[X_0] = m$.

1.4 Convergences monotone et dominée

On rappelle dans un contexte probabiliste les théorèmes fondamentaux de Calcul Intégral : convergence monotone, lemme de Fatou et théorème de convergence dominée. On renvoie au cours de Calcul Intégral pour les preuves.

Définition 1.4.1 *On dit que X_n converge vers X presque sûrement (ps) si l'ensemble des $\omega \in \Omega$ tel que $X_n(\omega) \rightarrow X(\omega)$ est de probabilité 1 :*

$$\mathbb{P}(X_n \rightarrow X) = 1.$$

Théorème 1.4.1 (Convergence monotone, Beppo Levi) *Soit X_n une suite croissante de v.a. positives ($0 \leq X_n \leq X_{n+1}$). Soit $X = \lim_{n \rightarrow +\infty} X_n$ la limite ps de X_n dans $[0, +\infty]$. Alors*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]. \quad (1.2)$$

Lemme 1.4.1 (Fatou) *Soit X_n une suite de v.a. positives. Alors*

$$\mathbb{E}[\liminf_{n \rightarrow +\infty} X_n] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n].$$

Théorème 1.4.2 (Convergence dominée) *Soit X_n une suite de v.a. telle que $X_n \rightarrow X$ ps quand $n \rightarrow +\infty$. S'il existe une v.a. Y intégrable ($\mathbb{E}[|Y|] < +\infty$) telle que pour tout n , $|X_n| \leq Y$ ps alors*

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_n] = \mathbb{E}[X]. \quad (1.3)$$

Conséquence. Si la convergence est dominée, on peut intervertir limite et espérance.

1.5 Moments des variables aléatoires

Définition 1.5.1 Une v.a. $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ a un moment d'ordre $p \geq 1$ ssi

$$\mathbb{E}[|X|^p] = \int_{\Omega} |X|^p d\mathbb{P} < +\infty.$$

Définition 1.5.2 $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R} \mid \mathbb{E}[|X|^p] < +\infty\}$

$L^p(\Omega)$ est un espace vectoriel normé avec pour norme

$$\|X\| = (\mathbb{E}[|X|^p])^{1/p}.$$

Les résultats généraux sur les espaces $L^p(X, \mathcal{A}, \mu)$ se formulent dans le cadre probabiliste de la façon suivante.

Proposition 1.5.1 On dispose de

- *Inégalité de Hölder* : $\|XY\|_1 \leq \|X\|_p \|Y\|_q$ pour p, q exposants conjugués ($1/p + 1/q = 1$).
- *Inégalité de Cauchy-Schwarz* : $\|XY\|_1 \leq \|X\|_2 \|Y\|_2$ ($p = q = 2$).
- *Inégalité de Minkowski* : $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ ($1 \leq p < +\infty$).
- *Si une v.a. est bornée, elle admet des moments de tous les ordres.*
- *Si X possède un moment d'ordre r , pour tout $n \leq r$, X en possède un d'ordre n .*
- *($L^p(\Omega, \mathcal{F}, \mathbb{P}), \|\cdot\|_p$) est un e.v.n. complet, c'est à dire un espace de Banach.*

Exercice : Faire la preuve des points 4 et 5.

En plus du moment d'ordre 1, lié à l'espérance, le plus important est le moment d'ordre 2, lié à la variance.

Définition 1.5.3 (Variance) Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, on définit la variance de X par

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (1.4)$$

On définit aussi l'écart-type $\sigma_X = \sqrt{\text{Var}(X)}$.

Remarque 1.5.1 L'espérance d'une v.a. aléatoire donne la valeur moyenne (au sens probabiliste) de la va. Sa variance (ou son écart-type) mesure la dispersion des valeurs de la v.a. autour de sa moyenne.

Il est équivalent de dire que la variance de X est finie et que X admet un moment d'ordre 2 fini.

La définition de la variance est unifiée entre les deux principaux cas (discret et à densité) grâce à la théorie de la mesure.

• Si X est discrète de domaine $X(\Omega) = \{x_i, i \in I\}$, avec $I = \{1, \dots, n\}$ ou $I = \mathbb{N}^*$, la loi de X est $\mathbb{P}_X = \sum_{i \in I} \mathbb{P}(X = x_i) \delta_{x_i}$ et la variance en (1.4) devient

$$\text{Var}(X) = \sum_{i \in I} (x_i - \mathbb{E}[X])^2 \mathbb{P}_X\{x_i\} = \sum_{i \in I} (x_i - \mathbb{E}[X])^2 \mathbb{P}(X = x_i).$$

• Si X est une v.a. de densité f alors la loi de X est la mesure de densité f , $d\mathbb{P}_X = f(x)dx$ et la variance en (1.4) devient

$$\text{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x) dx.$$

Proposition 1.5.2 (Propriétés de la variance)

- $\text{Var}(X) \geq 0$.
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (Formule de Koenig).
- $\text{Var}(aX) = a^2 \text{Var}(X)$.
- $\text{Var}(X + b) = \text{Var}(X)$ pour toute constante $b \in \mathbb{R}$.
- $\text{Var}(X) = 0$ ssi X est constante ps (et vaut alors $\mathbb{E}[X]$).

La variance est un opérateur quadratique non linéaire.

Démonstration : • Le premier point est clair.

• On développe $\text{Var}(X)$, en notant $\mu = \mathbb{E}[X]$:

$$\begin{aligned} \text{Var}(X) = E[(X - \mu)^2] &= E[(X^2 - 2X\mu + \mu^2)] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X\mu] + \mu^2 \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mu + \mu^2 \\ &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \end{aligned}$$

• Pour les troisième et quatrième points :

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

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

• Si $X = c$ une constante ps alors $\mathbb{E}[X] = \mathbb{E}[c] = c$ et $\mathbb{E}[X^2] = \mathbb{E}[c^2] = c^2$ si bien que $\text{Var}(X) = c^2 - c^2 = 0$. Réciproquement, si $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = 0$ alors la variable $(X - \mathbb{E}[X])^2$, positive d'espérance nulle, est elle même nulle ps, c'est à dire $X = \mathbb{E}[X]$ ps. □

Définition 1.5.4 (Covariance) Soient X, Y deux variables aléatoires avec des variances finies, on définit la covariance de X et de Y par

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

Remarque : • $(X, Y) \mapsto \text{Cov}(X, Y)$ est une application bilinéaire.

- Si X ou Y est centrée alors $\text{Cov}(X, Y) = \mathbb{E}[XY]$.
- $\text{Cov}(X, X) = \text{Var}(X)$.

En quelque sorte la variance Var est une forme quadratique sur $L^2(\Omega)$, d'application bilinéaire associée la covariance Cov .

Proposition 1.5.3 *Si X et Y sont deux v.a. avec des moments d'ordre 2 alors*

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

Démonstration : Il suffit de développer :

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \end{aligned}$$

□

Attention, en général, on n'a pas $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. Prendre par exemple $X = Y$. Par contre on verra que c'est vrai si X et Y sont des v.a. indépendantes.

Proposition 1.5.4

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}.$$

Démonstration : On applique l'inégalité de Cauchy-Schwarz

$$\begin{aligned} |\text{Cov}(X, Y)| &= |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \\ &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2] \mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\text{Var}(X) \text{Var}(Y)}. \end{aligned}$$

Définition 1.5.5 *Soit X, Y deux variables aléatoires, leur coefficient de corrélation est*

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

On constate facilement avec la Proposition 1.5.4 que $\rho(X, Y) \in [-1, 1]$.

Proposition 1.5.5 *Si $\rho(X, Y) = \pm 1$ alors il y a un lien linéaire entre X et Y : $Y = aX + b$, pour $a, b \in \mathbb{R}$. En plus, on montre que*

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad b = \mathbb{E}[Y] - a\mathbb{E}[X].$$

Démonstration : En effet, $\rho(X, Y) = \pm 1$ s'il y a égalité dans l'inégalité de Cauchy-Schwarz de la preuve de la Prop 1.5.4. Il est connu que c'est le cas s'il y a une relation linéaire entre X et Y .

Puis,

$$\begin{aligned}\text{Cov}(X, Y) &= \text{Cov}(X, aX + b) \\ &= a \text{Cov}(X, X) + \text{Cov}(X, b) \\ &= a \text{Var}(X) + 0\end{aligned}$$

car $\text{Cov}(X, b) = \mathbb{E}[Xb] - \mathbb{E}[X]\mathbb{E}[b] = 0$. On en déduit a . Puis comme $Y = aX + b$, on a $\mathbb{E}[Y] = a\mathbb{E}[X] + b$, d'où vient b aussi. \square

Moralement, si $\rho(X, Y)$ est proche de 1 ou -1 , alors c'est que X et Y prennent des valeurs « peu » dispersées par rapport à une relation linéaire affine et on pourra supposer que c'est le cas. On a donc à peu près $Y \simeq aX + b$.

De plus, si ρ est proche de 1 alors $a > 0$ et si ρ est proche de -1 alors $a < 0$.

Théorème 1.5.1 (Inégalité de Tchebychev) *Si $\text{Var}(X)$ existe, on a*

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

Démonstration : Par l'inégalité de Markov, on a

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

\square

Application : On jette 3600 fois un dé. Minorer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 480 et 720.

Notons S le nombre d'apparitions du 1. On peut voir S comme la somme de 3600 v.a. de Bernoulli indépendantes de paramètre $p = 1/6$ (probabilité d'apparition du 1 au cours d'un lancer). Par un raisonnement classique, S suit une loi $\mathcal{B}(3600, p)$. On cherche ici

$$\mathbb{P}(480 < S < 720) = \sum_{k=481}^{719} C_{3600}^k p^k (1-p)^{3600-k}.$$

Ce résultat exact ne peut être calculé en pratique, même un ordinateur très puissant ne pouvant calculer tous ces coefficients binomiaux pour des chiffres aussi grands.

On peut penser à approximer la loi $\mathcal{B}(3600, 1/6)$ par $\mathcal{P}(600)$ mais il resterait à calculer

$$\sum_{k=481}^{719} e^{-600} \frac{600^k}{k!},$$

ce qui n'est pas évident non plus.

On a alors recours à l'inégalité de Tchebychev : notons que $\mathbb{E}[S] = np = 3600/6 = 600$ et $\text{Var}(X) = npq = 3600 \times 5/6 \times 1/6 = 500$. Remarquons de plus que

$$480 < S < 720 \iff -120 < S - 600 < 120.$$

D'où

$$\begin{aligned} \mathbb{P}(480 < S < 720) &= \mathbb{P}(-120 < S - 600 < 120) = \mathbb{P}(|S - 600| < 120) \\ &= 1 - \mathbb{P}(|S - 600| \geq 120) \\ &\geq 1 - \frac{500}{120^2} \\ &\geq 0,95833\dots \end{aligned}$$

Remarque 1.5.2 Les valeurs 480 et 720 sont symétriques par rapport à la moyenne 600 de la v.a. considérée, ce sont 600 ± 120 . Ce n'est pas nécessaire : on peut aussi appliquer l'inégalité de Tchebychev sur un intervalle non centré autour de l'espérance. Il suffit pour cela d'utiliser le plus grand intervalle centré sur l'espérance qu'il contient. Ainsi pour minorer $\mathbb{P}(550 < S < 700)$, il suffit de remarquer que

$$550 < S < 700 \iff \underbrace{550 < S < 650}_{\text{intervalle centré autour de } 600} \iff -50 < S - 600 < 50.$$

et

$$\begin{aligned} \mathbb{P}(550 < S < 700) &\geq \mathbb{P}(550 < S < 650) = \mathbb{P}(-50 < S - 600 < 50) \\ &= \mathbb{P}(|S - 600| < 50) \\ &= 1 - \mathbb{P}(|S - 600| \geq 50) \\ &\geq 1 - \frac{500}{50^2} = 0,8. \end{aligned}$$

Tableau comparatif des formules pour des v.a. discrètes et continues à densité

Lorsque les intégrales et les séries concernées sont absolument convergentes, on a le tableau comparatif suivant entre le cas général et les déclinaisons discrètes et à densité (cas continu) :

X	Cas général	Variable discrète	Variable à densité f
$X(\Omega)$	quelconque	$\{x_1, x_2, \dots, x_k, \dots\}$	\mathbb{R} ou un borélien
$\mathbb{P}(a \leq X \leq b)$	$\int_{\Omega} \mathbf{1}_{[a,b]}(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{a \leq x_k \leq b} \mathbb{P}(X = x_k)$	$\int_a^b f(t) dt$
$F(x) = \mathbb{P}(X \leq x)$	$\int_{\Omega} X(\omega) d\mathbb{P}(\omega)$	$\sum_{x_k \leq x} \mathbb{P}(X = x_k)$	$\int_{-\infty}^x f(t) dt$
$\mathbb{E}[X]$	$\int \mathbf{1}_{]-\infty, x]}(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{k=1}^{+\infty} x_k \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} t f(t) dt$
$\mathbb{E}[g(X)]$	$\int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega)$	$\sum_{k=1}^{+\infty} g(x_k) \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} g(t) f(t) dt$
$\mathbb{E}[X^2]$	$\int_{\Omega} X(\omega)^2 d\mathbb{P}(\omega)$	$\sum_{k=1}^{+\infty} x_k^2 \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} t^2 f(t) dt$
$\text{Var}(X)$	$\int_{\Omega} (X(\omega) - \mathbb{E}[X])^2 d\mathbb{P}(\omega)$	$\sum_{k=1}^{+\infty} (x_k - \mathbb{E}[X])^2 \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} (t - \mathbb{E}[X])^2 f(t) dt$

Chapitre 2

Vecteurs aléatoires

2.1 Rappels d'intégration

Définition 2.1.1 (Espace produit) Soit (X, \mathcal{A}, μ) et (Y, \mathcal{B}, ν) deux espaces mesurés.

- Sur l'espace produit $X \times Y = \{(x, y), x \in X, y \in Y\}$, on définit la tribu produit $\mathcal{A} \otimes \mathcal{B}$ qui est la tribu de $X \times Y$ engendrée par les produits $A \times B$ pour $A \in \mathcal{A}, B \in \mathcal{B}$. Il s'agit de la tribu la plus naturelle sur $X \times Y$.

- Sur l'espace produit $X \times Y$ muni de la tribu produit $\mathcal{A} \times \mathcal{B}$, on définit la mesure produit $\mu \otimes \nu$ de la façon suivante

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B).$$

Exemple. Sur \mathbb{R}^n , la tribu borélienne de \mathbb{R}^n coïncide avec la tribu produit des tribus boréliennes des espaces facteurs \mathbb{R} :

$$\mathcal{B}(\mathbb{R}^n) = \underbrace{\mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})}_{n \text{ fois}}.$$

Un borélien typique de \mathbb{R}^n est le pavé $[a_1, b_1] \times \cdots \times [a_n, b_n]$.

En générale en dimension n , on utilise la mesure de Lebesgue de dimension n , il s'agit de la mesure produit des mesures de Lebesgue sur \mathbb{R} : $\lambda_n = \underbrace{\lambda \otimes \cdots \otimes \lambda}_{n \text{ fois}}$. Par exemple,

$$\lambda_n([a_1, b_1] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \times \cdots \times (b_n - a_n).$$

2.1.1 Théorèmes de Fubini

Sous de bonnes conditions, le théorème de Fubini permet de permuter les intégrations dans des intégrales multiples. Ainsi les intégrales multiples, ou par rapport à des mesures produits, se ramènent à des intégrales simples emboîtées.

Théorème 2.1.1 (Fubini-Tonelli) Soit $f : (X \times Y, \mathcal{A} \otimes \mathcal{B}) \rightarrow [0, +\infty]$ une fonction mesurable, avec μ et ν des mesures σ -finies sur (X, \mathcal{A}) et (Y, \mathcal{B}) . Alors

1) $x \mapsto \int_Y f(x, \cdot) d\nu$ est \mathcal{A} -mesurable et $y \mapsto \int_X f(\cdot, y) d\mu$ est \mathcal{B} -mesurable.

2) Puis, on a

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y).$$

Démonstration : cf poly de Calcul Intégral.

Le théorème de Fubini-Tonelli ne s'applique qu'à des fonctions mesurables positives (sans conditions supplémentaires). Pour des fonctions quelconques, on a le résultat suivant

Théorème 2.1.2 (Fubini) Soient (X, \mathcal{A}, μ) et (Y, \mathcal{B}, ν) des espaces σ -finis et $f : (X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu) \rightarrow \mathbb{R}$ ou \mathbb{C} une fonction $(\mu \otimes \nu)$ -intégrable, çàd

$$\int_{X \times Y} |f| d(\mu \otimes \nu) < +\infty.$$

Alors

1) Pour μ -presque chaque x , $f(x, \cdot)$ est ν -intégrable et pour ν -presque chaque y , $f(\cdot, y)$ est μ -intégrable.

Posons $I(x) = \int_Y f(x, \cdot) d\nu$ et $J(y) = \int_X f(\cdot, y) d\mu$, alors

2) I et J sont intégrables.

3)

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_X I d\mu = \int_Y J d\nu.$$

On a donc en écrivant les variables d'intégration

$$\int_{X \times Y} f(x, y) d(\mu \otimes \nu) = \int_X \int_Y f(x, y) d\nu(y) d\mu(x) = \int_Y \int_X f(x, y) d\mu(x) d\nu(y).$$

Démonstration : cf poly de Calcul Intégral.

Remarque 2.1.1 • En pratique, on raisonne de la façon suivante :

1) On montre que f est mesurable (arguments généraux),

2) pour montrer que f est intégrable, on calcule $\int |f| d(\mu \otimes \nu)$ en appliquant Fubini-Tonelli à la fonction positive $|f|$:

$$\int |f| d(\mu \otimes \nu) = \int_Y \left(\int_X |f(x, y)| d\mu \right) d\nu = \int_X \left(\int_Y |f(x, y)| d\nu \right) d\mu$$

en choisissant la forme la plus convenable (intégrer d'abord en x ou en y) pour faire le calcul.

3) On applique Fubini.

• Si F est positive, on peut intervertir directement les intégrations (par la version Fubini-Tonelli du résultat) . Si f ne l'est pas, il faut vérifier l'intégrabilité en calculant

l'intégrale de $|f|$ en appliquant par exemple la version Fubini-Tonelli à $|f| > 0$ pour se ramener à des intégrales simples.

- L'utilisation du théorème de Fubini permet de ramener de nombreux calculs d'intégrales doubles (ou triples ou plus généralement multiples) à des calculs successifs d'intégrales simples (aussi bien pour des calculs effectifs que pour montrer des convergences d'intégrales).

2.1.2 Changement de variable

Transfert

Soient (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurables et $\varphi : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ une fonction mesurable.

Si on considère une mesure μ sur (X, \mathcal{A}) alors la mesure image $\nu = \mu_\varphi$ est une mesure sur (Y, \mathcal{B}) . On a un lien entre les intégrales par rapport à μ sur X et celle par rapport à $\nu = \mu\varphi^{-1} = \mu_\varphi$ sur Y :

Théorème 2.1.3 (Transfert) *Soit $h : (Y, \mathcal{B}) \rightarrow K = \mathbb{R}, \bar{\mathbb{R}}, \mathbb{C}$ mesurable alors h est ν -intégrable ssi $h \circ \varphi$ est μ -intégrable et*

$$\int_X h \circ \varphi d\mu = \int_Y h d\nu. \quad (2.1)$$

Il s'agit d'une formule de changement de variable abstraite, très générale puisque la seule condition pour le changement de variable $y = \varphi(x)$ est que φ soit mesurable!

Malheureusement, la nouvelle mesure $\nu = \mu_\varphi$ n'est pas explicite du tout. Ce résultat est donc essentiellement abstrait et difficile à utiliser pour des calculs explicites.

On propose ici des résultats plus explicites, avec des conditions plus restrictives sur le changement de variables.

Rappel : intégrale de Riemann

Soit I un intervalle de \mathbb{R} et $\varphi : I \rightarrow \mathbb{R}$ strictement monotone et C^1 tel que φ' ne s'annule pas sur I . Alors on a

$$\int_I f(x) dx = \int_{\varphi(I)} f(\varphi^{-1}(y)) |(\varphi^{-1})'(y)| dy.$$

Pour cela, on pose $y = \varphi(x)$ ou $x = \varphi^{-1}(y)$ et en dérivant on a la relation entre dx et dy

$$\frac{dx}{dy} = (\varphi^{-1})'(y) \quad \text{c'est à dire} \quad dx = (\varphi^{-1})'(y) dy.$$

Changement de variable

Définition 2.1.2 Soit $F : D \subset \mathbb{R}^n \rightarrow D' \subset \mathbb{R}^n$ où D et D' sont des ouverts. F est appelé un *difféomorphisme* si c'est une bijection de classe C^1 dont la bijection réciproque est aussi de classe C^1 .

Définition 2.1.3 La matrice jacobienne d'un changement de variable

$$y = F(x) \iff (y_1, \dots, y_n) = (F_1(x_1, \dots, x_n), \dots, (F_n(x_1, \dots, x_n)))$$

est

$$J_F(x) = J_F(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1} & \dots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}.$$

Le jacobien est le déterminant de la matrice jacobienne.

La matrice jacobienne est la matrice des dérivées partielles.

Rappel : Calculs des déterminants d'ordre 2 et 3 :

- $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$

- Règle de Sarrus : $\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3 - a_3 b_2 c_1 - b_3 c_2 a_1 - c_3 a_2 b_1.$

- Développements selon une ligne ou une colonne pour se ramener à des déterminants d'ordre inférieur.

Théorème 2.1.4 (Changement de variable) Soit V un ouvert de \mathbb{R}^d et φ un C^1 -difféomorphisme de V dans $\varphi(V) \subset \mathbb{R}^d$. Alors, on a les formules de changements de variables

$$\begin{aligned} \int_V f(x) d\lambda(x) &= \int_{\varphi(V)} f(\varphi^{-1}(y)) |J_{\varphi^{-1}}(y)| d\lambda(y) \\ \int_V h(\varphi(x)) d\lambda(x) &= \int_{\varphi(V)} h(y) |J_{\varphi^{-1}}(y)| d\lambda(y) \end{aligned}$$

pour toutes fonctions f et h mesurables telle que f est λ -intégrable et $h \circ \varphi$ est λ -intégrable.

Démonstration : Admis.

Coordonnées polaires et sphériques

- Un changement de variables utile dans le plan \mathbb{R}^2 est le changement de variables en polaire qui consiste à passer de (x, y) représentant des coordonnées cartésiennes dans un repère orthonormé à (r, θ) les coordonnées polaires correspondantes données par

$$(r, \theta) = \varphi(x, y) \iff \varphi^{-1} : \begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}, \quad r \in [0, +\infty[, \theta \in [0, 2\pi[.$$

On remplace alors $dx dy$ par $r dr d\theta$ car le jacobien du changement de variable est r :

$$J_{\varphi^{-1}}(r, \theta) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta + r \sin^2 \theta = r.$$

Ainsi :

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) dx dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy \\ &= \int \int_{[0, +\infty[\times [0, 2\pi[} f(r \cos \theta, r \sin \theta) r dr d\theta. \end{aligned}$$

Exemples :

- Normalisation de la loi normale $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.

Notons $I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx$ et montrons que $I^2 = 2\pi$. On a

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-x^2/2} dx \times \int_{-\infty}^{+\infty} e^{-y^2/2} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{-y^2/2} dx dy = \int \int_{\mathbb{R} \times \mathbb{R}} e^{-(x^2+y^2)/2} dx dy \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{+\infty} r e^{-r^2/2} dr = 2\pi \left[-e^{-r^2/2} \right]_0^{+\infty} = 2\pi \end{aligned}$$

où on a utilisé le théorème de Fubini à la 2ème ligne puis on a fait un changement de variables en polaires à la 3ème ligne.

- Aire d'un disque : $\Delta = \{(x, y) : x^2 + y^2 \leq R^2\}$:

$$\lambda_2(\Delta) = \iint_{B(0, R)} dx dy = \iint_{[0, R] \times [0, 2\pi[} r dr d\theta = \int_0^{2\pi} d\theta \int_0^R r dr = 2\pi \left[\frac{r^2}{2} \right]_0^R = \pi R^2.$$

• En dimension 3, le changement de variables utile est le changement en coordonnées sphériques donné par

$$(r, \theta, \varphi) = \phi(x, y, z) \quad \iff \quad \phi^{-1} : \begin{cases} x &= r \cos \theta \cos \varphi \\ y &= r \cos \theta \sin \varphi \\ z &= r \sin \theta \end{cases}$$

où $\theta \in]-\pi/2, \pi/2[$ est la latitude, $\varphi \in [0, 2\pi[$ est la longitude et $r \in [0, +\infty[$ la distance à l'origine.

Le jacobien du changement de variable est

$$J_{\phi^{-1}}(r, \theta, \varphi) = \begin{bmatrix} \cos \theta \cos \varphi & -r \sin \theta \cos \varphi & -r \cos \theta \sin \varphi \\ \cos \theta \sin \varphi & -r \sin \theta \sin \varphi & r \sin \theta \cos \varphi \\ \sin \theta & r \cos \theta & 0 \end{bmatrix} = r^2 \cos \theta.$$

Ainsi :

$$\begin{aligned} & \int \int \int_{\mathbb{R}^3} f(x, y, z) dx dy dz \\ &= \int \int \int_{[0, +\infty[\times [0, 2\pi[\times]-\frac{\pi}{2}, \frac{\pi}{2}[} f(r \cos \theta \cos \varphi, r \cos \theta \sin \varphi, r \sin \theta) r^2 \cos \theta dr d\theta d\varphi. \end{aligned}$$

Ce type de changement de variable (polaire en dimension 2, sphérique en dimension 3) se généralise en dimension n avec

$$\begin{cases} x_1 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-2} \cos \theta_{n-1}, \\ x_2 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-2} \sin \theta_{n-1}, \\ x_3 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-3} \sin \theta_{n-2}, \\ x_4 &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{n-4} \sin \theta_{n-3}, \\ \dots &= \dots \\ x_{n-1} &= r \cos \theta_1 \sin \theta_2, \\ x_n &= r \sin \theta_1. \end{cases}$$

Exemple : Calcul du volume d'une boule euclidienne de rayon R est

$$\begin{aligned} \lambda_3(B(0, R)) &= \iiint_{B(0, R)} dx dy dz = \iiint_{[0, R] \times]-\pi/2, \pi/2[\times [0, 2\pi[} r^2 \cos \theta dr d\theta d\varphi \\ &= \int_0^{2\pi} d\varphi \int_{-\pi/2}^{\pi/2} \cos \theta d\theta \int_0^R r^2 dr = 2\pi [\sin \theta]_{-\pi/2}^{\pi/2} \left[\frac{r^3}{3} \right]_0^R \\ &= \frac{4}{3} \pi R^3 \end{aligned}$$

où λ_3 désigne la mesure de Lebesgue en dimension 3.

2.2 Vecteurs aléatoires

Définition 2.2.1 On appelle vecteur aléatoire toute application de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ mesurable.

Définition 2.2.2 Si on note $p_i : \mathbb{R}^n \rightarrow \mathbb{R}$ la i -ème projection qui à $x = (x_1, \dots, x_n)$ associe $p_i(x) = x_i$, on appelle i -ème marginale du vecteur X la variable aléatoire $X_i = p_i(X)$

$$X = (X_1, \dots, X_n), \quad X_i \text{ est la } i\text{-ème marginale.}$$

Définition 2.2.3 La loi du vecteur aléatoire X est la mesure image sur \mathbb{R}^n de la probabilité par X :

$$\mathbb{P}_X(A_1 \times \cdots \times A_n) = \mathbb{P}(X \in A_1 \times \cdots \times A_n) = \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n), A_i \in \mathcal{B}(\mathbb{R}), 1 \leq i \leq n.$$

C'est une mesure de probabilité dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Définition 2.2.4 • Un vecteur aléatoire X est discret si l'ensemble de ses valeurs $X(\Omega)$ est discret dans \mathbb{R}^n .

• Un vecteur aléatoire X de \mathbb{R}^n est de loi à densité si sa loi est absolument continue par rapport à la mesure de Lebesgue dans \mathbb{R}^n : $\mathbb{P}_X \ll \lambda^n$. Autrement dit, sa loi s'exprime comme une intégrale (multiple) :

$$d\mathbb{P}_X(x) = f(x_1, \dots, x_n) dx_1 \dots dx_n \iff \mathbb{P}_X(A) = \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n, A \in \mathcal{B}(\mathbb{R}^n).$$

On vérifie que comme $\mathbb{P}_X(\mathbb{R}^n) = \mathbb{P}(X \in \mathbb{R}^n) = 1$, une densité en dimension n satisfait $f(x_1, \dots, x_n) \geq 0$ et

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1.$$

Proposition 2.2.1 Si (X, Y) est un couple de loi $\mathbb{P}_{X,Y} = \mu$ alors les lois marginales \mathbb{P}_X et \mathbb{P}_Y de X et Y s'obtiennent par

$$\mathbb{P}_X(A) = \mu(A \times \mathbb{R}), \quad \mathbb{P}_Y(B) = \mu(\mathbb{R} \times B), \quad A, B \in \mathcal{B}(\mathbb{R}).$$

Démonstration : C'est évident si on remarque que $\{X \in A\} = \{(X, Y) \in A \times \mathbb{R}\}$. \square

Autrement dit les lois marginales s'obtiennent par intégration partielle, qu'elle soit discrète ou continues comme on le voit avec les cas particuliers suivants :

Proposition 2.2.2 (Cas discret) Si (X, Y) est un couple aléatoire de v.a. discrètes de domaine $(X, Y)(\Omega) = \{(x_1, y_1), \dots, (x_i, y_i), \dots\}$, les domaines des marginales X, Y s'obtiennent par projection :

$$X(\Omega) = p_1((X, Y)(\Omega)) = \{x_i, i \in I\}, \quad Y(\Omega) = p_2((X, Y)(\Omega)) = \{y_i, i \in I\}$$

où p_1, p_2 sont les première et seconde projections

$$p_1 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto x \end{cases}, \quad p_2 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto y \end{cases}.$$

Les lois marginales $\mathbb{P}_X, \mathbb{P}_Y$ (i.e. les lois de X et Y , ses marginales) sont données par :

$$\begin{aligned} \forall x_i \in X(\Omega), \quad \mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) &= \sum_{y_j \in Y(\Omega)} \mathbb{P}(X = x_i, Y = y_j), \\ \forall y_i \in Y(\Omega), \quad \mathbb{P}_Y(y_i) = \mathbb{P}(Y = y_i) &= \sum_{x_i \in X(\Omega)} \mathbb{P}(X = x_i, Y = y_i). \end{aligned}$$

Ici, l'intégration partielle est une somme.

Démonstration : Il suffit de faire la preuve pour le domaine et les probabilités de X . Or pour i fixé $\{X = x_i\}$ est la réunion de la famille dénombrable d'évènements deux à deux disjoints $\{X = x_i, Y = y_j\}$ pour tous les j tels que $y_j \in Y(\Omega)$ car $\{\omega; Y(\omega) = y_j\}_j$ est une partition de Ω . On conclut alors par σ -additivité de \mathbb{P} :

$$\begin{aligned}\mathbb{P}(X = x_i) &= \mathbb{P}\left(\{X = x_i\} \cap \bigcup_j \{Y = y_j\}\right) \\ &= \mathbb{P}\left(\bigcup_j \{X = x_i, Y = y_j\}\right) = \sum_{y_j \in Y(\Omega)} \mathbb{P}(X = x_i, Y = y_j).\end{aligned}$$

Puis $\{x_1, \dots, x_i, \dots\}$ et $\{y_1, \dots, y_j, \dots\}$ sont bien d'une part les projections de $(X, Y)(\Omega)$ sur les premier et second facteurs de $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ et d'autre part les domaines de X et Y . \square

Proposition 2.2.3 (Cas à densité) Si (X, Y) est un couple aléatoire de loi de densité f , ses lois marginales $\mathbb{P}_X, \mathbb{P}_Y$ sont à densité de densités données par :

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy, \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) dx.\end{aligned}$$

Les densités f_X de X et f_Y de Y sont appelées les densités marginales.

Démonstration : La preuve est une application directe du théorème de Fubini-Tonelli sur les intégrales doubles une fois qu'on a remarqué que

$$\begin{aligned}\mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) = \mathbb{P}_{(X, Y)}(A \times \mathbb{R}) \\ &= \int_{A \times \mathbb{R}} f(x, y) dx dy = \int_A \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_A f_X(x) dx\end{aligned}$$

avec la densité annoncée $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$. Il s'applique sans problème car par définition d'une densité, f est positive (et même intégrable sur \mathbb{R}^2). Idem pour Y . \square

De même si $X = (X_1, \dots, X_n)$ est de densité f , la i -ème marginale X_i est de densité

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

Remarque 2.2.1 Attention, si la connaissance de la loi du couple ou d'un vecteur permet d'en déduire celle des lois marginales, la réciproque est en général fautive. Voir l'exemple suivant.

Exemples : Pour des variables discrètes.

On donne le tableau de $\mathbb{P}(X = x_i, Y = y_j)$:

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0, 1	0, 05	0, 15	0	0, 3
$x_2 = 2$	0, 05	0, 2	0, 05	0, 1	0, 4
$x_3 = 3$	0, 1	0	0, 1	0, 1	0, 3
	0, 25	0, 25	0, 3	0, 2	1

On en déduit la loi de X : $X(\Omega) = \{0, 2, 3\}$ et

$$\mathbb{P}(X = 0) = 0, 3, \quad \mathbb{P}(X = 2) = 0, 4, \quad \mathbb{P}(X = 3) = 0, 3$$

et celle de Y : $Y(\Omega) = \{-1, 2, 3, 5\}$ et

$$\mathbb{P}(Y = -1) = 0, 25, \quad \mathbb{P}(Y = 2) = 0, 25, \quad \mathbb{P}(Y = 3) = 0, 3, \quad \mathbb{P}(Y = 5) = 0, 2.$$

Notons qu'il n'y a pas unicité des couples (X, Y) donnant les mêmes marginales. Ainsi, le couple suivant différent du précédent partage les mêmes marginales.

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0, 1	0, 1	0	0, 1	0, 3
$x_2 = 2$	0, 1	0, 1	0, 1	0, 1	0, 4
$x_3 = 3$	0, 05	0, 05	0, 2	0	0, 3
	0, 25	0, 25	0, 3	0, 2	1

Exemples : Pour des variables à densité.

• Considérons $f(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$. Il s'agit bien d'une densité car f est positive et

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) \, dx dy &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) \, dx dy \\ &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) \, dx dy \\ &= \frac{1}{3} \underbrace{\int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \, dx}_{=1} \underbrace{\int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) \, dy}_{=2 - (-1) = 3} \\ &= 1. \end{aligned}$$

Considérons un couple (X, Y) de loi de densité f . La loi de X est alors de densité donnée par :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) dy$$

$$\begin{aligned}
&= \mathbf{1}_{[0,1]}(x) \times \underbrace{\frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) dy}_{=1} \\
&= \mathbf{1}_{[0,1]}(x).
\end{aligned}$$

De la même façon, $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$.

• Soit $f(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$. Il s'agit d'une densité car

$$\begin{aligned}
\int \int_{\mathbb{R}^2} f(x, y) dx dy &= \int \int_{\mathbb{R}^2} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx dy}{3\pi} \\
&= \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx dy}{3\pi} = \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2}{2 \times 3}} e^{-\frac{4y^2}{2 \times 3}} \frac{dx dy}{3\pi} \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} dx \right) e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-\frac{z^2}{2 \times 3}} dz \right) e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} \\
&= \int_{\mathbb{R}} \sqrt{2\pi} \times 3 e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{y^2}{2 \times (3/4)}} \frac{dy}{\sqrt{2\pi} \times 3/4} = 1
\end{aligned}$$

en utilisant que $\int_{\mathbb{R}} e^{-\frac{t^2}{2\sigma^2}} dt = \sqrt{2\pi\sigma^2}$ d'après la normalisation de la loi normale $\mathcal{N}(0, \sigma^2)$.

Considérons un couple (X, Y) de densité f , alors X est de densité

$$\begin{aligned}
f_X(x) &= \int_{\mathbb{R}} f(x, y) dy = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x + \sqrt{5}y)^2 + 4x^2/5}{6}} \frac{dy}{3\pi} \\
&= \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x + \sqrt{5}y)^2}{6}} e^{-\frac{4x^2}{30}} \frac{dy}{3\pi} = e^{-\frac{4x^2}{30}} \int_{\mathbb{R}} e^{-\frac{z^2}{2 \times 3}} \frac{dz}{3\pi\sqrt{5}} = e^{-\frac{4x^2}{30}} \frac{\sqrt{2\pi} \times 3}{3\pi\sqrt{5}} \\
&= \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}.
\end{aligned}$$

et Y de densité

$$\begin{aligned}
f_Y(y) &= \int_{\mathbb{R}} f(x, y) dx = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} e^{-\frac{4y^2}{6}} \frac{dx}{3\pi} \\
&= e^{-\frac{4y^2}{6}} \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} \frac{dx}{3\pi} = e^{-\frac{4y^2}{6}} \frac{\sqrt{2\pi} \times 3}{3\pi} = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}.
\end{aligned}$$

Les marginales X et Y sont donc de lois $\mathcal{N}(0; 15/4)$ et $\mathcal{N}(0; 3/4)$.

Généralisation au cas gaussien.

Soit $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ une forme quadratique positive. Notons

$$C = \int_{\mathbb{R}^n} e^{-\Phi(x_1, \dots, x_n)} dx_1 \dots dx_n.$$

Alors tout vecteur aléatoire de densité

$$f(x_1, \dots, x_n) = \frac{1}{C} e^{-\Phi(x_1, \dots, x_n)}$$

est dit gaussien. Ses marginales sont des v.a. gaussiennes en dimension 1. On traitera ce type de vecteur en détail plus tard.

2.3 Indépendance de variables aléatoires

2.3.1 Définition

Cette notion a déjà été abordée en L2. Il s'agit d'une notion fondamentale en probabilité. Rappelons d'abord que

Définition 2.3.1 (Indépendance d'évènements) • Deux évènements $A, B \in \mathcal{F}$ d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sont indépendants si

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

On note $A \perp B$.

• n évènements observables A_1, \dots, A_n sont mutuellement indépendants si pour toute sous-famille A_{i_1}, \dots, A_{i_p} avec $1 \leq i_1 < \dots < i_p \leq n$, on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_p}).$$

• Une suite infinie d'évènements est dite (mutuellement) indépendante si toute sous-famille finie est formée d'évènements mutuellement indépendants (au sens précédent).

En particulier, A et B incompatibles ne peuvent pas être indépendants à moins que l'un des deux ne soit de probabilité nulle. Sinon $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$, tandis que $\mathbb{P}(A)\mathbb{P}(B) > 0$.

Il ne faut donc pas confondre les deux notions.

La notion d'indépendance se généralise aux tribus de la façon suivante :

Définition 2.3.2 (Indépendance de tribus)

• Deux tribus \mathcal{F} et \mathcal{G} sur un même espace Ω sont indépendantes si pour tout $A \in \mathcal{F}$ et $B \in \mathcal{G}$ on a

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

On note toujours $\mathcal{F} \perp \mathcal{G}$.

• Une famille de tribus $(\mathcal{F}_i)_{i \in I}$ est dite indépendante si pour tous $A_i \in \mathcal{F}_i$, la famille d'évènements $(A_i)_{i \in I}$ est (mutuellement) indépendante.

Enfin en associant à chaque v.a. X une tribu $\sigma(X)$, on définit l'indépendance pour des va.

Définition 2.3.3 (Tribu engendrée par une va) Soit X une v.a. sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ alors la tribu engendrée par X est celle engendrée par les ensembles $X^{-1}(A)$ pour tout $A \in \mathcal{B}(\mathbb{R})$. On la note $\sigma(X)$.

Heuristiquement, la tribu $\sigma(X)$ est la tribu qui contient toutes les informations liées à la v.a. X .

Définition 2.3.4 (Indépendance de va)

• Deux v.a. X et Y sont indépendantes ssi leur tribu engendrée sont indépendantes. On note toujours $\sigma(X) \perp \sigma(Y)$.

• Toute suite $(X_i)_{i \in I}$ de v.a. est indépendante ssi leur tribu associée sont indépendantes.

2.3.2 Critères et exemples

Par définition des tribus engendrées par les $v.a.$, il est immédiat de voir que des critères plus concrets d'indépendance sont :

Proposition 2.3.1 (Indépendance de $v.a.$)

• **Indépendance de deux $v.a.$** Deux $v.a.$ X, Y sont dites indépendantes si pour $A, B \in \mathcal{B}(\mathbb{R})$, mesurables de \mathbb{R} , les évènements $\{X \in A\}, \{Y \in B\}$ sont indépendants :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

• **Indépendance d'une famille finie de $v.a.$** Les m variables aléatoires X_1, \dots, X_m sont dites (mutuellement) indépendantes si pour tout boréliens A_1, \dots, A_m , les évènements $\{X_1 \in A_1\}, \dots, \{X_m \in A_m\}$ sont mutuellement indépendants :

$$\mathbb{P}(X_1 \in A_1, \dots, X_m \in A_m) = \mathbb{P}(X_1 \in A_1) \dots \mathbb{P}(X_m \in A_m).$$

• **Indépendance d'une suite de $v.a.$** Une suite $(X_i)_{i \in \mathbb{N}}$ de $v.a.$ est dite indépendante si toute sous-suite finie de $(X_i)_{i \in \mathbb{N}}$, la propriété précédente est vraie.

On a d'autres critères pour l'indépendance des $v.a.$ Ils portent sur la structure de la loi du vecteur associé, considérée comme mesure dans l'espace produit \mathbb{R}^n (cf. la définition 2.1.1).

Proposition 2.3.2 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est à composantes indépendantes ssi sa loi \mathbb{P}_X est une loi produit (de ses lois marginales) :

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}. \quad (2.2)$$

Démonstration : Soit $B = B_1 \times \dots \times B_n$ pavé de $\mathcal{B}(\mathbb{R}^n)$, alors

$$\begin{aligned} \mathbb{P}_{(X_1, \dots, X_n)}(B) &= \mathbb{P}((X_1, \dots, X_n) \in B) = \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \dots \times B_n) \\ &= \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n) \\ &= \mathbb{P}_{X_1}(B_1) \times \dots \times \mathbb{P}_{X_n}(B_n) = (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(B_1 \times \dots \times B_n) \\ &= (\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n})(B). \end{aligned}$$

Comme les pavés $B = B_1 \times \dots \times B_n$ engendrent $\mathcal{B}(\mathbb{R}^n)$ et $\mathbb{P}_{(X_1, \dots, X_n)}$ et $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ coïncident sur les pavés alors ces mesures coïncident sur $\mathcal{B}(\mathbb{R}^n)$.

Pour la réciproque, prendre B un pavé et remonter les étapes. \square

Corollaire 2.3.1 Les $v.a.$ X_1, \dots, X_n sont indépendantes ssi pour tout x_1, \dots, x_n , on a

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \times \dots \times \mathbb{P}(X_n \leq x_n).$$

Démonstration : Appliquer le résultat précédent avec la famille d'ensembles $B =]-\infty, x_1] \times \dots \times]-\infty, x_n]$. C'est suffisant car on montre que cette famille d'ensembles engendre la tribu $\mathcal{B}(\mathbb{R}^n)$. \square

Remarque 2.3.1 Dans les cas discret et à densité, on peut préciser les critères d'indépendances :

— Les v.a. discrètes X et Y sont indépendantes si et seulement si

$$\forall x_i \in X(\Omega), \forall y_j \in Y(\Omega), \quad \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j). \quad (2.3)$$

— Les v.a. X, Y de densités respectives f et g sont indépendantes si et seulement si le couple (X, Y) est de densité

$$f \otimes g : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto f(x)g(y). \quad (2.4)$$

Remarque 2.3.2 Une conséquence importante : si on connaît les lois de X et Y , des variables supposées **indépendantes**, on peut reconstruire la loi du couple (X, Y) à partir des marginales par (2.2) (dans le cas discret par (2.3) et dans le cas à densité par (2.4)). Insistons sur le fait que ce n'est pas vrai en général quand X et Y ne sont pas indépendantes.

Dans les deux exemples de la page 28, X et Y ne sont pas indépendantes car par exemple pour le premier :

$$\mathbb{P}(X = 2, Y = 2) = 0,2, \quad \text{tandis que} \quad \mathbb{P}(X = 2) \times \mathbb{P}(Y = 2) = 0,4 \times 0,25 = 0,1.$$

et pour le second :

$$\mathbb{P}(X = 3, Y = 5) = 0, \quad \text{tandis que} \quad \mathbb{P}(X = 3) \times \mathbb{P}(Y = 5) = 0,3 \times 0,2 = 0,06.$$

Exemples : • On donne le tableau de la loi d'un couple (X, Y) en donnant les probabilités ponctuelles $\mathbb{P}(X = x_i, Y = y_j)$:

$X \setminus Y$	y_1	y_2	y_3	
x_1	0,12	0,08	0,20	0,4
x_2	0,18	0,12	0,30	0,6
	0,3	0,2	0,5	= 1

On vérifie ici que X et Y sont indépendantes car pour tout $i = 1, 2$ et $j = 1, 2, 3$, on a

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j).$$

• Considérons le couple (X, Y) de loi donnée par la densité $f_{(X,Y)}(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$. On a vu que X et Y avaient pour densité $f_X(x) = \mathbf{1}_{[0,1]}(x)$ et $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$. On a alors

$$f_{(X,Y)}(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) = \mathbf{1}_{[0,1]}(x) \times \frac{1}{3} \mathbf{1}_{[-1,2]}(y) = f_X(x) f_Y(y).$$

Les variables X et Y sont donc indépendantes.

• Soit (X, Y) le couple aléatoire de loi donnée par la densité $f_{(X,Y)}(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$. On a vu que les densités marginales sont

$$f_X(x) = \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}, \quad f_Y(y) = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}.$$

On a alors

$$f_X(x)f_Y(y) = \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}} \times \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}} \neq \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}} = f_{(X,Y)}(x, y).$$

Dans ce cas, X et Y ne sont pas indépendantes.

Proposition 2.3.3 Soient X_i $1 \leq i \leq n$, des v.a. indépendantes et $h_i : \mathbb{R} \rightarrow \mathbb{C}$ des fonctions boréliennes telles que $h_i(X_i)$ soit \mathbb{P} -intégrable. Alors $\prod_{i=1}^n h_i(X_i)$ est \mathbb{P} -intégrable et

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

Démonstration : On utilise le théorème de transfert, l'indépendance et Fubini :

$$\begin{aligned} E\left[\prod_{i=1}^n h_i(X_i)\right] &= \int \prod_{i=1}^n h_i(x_i) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int \prod_{i=1}^n h_i(x_i) d\mathbb{P}_{X_1}(x_1) \dots d\mathbb{P}_{X_n}(x_n) \\ &= \int h(x_1) d\mathbb{P}_{X_1}(x_1) \times \dots \times \int h(x_n) d\mathbb{P}_{X_n}(x_n) \\ &= \prod_{i=1}^n \mathbb{E}[h_i(X_i)]. \end{aligned}$$

□

Corollaire 2.3.2 Soient X_1, \dots, X_n des v.a. réelles (ou complexes) indépendantes avec des moments d'ordre 1

$$\mathbb{E}[X_1 \dots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n].$$

Démonstration : Appliquer le résultat précédent avec la fonction $h_i(x) = x$. □

La réciproque est fautive : soit X_1 de loi uniforme sur $[-1, 1]$ et $X_2 = X_1^2$, on a $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1^3] = \int_{-1}^+ x_1^3 dx_1 = 0$ et $\mathbb{E}[X_1] = 0$ si bien qu'on a $\mathbb{E}[X_1 X_2] = 0 = \mathbb{E}[X_1] \mathbb{E}[X_2]$ mais X_1 et X_2 ne sont pas indépendantes car par exemple

$$\mathbb{P}(X_1 \in [0, 1/2], X_2 \in [0, 1/4]) = 1/4, \quad \mathbb{P}(X_1 \in [0, 1/2]) \times \mathbb{P}(X_2 \in [0, 1/4]) = 1/8.$$

2.3.3 Covariance et indépendance

Proposition 2.3.4 *Soient X et Y deux vecteurs aléatoires indépendants de variances finies. Alors $\text{Cov}(X, Y) = 0$.*

Démonstration :

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y\mathbb{E}[X]] - \mathbb{E}[X\mathbb{E}[Y]] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 \end{aligned}$$

car par indépendance $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. □

La réciproque est fautive : si X et Y sont de covariance nulle alors ils sont nécessairement indépendants. Cependant dans le cas de variables X, Y gaussiennes, on verra que la réciproque est vraie.

Pour la somme d'une variance, on déduit

Corollaire 2.3.3 *Si X et Y sont deux v.a. indépendantes avec des moments d'ordre deux alors*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Démonstration : C'est immédiat puisque

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

et $\text{Cov}(X, Y) = 0$ quand $X \perp Y$. □

Chapitre 3

Somme de deux variables aléatoires indépendantes et convolution

3.1 Convolution de mesures

On considère un espace vectoriel mesurable (X, \mathcal{A}) . Il y a donc sur X une structure d'espace vectoriel et une tribu. L'exemple typique est \mathbb{R} ou \mathbb{R}^n .

Définition 3.1.1 Soient μ et ν deux mesures quelconques sur un espace vectoriel mesurable (X, \mathcal{A}) . La convolée $\mu * \nu$ de ces deux mesures est

$$\mu * \nu(A) = \int_X \mu(A - x) d\nu(x) \quad (3.1)$$

où $A - x = \{a - x : a \in A\}$.

On vérifie facilement que $\mu * \nu$ est une mesure (la σ -additivité vient de celle de μ et du théorème de convergence monotone).

Proposition 3.1.1 La convolution est commutative : $\mu * \nu = \nu * \mu$.

Démonstration : En effet par le théorème de Fubini, on a

$$\mu * \nu(A) = \int \int \mathbf{1}_A(x + y) d\nu(x) d\mu(y).$$

On a obtenu une expression symétrique en μ et ν , ce qui justifie la proposition. \square

Proposition 3.1.2 δ_0 est l'élément neutre de la convolution des mesures : $\mu * \delta_0 = \mu$.

Démonstration : En effet

$$\mu * \delta_0(A) = \int_X \delta_0(A - x) d\mu(x) = \int_A 1 d\mu(x) = \mu(A)$$

car $\delta_0(A - x) = 1$ ssi $0 \in A - x$ c'est à dire ssi $x \in A$, cela vaut 0 sinon.

Proposition 3.1.3 *Si μ et ν sont des mesures de probabilité alors $\mu * \nu$ l'est aussi.*

Démonstration :

$$\mu * \nu(X) = \int_X \mu(X - x) d\nu(x) = \int_X d\nu(x) = \nu(X) = 1$$

car $X - x = X$ et donc $\mu(X - x) = \mu(X) = 1$. La mesure $\mu * \nu$ est donc de poids total égale à 1 : c'est une mesure de probabilité. \square

Désormais, nous considérons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel sont définies des variables aléatoires X . Leur lois \mathbb{P}_X sont des mesures sur l'ev \mathbb{R} . On va les convoler.

Proposition 3.1.4 *Soient X et Y deux variables aléatoires **indépendantes** de loi \mathbb{P}_X et \mathbb{P}_Y . Alors la loi de $X + Y$ est $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.*

Démonstration : Soit $A \in \mathcal{B}(\mathbb{R})$, on a

$$\begin{aligned} \mathbb{P}_{X+Y}(A) &= \mathbb{P}(X + Y \in A) = \mathbb{E}[\mathbf{1}_A(X + Y)] = \int \int \mathbf{1}_A(x + y) d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int \int \mathbf{1}_A(x + y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) = \int \left(\int \mathbf{1}_{A-y}(x) d\mathbb{P}_X(x) \right) d\mathbb{P}_Y(y) \\ &= \int \mathbb{P}_X(A - y) d\mathbb{P}_Y(y) = \mathbb{P}_X * \mathbb{P}_Y(A). \end{aligned}$$

\square

Par une récurrence immédiate, on connaît donc la loi d'une somme de n variables indépendantes $X_1 + \dots + X_n$ lorsqu'on a celles des termes X_i de la somme. Lorsqu'il n'y a pas indépendance, ce n'est pas suffisant : il faut connaître l'imbrication des lois c'est à dire la loi du vecteur associé (X_1, \dots, X_n) .

Ainsi de façon générale, la loi d'une somme $X_1 + \dots + X_n$ de variables aléatoires non indépendantes est donnée par

$$\mathbb{P}_{X_1 + \dots + X_n}(A) = \mathbb{P}(X_1 + \dots + X_n \in A) = \int \int \mathbf{1}_A(x_1 + \dots + x_n) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n).$$

Étant donné deux mesures μ et ν sur un espace (X, \mathcal{A}) , il ne faut pas confondre la mesure produit $\mu \otimes \nu$ et la mesure produit de convolution $\mu * \nu$.

- La mesure produit $\mu \otimes \nu$ est la mesure sur l'espace produit $X \times X$ définie par

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), \quad \forall A \times B \in \mathcal{A} \otimes \mathcal{A}.$$

Dans le contexte probabiliste, si μ et ν sont les lois de probabilité de variables aléatoires X, Y indépendantes alors $\mu \otimes \nu$ est la loi du vecteur (X, Y) .

• La mesure produit de convolution $\mu * \nu$ est la mesure sur l'espace X définie par (3.1). Dans le contexte probabiliste, si μ et ν sont les lois de probabilités de variables aléatoires X, Y indépendantes alors $\mu * \nu$ est la loi de la variable $X + Y$.

On considère dans la suite les deux cas particuliers des variables aléatoires discrètes et à densité.

Dans le cas à densité, la convolution des lois à densité devient la convolution des densités.

On décrit pour commencer la convolution de deux fonctions.

3.2 Convolution de deux fonctions

On considère dans ce chapitre $(X, \mathcal{A}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda^n)$. Les fonctions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ sont à voir comme des densités de vecteurs aléatoires de \mathbb{R}^n . Pour simplifier les notations, on considère dans la suite $n = 1$.

Définition 3.2.1 *La convolution de deux fonctions f et g réelles est la fonction $f * g$ réelle donnée par*

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y) dy.$$

On parle encore de la convolée $f * g$ de f et de g .

À ce stade, rien ne garantit que la fonction $f * g$ soit bien définie. Il n'est pas clair en effet pour quels x l'intégrale qui la définit converge. Elle est convergente, par exemple, si les fonctions f et g sont à supports compacts.

Proposition 3.2.1 *Soient f, g des fonctions mesurables, alors*

- $(f * g)(x) = (g * f)(x)$
- $(f, g) \mapsto f * g$ est bilinéaire.

Par contre, il n'y a pas de fonction e élément neutre pour ce produit ($f * e = e * f = f$).

C'est un problème en analyse que de construire des « approximations de l'unité », c'est-à-dire des fonctions e_k qui asymptotiquement se comportent comme une unité pour la convolution des fonctions.

Démonstration : • Pour le premier point, le changement de variable $z = x - y$ donne

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y)dy = \int_{\mathbb{R}} f(z)g(x - z)|-1|dz = (g * f)(x).$$

• Pour le second point, par linéarité de l'intégrale, il est facile de voir que

$$(a_1f_1 + a_2f_2) * g(x) = a_1(f_1 * g)(x) + a_2(f_2 * g)(x).$$

La linéarité par rapport à g se montre de la même façon ou en utilisant celle par rapport à f avec la symétrie de $*$ vue au premier point. \square

Le résultat suivant est une version de l'inégalité de Hölder pour la convolution.

Proposition 3.2.2 Soient $f, g \in L^1(\mathbb{R})$, alors $f * g$ est bien définie pp et

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

Démonstration : En appliquant le théorème de Fubini-Tonelli, on a

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}^n} |f(y)| |g(x-y)| dy dx &= \int_{\mathbb{R}} \int_{\mathbb{R}^n} |f(y)| |g(x-y)| dx dy \\ &= \int_{\mathbb{R}} |f(y)| \left(\int_{\mathbb{R}^n} |g(x-y)| dx \right) dy \\ &= \int_{\mathbb{R}} |f(y)| \left(\int_{\mathbb{R}^n} |g(z)| dz \right) dy \\ &= \int_{\mathbb{R}} |f(y)| dy \int_{\mathbb{R}^n} |g(z)| dz \\ &= \|f\|_1 \|g\|_1. \end{aligned}$$

Par conséquent, $I(x) = \int_{\mathbb{R}} |f(y)| |g(x-y)| dy$ est d'intégrale finie bornée par $\|f\|_1 \|g\|_1$, c'est donc que $I(x) = +\infty$ sur un ensemble de mesure nulle et $f * g$ est bien définie pp. Puis

$$\begin{aligned} \|f * g\|_1 &= \int_{\mathbb{R}} |(f * g)(x)| dx = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(y) g(x-y) dy \right| dx \\ &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |f(y) g(x-y)| dy dx \\ &\leq \|f\|_1 \|g\|_1. \end{aligned}$$

□

Ainsi $*$ est une loi de composition interne de $L^1(\mathbb{R})$. On peut vérifier qu'elle est associative, $(L^1(\mathbb{R}^n), +, \cdot, *)$ est alors une algèbre, comme elle est complète pour $\|\cdot\|_1$, il s'agit même d'une algèbre de Banach.

Le résultat précédent est un cas particulier du résultat suivant plus général (admis, cf. poly de Calcul Intégral) :

Théorème 3.2.1 Soit $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ avec $r \neq +\infty$ alors pour $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$, $f * g$ existe pp et $f * g \in L^r(\mathbb{R})$, avec

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

Si $q = 1$, on constate que la convolution par une fonction $g \in L^1(\mathbb{R})$ est un opérateur de $L^p(\mathbb{R})$ dans $L^p(\mathbb{R})$:

$$f \in L^p(\mathbb{R}) \mapsto f * g \in L^p(\mathbb{R}).$$

Le résultat suivant explique le lien entre la convolution de mesures avec des densités f et g et la convolution des fonctions f, g .

Proposition 3.2.3 Soient μ et ν des mesures sur \mathbb{R} de densités respectives f et g ($d\mu = f d\lambda$, $d\nu = g d\lambda$). La convolée $\mu * \nu$ est de densité $f * g$.

Démonstration : On a

$$\begin{aligned} \mu * \nu(A) &= \int \int \mathbf{1}_A(x+y) d\mu(x) d\nu(y) \\ &= \int \int_{x+y \in A} f(x)g(y) dx dy \\ &= \int \int_{\mathbb{R} \times A} f(t)g(s-t) dt ds \\ &= \int_A \left(\int_{\mathbb{R}} f(t)g(s-t) ds \right) dt \\ &= \int_A (f * g)(t) dt. \end{aligned}$$

On a fait le changement de variable $(x, y) \longrightarrow (t, s) = (x, x+y)$. Comme (x, y) varie dans \mathbb{R}^2 de façon que $x+y \in A$, t décrit tout \mathbb{R} et s décrit A . Noter que le jacobien du changement de variable est

$$Jac = \begin{vmatrix} \frac{\partial t}{\partial x} & \frac{\partial s}{\partial x} \\ \frac{\partial t}{\partial y} & \frac{\partial s}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

□

3.3 Loi d'une somme de variables aléatoires à densité indépendantes

La convolution de deux densités permet de calculer la densité de la somme de deux variables aléatoires indépendantes qui ont des densités :

Proposition 3.3.1 Soient X, Y deux v.a. indépendantes et de densités f et g . La loi de $X+Y$ est donnée par :

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}_{X+Y}(A) = \mathbb{P}(X+Y \in A) = \int_A (f * g)(x) dx.$$

Autrement dit, $X+Y$ a pour densité la fonction $f * g$.

Démonstration : C'est une conséquence immédiate des Propositions 3.1.4 et 3.2.3.

Remarque 3.3.1 On connaît bien la loi de la somme $X + Y$ si X et Y sont indépendantes, sinon, il faut connaître la loi du couple (X, Y) et par exemple sa densité $h(x, y)$ si elle existe pour avoir la loi de $X + Y$ par

$$\mathbb{P}(X + Y \in A) = \int_{(x,y); x+y \in A} h(x, y) dx dy = \int_A \int_{-\infty}^{+\infty} h(x, y - x) dx dy.$$

Exemples : • Soient X, Y des v.a. indépendantes suivant des lois exponentielles de paramètres respectifs a et b . Quelle est la loi de $S = X + Y$?

Les lois de X et de Y sont de densités

$$f(x) = ae^{-ax} \mathbf{1}_{[0, +\infty[}(x), \quad g(y) = be^{-by} \mathbf{1}_{[0, +\infty[}(y).$$

Comme X et Y sont indépendantes, la densité de $X + Y$ est, si $a \neq b$:

$$\begin{aligned} (f * g)(x) &= \int_{-\infty}^{+\infty} g(y) f(x - y) dy \\ &= \int_{-\infty}^{+\infty} be^{-by} \mathbf{1}_{[0, +\infty[}(y) ae^{-a(x-y)} \mathbf{1}_{[0, +\infty[}(x - y) dy \\ &= abe^{-ax} \mathbf{1}_{[0, +\infty[}(x) \int_0^x e^{(a-b)y} dy \\ &= \frac{abe^{-ax}}{a - b} (e^{(a-b)x} - 1) \mathbf{1}_{[0, +\infty[}(x) \\ &= \frac{ab}{a - b} (e^{-bx} - e^{-ax}) \mathbf{1}_{[0, +\infty[}(x). \end{aligned}$$

Noter que dans ce calcul, on a utilisé l'égalité entre produit d'indicatrices

$$\mathbf{1}_{[0, +\infty[}(x - y) \mathbf{1}_{[0, +\infty[}(y) = \mathbf{1}_{[0, x]}(y) \mathbf{1}_{[0, +\infty[}(x).$$

Si $a = b$, la densité est

$$\begin{aligned} (f * g)(x) &= \int_{-\infty}^{+\infty} g(y) f(x - y) dy = a^2 \int_{-\infty}^{+\infty} e^{-ay} \mathbf{1}_{\{y \geq 0\}} e^{-a(x-y)} \mathbf{1}_{\{x-y \geq 0\}} dy \\ &= a^2 \mathbf{1}_{\{x \geq 0\}} \int_0^x e^{-ay} e^{-a(x-y)} dy = a^2 \mathbf{1}_{\{x \geq 0\}} e^{-ax} \int_0^x dy = a^2 x \mathbf{1}_{\{x \geq 0\}} e^{-ax}. \end{aligned}$$

•

Proposition 3.3.2 Soient X_1 de loi $\mathcal{N}(m_1, \sigma_1^2)$ et X_2 de loi $\mathcal{N}(m_2, \sigma_2^2)$ alors $X_1 + X_2$ est de loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Comme $X_1 - m_1 \simeq \mathcal{N}(0, \sigma_1^2)$ et $X_2 - m_2 \simeq \mathcal{N}(0, \sigma_2^2)$, il suffit, pour simplifier, de voir le cas où $m_1 = m_2 = 0$. Le point clef dans ce calcul est donné par la normalisation de la loi normale $\mathcal{N}(0, \sigma^2)$:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}.$$

Notons alors f_1 et f_2 les densités de X_1 et de X_2 . Celle de $X_1 + X_2$ est donnée par

$$\begin{aligned}
f_1 * f_2(x) &= \int_{-\infty}^{+\infty} f_1(t) f_2(x-t) dt = \int_{-\infty}^{+\infty} e^{-t^2/(2\sigma_1^2)} e^{-(x-t)^2/(2\sigma_2^2)} \frac{dt}{\sqrt{2\pi\sigma_1^2} \sqrt{2\pi\sigma_2^2}} \\
&= \int_{-\infty}^{+\infty} \exp - \left(\frac{(\sigma_1^2 + \sigma_2^2)t^2 - 2\sigma_1^2 xt + \sigma_1^2 x^2}{2\sigma_1^2 \sigma_2^2} \right) \frac{dt}{2\pi\sigma_1\sigma_2} \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp - \left(\frac{((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x)^2 - \frac{\sigma_1^4}{(\sigma_1^2 + \sigma_2^2)}x^2 + \sigma_1^2 x^2}{2\sigma_1^2 \sigma_2^2} \right) dt \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp - \left(\frac{((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x)^2 + \frac{\sigma_1^2 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}x^2}{2\sigma_1^2 \sigma_2^2} \right) dt \\
&= \frac{\exp - \frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp - \left(\frac{((\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x)^2}{2\sigma_1^2 \sigma_2^2} \right) dt \\
&= \frac{\exp - \frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp - \left(\frac{u^2}{2\sigma_1^2 \sigma_2^2} \right) \frac{du}{(\sigma_1^2 + \sigma_2^2)^{1/2}}
\end{aligned}$$

avec le changement de variable $u = (\sigma_1^2 + \sigma_2^2)^{1/2}t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}}x$. Puis d'après la normalisation de la loi normale $\mathcal{N}(0, \sigma_1^2 \sigma_2^2)$, la dernière intégrale vaut $\frac{\sqrt{2\pi\sigma_1^2 \sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}}$, on a finalement :

$$f_1 * f_2(x) = \frac{\exp - \frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} \frac{\sqrt{2\pi\sigma_1^2 \sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}} = \frac{\exp - \frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}.$$

On a obtenu la densité de la loi $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

3.4 Cas de variables aléatoires discrètes indépendantes

Si X et Y sont des variables aléatoires discrètes indépendantes, alors $X+Y$ est encore une variable discrète de domaine

$$(X + Y)(\Omega) = X(\Omega) + Y(\Omega) = \{x_i + y_j : x_i \in X(\Omega), y_j \in Y(\Omega)\}$$

et on peut calculer ses probabilités ponctuelles par une « convolution discrète ».

Supposons pour simplifier que $X(\Omega) = Y(\Omega) = \mathbb{N}$, et notons $(p_k)_{k \in \mathbb{N}}$, $(q_k)_{k \in \mathbb{N}}$ les probabilités ponctuelles de X et de Y alors $(X+Y)(\Omega) = \mathbb{N}$ et comme on a pour chaque $k \in \mathbb{N}$ la partition suivante $\{X + Y = n\} = \bigcup_{k=0}^n \{X = k, Y = n - k\}$, il vient

$$\mathbb{P}(X + Y = n) = \mathbb{P} \left(\bigcup_{k=0}^n \{X = k, Y = n - k\} \right)$$

$$\begin{aligned}
&= \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) \\
&= \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) \\
&= \sum_{k=0}^n p_k q_{n-k}.
\end{aligned}$$

Les probabilités ponctuelles de $X + Y$ sont donc données par la suite $\left(\sum_{k=0}^n p_k q_{n-k} \right)_{n \in \mathbb{N}}$.

Exemple : Soient X, Y des v.a. indépendantes suivant des lois de Poisson de paramètres respectifs α et β . Quelle est la loi de $S = X + Y$?

Les lois de X et de Y sont données par

$$\mathbb{P}(X = i) = \frac{e^{-\alpha} \alpha^i}{i!}, \quad \mathbb{P}(Y = j) = \frac{e^{-\beta} \beta^j}{j!}, \quad i, j \in \mathbb{N}.$$

Comme X et Y sont indépendantes, on a en utilisant la formule du binôme de Newton :

$$\begin{aligned}
\mathbb{P}(S = n) &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i) = \sum_{i=0}^n \frac{e^{-\alpha} \alpha^i}{i!} \frac{e^{-\beta} \beta^{n-i}}{(n-i)!} \\
&= \frac{e^{-(\alpha+\beta)}}{n!} \sum_{i=0}^n C_n^i \alpha^i \beta^{n-i} \\
&= \frac{e^{-(\alpha+\beta)} (\alpha + \beta)^n}{n!}.
\end{aligned}$$

Ainsi $S = X + Y$ suit la loi de Poisson de paramètre $\alpha + \beta$.

Chapitre 4

Loi du 0/1 de Kolmogorov

On rappelle qu'une tribu (ou σ -algèbre) sur X est une famille \mathcal{F} d'ensemble qui contient l'espace X , stable par complémentaire (si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$) et stable par réunion dénombrable (si $A_i \in \mathcal{F}$ pour $i \in \mathbb{N}$, alors $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$).

Si on remplace la stabilité par réunion dénombrable par la stabilité par réunion finie, alors on définit une algèbre.

On rappelle que $\sigma(\mathcal{M})$ est la tribu engendrée par \mathcal{M} . Il s'agit de la plus petite tribu contenant \mathcal{M} et elle s'écrit $\sigma(\mathcal{M}) = \bigcap_{\mathcal{F} \text{ tribu } \supset \mathcal{M}} \mathcal{F}$.

4.1 π -système et d-système

4.1.1 En théorie de la mesure

Définition 4.1.1 (π -système) Une famille d'ensemble $\mathcal{M} \subset \mathcal{P}(X)$ est un π -système si elle est stable par intersection finie :

$$\forall A_1, A_2 \in \mathcal{M}, \quad A_1 \cap A_2 \in \mathcal{M}.$$

Définition 4.1.2 (d-système) Une famille d'ensemble $\mathcal{M} \subset \mathcal{P}(X)$ est un d-système si

- $X \in \mathcal{M}$.
- elle est stable par différence : pour $A_1, A_2 \in \mathcal{M}$ et $A_1 \subset A_2$ alors $A_2 \setminus A_1 \in \mathcal{M}$.
- elle est stable par réunion dénombrable croissante : pour $A_i \in \mathcal{M}$ avec $A_i \subset A_{i+1}$ alors $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{M}$.

On montre que si on a deux mesures finies μ_1 et μ_2 sur (X, \mathcal{A}) alors la famille des ensembles où elles coïncident $\mathcal{D} = \{A \in \mathcal{A}, \mu_1(A) = \mu_2(A)\}$ est un d-système (cf. ci-dessous, le théorème de Dynkin)

Définition 4.1.3 Étant donnée une famille d'ensemble \mathcal{M} , on note respectivement

$$d(\mathcal{M}) = \bigcap_{\mathcal{D} \text{ d-système } \supset \mathcal{M}} \mathcal{D}$$

$$\pi(\mathcal{M}) = \bigcap_{\mathcal{D} \text{ } \pi\text{-système} \supset \mathcal{M}} \mathcal{D}$$

les plus petits d -système et π -système contenant \mathcal{M} .

Proposition 4.1.1 Soit $\mathcal{M} \subset \mathcal{P}(X)$ stable par intersection finie. Alors $d(\mathcal{M}) = \sigma(\mathcal{M})$.

Démonstration : Comme une tribu est un d -système, on a $d(\mathcal{M}) \subset \sigma(\mathcal{M})$. Il reste à voir que $\sigma(\mathcal{M})$ est un d -système pour avoir l'égalité.

Soit $C \in \mathcal{M}$ alors $\mathcal{D}_1 = \{B \in d(\mathcal{M}), B \cap C \in d(\mathcal{M})\}$ est un d -système.

En effet,

- $X \in \mathcal{D}_1$ car $X \cap C = C \in \mathcal{M} \subset d(\mathcal{M})$.
- Si $A, B \in \mathcal{D}_1$ avec $B \subset A$, alors $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C) \in d(\mathcal{M})$ car $A \cap C, B \cap C \in d(\mathcal{M})$ et $d(\mathcal{M})$ est stable par différence.
- Si $A_i \in \mathcal{D}_1$ avec $A_i \subset A_{i+1}$ alors $(\cup_i A_i) \cap C = \cup_i (A_i \cap C) \in d(\mathcal{M})$ car $A_i \cap C \in d(\mathcal{M})$ et $d(\mathcal{M})$ est stable par réunion croissante.

Puis $\mathcal{M} \subset \mathcal{D}_1$ car si $B \in \mathcal{M}$ alors comme \mathcal{M} est stable par intersection finie, $B \cap C \in \mathcal{M} \subset d(\mathcal{M})$.

On a donc $d(\mathcal{M}) \subset \mathcal{D}_1$, et donc $d(\mathcal{M}) = \mathcal{D}_1$. C'est à dire : pour tout $B \in d(\mathcal{M})$ et $C \in \mathcal{M}$, on a $B \cap C \in d(\mathcal{M})$.

Soit $D \in d(\mathcal{M})$. Alors $\mathcal{D}_2 = \{B \in d(\mathcal{M}), B \cap D \in d(\mathcal{M})\}$ est un d -système.

Cela se justifie comme pour \mathcal{D}_1 .

Puis d'après la première étape : pour $B \in \mathcal{M}, B \cap D \in d(\mathcal{M})$, donc $B \in \mathcal{D}_2$ et en fait $\mathcal{M} \subset \mathcal{D}_2$. Comme \mathcal{D}_2 est un d -système, on a aussi $d(\mathcal{M}) = \mathcal{D}_2$.

On a montré que pour tout $B, D \in d(\mathcal{M})$, on a $B \cap D \in d(\mathcal{M})$, càd $d(\mathcal{M})$ est stable par intersection finie.

On a donc :

- $X \in d(\mathcal{M})$.
- $d(\mathcal{M})$ est stable par complémentaire : car si $A \in d(\mathcal{M})$ alors $A^c = X \setminus A \in d(\mathcal{M})$ par stabilité par différence de $d(\mathcal{M})$.
- $d(\mathcal{M})$ est stable par intersections finies donc aussi par réunions finies d'après le point précédent. Puis comme $d(\mathcal{M})$ est stable par réunions croissantes, $d(\mathcal{M})$ est finalement stable par réunions dénombrables.

Finalement $d(\mathcal{M})$ est une tribu or $\sigma(\mathcal{M}) \subset d(\mathcal{M})$. D'où l'égalité cherchée. \square

Théorème 4.1.1 (Dynkin) Soient P, Q deux mesures de probabilité sur $\sigma(C)$ la tribu engendrée par un π -système C . On suppose que pour tout $A \in C$, on a $P(A) = Q(A)$. Alors $P = Q$.

Démonstration : Soit $L = \{A \in \sigma(C) \mid P(A) = Q(A)\}$. On montre facilement que L est un d -système qui contient (par hypothèse) C . Comme C est stable par intersection, on a d'après la Prop. 4.1.1 $\sigma(C) = d(C)$. Mais alors $\sigma(C) = d(C) \subset d(L) = L$, càd

$\sigma(C) \subset L$. On a même $\sigma(C) = L$. \square

Exemple. La fonction de répartition $F(x) = \mathbb{P}(X \leq x)$ suffit à déterminer la mesure \mathbb{P}_X d'une v.a. X .

Proposition 4.1.2 *Soient \mathcal{D} un d-système et Π un π -système. Alors si $\Pi \subset \mathcal{D}$, on a $\sigma(\Pi) \subset \mathcal{D}$.*

Démonstration : De $\Pi \subset \mathcal{D}$, on déduit $d(\Pi) \subset d(\mathcal{D}) = \mathcal{D}$.

Mais Π est stable par intersection finie donc d'après la Prop. 4.1.1, on a $d(\Pi) = \sigma(\Pi)$, ce qui conclut. \square

Définition 4.1.4 (Classe monotone) *Une famille d'ensemble $C \subset \mathcal{P}(\Omega)$ est une classe monotone si elle est stable par réunion dénombrable croissante et par intersection dénombrable décroissante.*

Théorème 4.1.2 (classe monotone) *Soient \mathcal{A} une algèbre et \mathcal{M} une classe monotone avec $\mathcal{A} \subset \mathcal{M}$. Alors $\sigma(\mathcal{A}) \subset \mathcal{M}$.*

Démonstration : Notons $m(\mathcal{A})$ la plus petite classe monotone contenant \mathcal{A} . La classe monotone $m(\mathcal{A})$ est stable par complémentaire car \mathcal{M}' la famille d'ensembles de $m(\mathcal{A})$ de complémentaire toujours dans $m(\mathcal{A})$ est une classe monotone, ce qui garantit $m(\mathcal{A}) \subset \mathcal{M}' \subset m(\mathcal{A})$.

Comme \mathcal{A} est une algèbre, on a aussi $\Omega \in \mathcal{A} \subset m(\mathcal{A})$ et la classe monotone $m(\mathcal{A})$ est finalement un d-système. Comme une algèbre est un cas particulier de π -système, d'après la Prop. 4.1.2, on a $\sigma(\mathcal{A}) \subset m(\mathcal{A}) \subset \mathcal{M}$. \square

4.1.2 En probabilité

Proposition 4.1.3 *Soient $(C_i)_{i \in I}$ des π -systèmes. Alors les $(C_i)_{i \in I}$ sont indépendants ssi $(\sigma(C_i))_{i \in I}$ le sont.*

Il suffit donc de vérifier l'indépendance des tribus sur des π -systèmes qui les engendrent.

Démonstration : Le sens indirect est immédiat.

Pour le sens direct, soit $J = \{1, \dots, n\} \subset I$. Pour $A_i \in \sigma(C_i)$, $i = 1, \dots, n$, notons

$$L_1 = \{B \in \sigma(C_1) \mid \forall A_2, \dots, A_n \in C_2 \times \dots \times C_n, \mathbb{P}(B \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(B)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)\}.$$

On a $C_1 \subset L_1 \subset \sigma(C_1)$. Puis L_1 est un d-système car

— $\Omega \in L_1$, c'est clair.

— Si $B_1, B_2 \in L_1$ avec $B_2 \subset B_1$ alors

$$\begin{aligned}
 \mathbb{P}((B_1 \setminus B_2) \cap A_2 \cap \dots \cap A_n) &= \mathbb{P}((B_1 \cap A_2 \cap \dots \cap A_n) \setminus (B_2 \cap A_2 \cap \dots \cap A_n)) \\
 &= \mathbb{P}((B_1 \cap A_2 \cap \dots \cap A_n)) - \mathbb{P}((B_2 \cap A_2 \cap \dots \cap A_n)) \\
 &= \mathbb{P}(B_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n) - \mathbb{P}(B_2)\mathbb{P}(A_2) \dots \mathbb{P}(A_n) \\
 &= (\mathbb{P}(B_1) - \mathbb{P}(B_2))\mathbb{P}(A_2) \dots \mathbb{P}(A_n) \\
 &= \mathbb{P}(B_1 \setminus B_2)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)
 \end{aligned}$$

on a donc $B_1 \setminus B_2 \in L_1$.

— Si $B_i \in L_1$ avec $B_i \subset B_{i+1}$, on a aussi $\cup_i B_i \in L_1$, en effet

$$\begin{aligned}
 \mathbb{P}((\cup_i B_i) \cap A_2 \cap \dots \cap A_n) &= \mathbb{P}(\cup_i (B_i \cap A_2 \cap \dots \cap A_n)) \\
 &= \lim_i \mathbb{P}(B_i \cap A_2 \cap \dots \cap A_n) \text{ (monotonie séquentielle des proba)} \\
 &= \lim_i \mathbb{P}(B_i)\mathbb{P}(A_2) \dots \mathbb{P}(A_n) \\
 &= \mathbb{P}(\cup_i B_i)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).
 \end{aligned}$$

L_1 est donc un d système, stable par intersection finie. D'après la proposition 4.1.1, on a $L_1 = d(L_1) = \sigma(L_1)$. On a donc pour tout $B_1 \in \sigma(C_1)$, pour tout $A_2 \in C_2 \dots, A_n \in C_n$,

$$\mathbb{P}(B \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(B)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

On définit maintenant

$$\begin{aligned}
 L_2 &= \{B_2 \in \sigma(C_2) \mid \forall B_1 \in \sigma(C_1), A_3, \dots, A_n \in C_3 \times \dots \times C_n, \\
 &\quad \mathbb{P}(B_1 \cap B_2 \cap A_3 \dots \cap A_n) = \mathbb{P}(B_1)\mathbb{P}(B_2)\mathbb{P}(A_3) \dots \mathbb{P}(A_n)\}.
 \end{aligned}$$

Comme pour L_1 , on montre que $L_2 = \sigma(C_2)$, puis en raisonnant par récurrence, on montre que

$$\forall B_1, \dots, B_n \in \sigma(C_1) \times \dots \times \sigma(C_n), \mathbb{P}(B_1 \cap B_2 \dots \cap A_n) = \mathbb{P}(B_1) \dots \mathbb{P}(A_n).$$

□

Application : Deux v.a. X et Y sont indépendantes ssi

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

En effet $\{\{X \leq x\}, x \in \mathbb{R}\}$ est un π -système qui engendre la tribu $\sigma(X)$. De même pour $\sigma(Y)$.

Théorème 4.1.3 (Coalitions) Soient $(\mathcal{F}_i)_{i \in I}$ des sous tribus de \mathcal{F} . On suppose que les \mathcal{F}_i sont indépendantes. Soit $I = \bigcup_{a \in A} I_a$ une partition de l'ensemble d'indexation I ($I_a \cap I_b = \emptyset$) et $\mathcal{G}_a = \sigma(\bigcup_{i \in I_a} \mathcal{F}_i)$. Alors les tribus $(\mathcal{G}_a)_{a \in A}$ sont indépendantes.

Heuristiquement, le théorème des coalitions explique qu'en réunissant des tribus indépendantes, on construit de nouvelles tribus indépendantes.

Démonstration : Soit $C_a = \{B \in \mathcal{F} \text{ tel que } \exists J \text{ finie } \subset I_a, \exists A_i \in \mathcal{F}, i \in J, B = \bigcap_{i \in J} A_i\}$.

Alors $\bigcup_{i \in I_a} \mathcal{F}_i \subset C_a \subset \mathcal{G}_a$. Les $(C_a)_{a \in A}$ sont clairement indépendants (car l'intersection est finie) et ce sont des π -systèmes. C'est donc encore vraie pour les tribus engendrées $\sigma(C_a) = \mathcal{G}_a$. \square

4.2 Tribus du futur et tribu asymptotique

Définition 4.2.1 Soit $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une suite de tribus de \mathcal{F} .

- On appelle tribu du futur : $\mathcal{F}^n = \sigma\left(\bigcup_{k \geq n} \mathcal{F}_k\right)$, $n \in \mathbb{N}$.
- On appelle tribu asymptotique : $\mathcal{F}^\infty = \bigcap_{n \in \mathbb{N}} \mathcal{F}^n$.

D'habitude on a $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ et la famille de tribus croissantes $(\mathcal{F}_n)_n$ s'appelle une filtration. Dans ce cas, heuristiquement, si \mathbb{N} désigne le temps et $n \in \mathbb{N}$ la date n , la tribu \mathcal{F}_n représente la tribu des événements du passé par rapport au temps n (elle comporte toutes les informations sur ce qu'il s'est passé jusqu'au temps n), la tribu \mathcal{F}^n représente la tribu des événements du futur par rapport au temps n (elle comporte toutes les informations sur ce qu'il se passera après le temps n). La tribu \mathcal{F}^∞ représente celle des événements asymptotiques.

Exemples. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. indépendantes. On considère la suite de tribus $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Il s'agit de la filtration habituelle associée à une suite de va.

• Notons A l'évènement « X_n repasse par 0 une infinité de fois ». Alors A est dans la tribu asymptotique des \mathcal{F}_n . En effet,

$$A = \bigcap_n \bigcup_{k \geq n} \{X_k = 0\}.$$

et $\bigcup_{k \geq n} \{X_k = 0\} \in \mathcal{F}^n$.

• $B = \{\text{les } X_n \text{ sont bornés}\}$ n'est pas dans \mathcal{F}^∞ car par exemple X_1 borné n'est pas dans \mathcal{F}^2 ni dans aucun \mathcal{F}^n ($n \geq 2$).

• $C = \{\text{la suite } (X_n)_n \text{ converge}\} \in \mathcal{F}^\infty$. Heuristiquement, la convergence de la suite X_n est un résultat asymptotique (il ne concerne que les X_n pour n arbitrairement grand) donc il est dans la tribu asymptotique. Plus précisément, comme dans \mathbb{R} la convergence d'une suite est équivalente à dire que la suite est de Cauchy, $\omega \in C$ s'écrit :

$$\forall r \in \mathbb{N}, \exists N \in \mathbb{N}, \forall p, q \geq N, |X_p(\omega) - X_q(\omega)| < 1/r.$$

On a donc l'écriture suivante de C :

$$C = \bigcap_{r \in \mathbb{N}} \bigcup_{N \in \mathbb{N}} \bigcap_{p, q \geq N} \{|X_p - X_q| < 1/r\}$$

avec $\{|X_p - X_q| < 1/r\} \in \mathcal{F}^p$ (si $p < q$).

Le résultat clef sur les tribus asymptotiques est

Théorème 4.2.1 (Loi du 0/1 de Kolmogorov) *Soit $(\mathcal{F}_n)_n$ une suite de tribus indépendantes alors la tribu asymptotique est triviale : c'est à dire*

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

Démonstration : On montre que \mathcal{F}^∞ est indépendante de \mathcal{F}^∞ . Ainsi si $A \in \mathcal{F}^\infty$, on a $A = A \cap A$ et par indépendance de A avec A : $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, d'où $\mathbb{P}(A) = 0$ ou 1.

Par le théorème des coalitions, $\mathcal{F}^n \perp \mathcal{F}_k$ pour $k < n$ si bien que $\mathcal{F}^n \perp \sigma(\bigcup_{k < n} \mathcal{F}_k)$ et comme $\mathcal{F}^\infty \subset \mathcal{F}^n$, on a $\mathcal{F}^\infty \perp \sigma(\bigcup_{k < n} \mathcal{F}_k)$.

Soit $\mathcal{C} = \bigcup_n \sigma(\bigcup_{k < n} \mathcal{F}_k)$, c'est un π -système car \mathcal{C} est stable par intersection finie :

En effet, soient $A, B \in \mathcal{C}$, on a $A \in \sigma(\bigcup_{k < n} \mathcal{F}_k)$ et $B \in \sigma(\bigcup_{k < n'} \mathcal{F}_k)$. En notant $N = \max(n, n')$, on a $A, B \in \sigma(\bigcup_{k < N} \mathcal{F}_k)$. Par stabilité d'une tribu par \cap , on a aussi $A \cap B \in \sigma(\bigcup_{k < N} \mathcal{F}_k)$ et donc $A \cap B \in \mathcal{C}$.

Comme \mathcal{C} est un π -système, de $\mathcal{C} \perp \mathcal{F}^\infty$ on déduit $\sigma(\mathcal{C}) \perp \mathcal{F}^\infty$ par la Prop. 4.1.3.

Pour tout k , on a $\mathcal{F}_k \subset \mathcal{C}$ donc aussi $\bigcup_k \mathcal{F}_k \subset \mathcal{C}$. Ainsi la tribu $\sigma(\mathcal{C})$ contient $\sigma(\bigcup_k \mathcal{F}_k) = \mathcal{F}^1$ donc aussi \mathcal{F}^∞ .

On déduit alors de $\sigma(\mathcal{C}) \perp \mathcal{F}^\infty$ que $\mathcal{F}^\infty \perp \mathcal{F}^\infty$, ce qui permet de conclure que pour tout $A \in \mathcal{F}^\infty$, on a $A \perp A$. \square

Exemples. Dans les exemples précédents, on déduit de la loi de Kolmogorov que $\mathbb{P}(A) = 0$ ou 1, de même pour $\mathbb{P}(C)$. Par contre, on ne peut rien dire de $\mathbb{P}(B)$.

4.3 Liminf et limsup d'ensembles

Définition 4.3.1 *Soit $(A_n)_n$ une suite d'évènements observables d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On pose*

$$\overline{\lim}_{n \rightarrow +\infty} A_n = \bigcap_{n \geq 1} \bigcup_{k > n} A_k,$$

$$\underline{\lim}_{n \rightarrow +\infty} A_n = \bigcup_{n \geq 1} \bigcap_{k > n} A_k.$$

On parle respectivement de limites supérieure et inférieure de la suite d'ensembles $(A_n)_n$.

Notez que

$$\underline{\lim}_{n \rightarrow +\infty} A_n \subset \overline{\lim}_{n \rightarrow +\infty} A_n. \quad (4.1)$$

En effet $\bigcap_{k > n} A_k \subset A_q$ pour tout $q > n$. On a donc $\bigcap_{k > n} A_k \subset \bigcup_{q > p} A_q$ pour tout p .

On a alors pour tout n :

$$\bigcap_{k > n} A_k \subset \bigcap_{p \geq 1} \bigcup_{q > p} A_q = \overline{\lim}_{n \rightarrow +\infty} A_n.$$

Finalement

$$\bigcup_{n \geq 1} \bigcap_{k > n} A_k \subset \overline{\lim_{n \rightarrow +\infty} A_n}.$$

C'est à dire (4.1).

Puis, les règles élémentaires sur les \bigcup , \bigcap et c donnent :

$$(\overline{\lim_{n \rightarrow +\infty} A_n})^c = \underline{\lim_{n \rightarrow +\infty} A_n^c}.$$

De plus, les limites supérieure et inférieure d'ensembles ont les interprétations suivantes :

Proposition 4.3.1 Soit $A_i, i \geq 0$, une collection infinie d'ensembles. Alors

— « À partir d'un certain rang, ω est dans tous les A_i » s'écrit

$$\omega \in \bigcup_{i \geq 0} \bigcap_{j > i} A_j \quad (= \underline{\lim}_i A_i).$$

— « ω est dans une infinité de A_i » s'écrit

$$\omega \in \bigcap_{i \geq 0} \bigcup_{j > i} A_j \quad (= \overline{\lim}_i A_i).$$

Démonstration :

• Pour le premier point : Soit ω qui, à partir d'un certain rang, est dans tous les A_i . On traduit cela de la façon suivante : il existe un rang i tel que pour tout rang $j > i$, ω est dans A_j . D'après la signification des symboles $\forall, \exists, \cap, \cup$, cela revient à écrire

$$\omega \in \underbrace{\bigcup_{i \geq 0}}_{\text{il existe } i \geq 0} \underbrace{\bigcap_{j > i}}_{\text{pour tout } j > i} \underbrace{A_j}_{\omega \text{ est dans } A_j}.$$

• Pour le second point, dire que ω est dans une infinité de A_i est équivalent à dire que

« pour tout p , il existe $q > p$ avec ω dans A_q . »

En effet, si tel est le cas, ω est bien dans une infinité de A_i car, d'après cette propriété,

- avec $p = 0$, il existe $p_1 > p$ tel que ω est dans A_{p_1}
- avec $p = p_1$, il existe $p_2 > p_1$ tel que ω est dans A_{p_2}
- avec $p = p_2$, il existe $p_3 > p_2$ tel que ω est dans A_{p_3}
- ...
- avec $p = p_n$, il existe $p_{n+1} > p_n$ tel que ω est dans $A_{p_{n+1}}$
- ...

et finalement, ω est dans chaque A_{p_n} , $n \in \mathbb{N}^*$, c'est à dire dans une infinité de A_i . Réciproquement, s'il est dans une infinité de A_i , alors pour tout p , on trouve $q > p$ tel que $\omega \in A_q$, sinon, ce serait qu'il existe p tel que pour $q > p$, ω n'est pas dans A_q . Ou encore : ω ne peut appartenir qu'aux A_i d'indice $i \leq p$, c'est à dire seulement à un nombre fini d'entre eux, ce qui est faux.

Donc, pour ce deuxième point, pour tout p , on trouve $q > p$, tel que $\omega \in A_q$, en langage \forall, \exists , cela s'écrit

$$\omega \in \underbrace{\bigcap_{p \geq 0}}_{\text{pour tout } p \geq 0} \bigcup_{\substack{q > p \\ \text{il existe } q > p}} \underbrace{A_q}_{\substack{\omega \text{ est} \\ \text{dans } A_q}}.$$

□

Lien entre liminf, limsup d'ensembles et de fonctions

Rappel. Pour une suite réelle $u = (u_n)_n$, on définit ses limites supérieures et inférieures

$$\begin{aligned} \underline{\lim}_{n \rightarrow +\infty} u_n &= \sup_n \inf_{k \geq n} u_k, \\ \overline{\lim}_{n \rightarrow +\infty} u_n &= \inf_n \sup_{k \geq n} u_k. \end{aligned}$$

Ce sont les plus petites et plus grandes valeur d'adhérence de la suite u . On a toujours

$$\underline{\lim}_{n \rightarrow +\infty} u_n \leq \overline{\lim}_{n \rightarrow +\infty} u_n$$

et il y a égalité ssi la suite u converge ; de plus si tel est le cas

$$\lim_{n \rightarrow +\infty} u_n = \underline{\lim}_{n \rightarrow +\infty} u_n = \overline{\lim}_{n \rightarrow +\infty} u_n.$$

En plus, en changeant le signe, les limites inférieure et supérieure s'échangent :

$$\begin{aligned} \underline{\lim}_{n \rightarrow +\infty} (-u_n) &= -\overline{\lim}_{n \rightarrow +\infty} u_n, \\ \overline{\lim}_{n \rightarrow +\infty} (-u_n) &= -\underline{\lim}_{n \rightarrow +\infty} u_n. \end{aligned}$$

Pour une suite de fonctions $(f_n)_n$, on définit des fonctions limites inférieure et supérieure de la façon suivante :

$$(\underline{\lim}_{n \rightarrow +\infty} f_n)(x) = \underline{\lim}_{n \rightarrow +\infty} f_n(x), \quad (\overline{\lim}_{n \rightarrow +\infty} f_n)(x) = \overline{\lim}_{n \rightarrow +\infty} f_n(x).$$

Proposition 4.3.2 Soit $(A_n)_n$ une suite d'ensembles mesurables, on a

$$\underline{\lim} \mathbf{1}_{A_n} = \mathbf{1}_{\{\underline{\lim} A_n\}}, \quad \overline{\lim} \mathbf{1}_{A_n} = \mathbf{1}_{\{\overline{\lim} A_n\}}.$$

Démonstration : Il suffit de le faire pour $\overline{\lim} \mathbf{1}_{A_n}$. Or $\omega \in \overline{\lim} A_n$ si $\forall k, \exists n \geq k$ tel que $\omega \in A_n$. On a donc $\forall k, \exists n \geq k$ avec $\mathbf{1}_{A_n}(\omega) = 1$ ou encore $\inf_k \sup_{n \geq k} \mathbf{1}_{A_n}(\omega) = 1$, c'est à dire $\overline{\lim} \mathbf{1}_{A_n}(\omega) = 1$. \square

De plus, on a

Proposition 4.3.3

$$\begin{aligned} \underline{\lim}\{f_n \leq t\} &= \{\overline{\lim} f_n \leq t\}, & \overline{\lim}\{f_n \leq t\} &= \{\underline{\lim} f_n \leq t\}, \\ \underline{\lim}\{f_n \geq t\} &= \{\underline{\lim} f_n \geq t\}, & \overline{\lim}\{f_n \geq t\} &= \{\overline{\lim} f_n \geq t\}. \end{aligned}$$

Démonstration : exo

On montre aussi :

Proposition 4.3.4

$$\mathbb{P}(\underline{\lim}_n A_n) \leq \underline{\lim}_n \mathbb{P}(A_n) \leq \overline{\lim}_n \mathbb{P}(A_n) \leq \mathbb{P}(\overline{\lim}_n A_n).$$

On en déduit que si $A_n \rightarrow A$ alors $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$.

Démonstration : Soient $B_n = \bigcup_{k \geq n} A_k$ et $C_n = \bigcap_{k \geq n} A_k$. Alors B_n est une suite décroissante d'ensembles, de limite $\overline{\lim}_n A_n$ et C_n est une suite croissante d'ensembles, de limite $\underline{\lim}_n A_n$. D'après la propriété de convergence monotone des mesures (de probabilités), on a

$$\begin{aligned} \mathbb{P}(A_n) &\leq \mathbb{P}(B_n) \longrightarrow \mathbb{P}(\overline{\lim}_n A_n) \\ \mathbb{P}(A_n) &\geq \mathbb{P}(C_n) \longrightarrow \mathbb{P}(\underline{\lim}_n A_n) \end{aligned}$$

On en déduit $\overline{\lim}_n \mathbb{P}(A_n) \leq \mathbb{P}(\overline{\lim}_n A_n)$ et $\mathbb{P}(\underline{\lim}_n A_n) \leq \underline{\lim}_n \mathbb{P}(A_n)$.

Le second point est clair car la limite supérieure et inférieure coïncident lorsque $A_n \rightarrow A$. \square

4.4 Lemmes de Borel-Cantelli

Les lemmes de Borel-Cantelli complètent d'une certaine façon la loi du 0/1 de Kolmogorov en donnant des cas où la probabilité vaut 1 d'autres où elle vaut 0.

Théorème 4.4.1 (Premier lemme de Borel-Cantelli) Soit $(A_n)_n$ une suite (infinie) d'évènements. On suppose que $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty$. Alors

$$\mathbb{P}(\overline{\lim}_n A_n) = 0.$$

C'est à dire presque sûrement, seuls un nombre fini d'évènements A_n sont réalisés.

Démonstration : Par définition, $\overline{\lim}_n A_n \subset \bigcup_{j \geq i} A_j$ pour tout i . Or

$$\mathbb{P}(\overline{\lim}_n A_n) \leq \mathbb{P}\left(\bigcup_{j \geq i} A_j\right) \leq \sum_{j \geq i} \mathbb{P}(A_j). \quad (4.2)$$

Or $\sum_{j \geq i} \mathbb{P}(A_j)$ est le reste d'ordre i de la série convergente $\sum_n \mathbb{P}(A_n)$. Comme le reste d'une série convergente tend vers 0, on en déduit $\lim_{i \rightarrow +\infty} \mathbb{P}\left(\bigcup_{j \geq i} A_j\right) = 0$. Finalement, un passage à la limite $i \rightarrow +\infty$ dans (4.2) conclut. \square

Ce résultat admet un complément quand les évènements A_n sont indépendants :

Théorème 4.4.2 (Second lemme de Borel-Cantelli) *Soit $(A_n)_n$ une suite (infinie) d'évènements indépendants. On suppose que $\sum_{n=1}^{+\infty} \mathbb{P}(A_n) = +\infty$. Alors*

$$\mathbb{P}(\overline{\lim}_n A_n) = 1.$$

C'est à dire presque sûrement, une infinité d'évènements A_n sont réalisés.

Remarque : Le résultat reste vrai si l'indépendance des évènements A_n n'est vraie que deux à deux.

Démonstration : On a

$$\mathbb{P}(\overline{\lim}_n A_n) = 1 - \mathbb{P}(\underline{\lim}_n A_n^c).$$

Puis par monotonie séquentielle croissante en (4.3), décroissante en (4.4) et comme les évènements A_n^c sont aussi indépendants, on a :

$$\mathbb{P}(\underline{\lim}_n A_n^c) = \mathbb{P}\left(\bigcup_{n \geq 0} \bigcap_{k \geq n} A_k^c\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) \quad (4.3)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k=n}^q A_k^c\right) \quad (4.4)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \prod_{k=n}^q \mathbb{P}(A_k^c) \quad (4.5)$$

$$= \lim_{n \rightarrow +\infty} \lim_{q \rightarrow +\infty} \prod_{k=n}^q (1 - \mathbb{P}(A_k)).$$

Mais, pour tout $x \in \mathbb{R}$, on a $1 - x \leq e^{-x}$, il vient donc

$$0 \leq \prod_{k=n}^q (1 - \mathbb{P}(A_k)) \leq \exp\left(-\sum_{k=n}^q \mathbb{P}(A_k)\right)$$

Comme $\sum_k \mathbb{P}(A_k) = +\infty$, le passage à la limite $q \rightarrow +\infty$ donne $\lim_{q \rightarrow +\infty} \prod_{k=n}^q (1 - \mathbb{P}(A_k)) = 0$. On a donc $\mathbb{P}(\underline{\lim}_n A_n^c) = 0$, ce qui conclut. \square

Chapitre 5

Convergences de variables aléatoires

5.1 Convergence presque sûre

La convergence de variables aléatoires correspond à la convergence de fonctions. La convergence la plus faible est la convergence simple qui s'énoncerait : pour tout $\omega \in \Omega$, $X_n(\omega) \rightarrow X(\omega)$ quand $n \rightarrow +\infty$. En probabilité (et plus généralement en théorie de la mesure), il est trop restrictif de demander la convergence pour **tous** les $\omega \in \Omega$. À la place, on définit la convergence presque sûre :

Définition 5.1.1 Une suite $(X_n)_n$ de variables aléatoires converge presque sûrement vers X si la convergence est vraie avec une probabilité 1

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

On note la convergence presque sûre : $X_n \rightarrow X$ p.s.

D'un point de vue analyse, c'est une convergence plus faible que la convergence simple (déjà très faible). Mais d'un point de vue probabiliste, ce sera une des convergences les plus fortes que nous manipulerons.

Dans la suite, on étudiera notamment les liens entre cette convergence et les autres.

5.2 Convergence en norme p

On rappelle que pour $p > 1$, on définit les espaces L^p :

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{R} : \|X\|_p = (\mathbb{E}[|X|^p])^{1/p} < +\infty\}.$$

Ce sont des espaces de Banach pour la norme $\|\cdot\|_p$. Remarquez que pour $p = 2$, $L^2(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Hilbert car la norme 2 dérive du produit scalaire suivant

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

Définition 5.2.1 On dit qu'une suite de v.a. $(X_n)_n$ converge en norme p vers X si $\|X_n - X\|_p \rightarrow 0$ quand $n \rightarrow +\infty$, càd

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

On note la convergence en norme p : note $X_n \xrightarrow{L^p} X$.

Proposition 5.2.1 Si $p \geq q$, la convergence en norme p entraîne la convergence en norme q .

Démonstration : Soit $p > q$. Avec $\alpha = \frac{p}{q} > 1$ et β l'indice conjugué ($\frac{1}{\alpha} + \frac{1}{\beta} = 1$), on a d'après l'inégalité de Hölder pour toute v.a. X :

$$\begin{aligned} \mathbb{E}[|X|^q] &\leq \mathbb{E}[|X|^{q\alpha}]^{1/\alpha} \mathbb{E}[1^\beta]^{1/\beta} \\ &\leq \mathbb{E}[|X|^p]^{q/p}. \end{aligned}$$

On en déduit $\|X\|_q \leq \|X\|_p$. Appliquée à la v.a. $X_n - X$, on a $\|X_n - X\|_q \leq \|X_n - X\|_p$, ce qui conclut. Au passage, noter qu'on a montré que si $p \geq q$: $L^p(\Omega, \mathcal{F}, \mathbb{P}) \subset L^q(\Omega, \mathcal{F}, \mathbb{P})$. \square

En particulier, la convergence en norme p entraîne toujours la convergence en norme 1.

5.3 Convergence en loi

Définition 5.3.1 Soit $(X_n)_n$ une suite de variables aléatoires réelles. On dit que $(X_n)_n$ converge en loi vers la variable aléatoire X si la fonction de répartition $F_n(t)$ de X_n converge vers $F(t)$, celle de X en tout point t où F est continue (c'est à dire en les points t tels que $\mathbb{P}(X = t) = 0$). On note la convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$.

Remarque : Si X est une variable à densité (de fonction de répartition continue), il n'y a plus d'hypothèse restrictive à faire sur les points de continuité t dans la définition de la convergence en loi, car pour tout t , $\mathbb{P}(X = t) = 0$.

Exemple : Soit X_n de loi de Bernoulli donnée par $\mathbb{P}(X_n = 1 + 1/n) = \mathbb{P}(X_n = 0) = \frac{1}{2}$. Alors la suite $(X_n)_n$ converge en loi vers X de loi $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$.

Remarque : La convergence en loi ne vérifie pas de bonnes propriétés arithmétiques : si par exemple X_n tend en loi vers X , il n'est pas vrai que $X_n - X$ tend vers 0. En effet, prendre par exemple, X_n et X des v.a. de loi normale centrée réduite $\mathcal{N}(0, 1)$ alors la loi des X_n coïncide avec celle de X mais aussi celle de $-X$ (car X et $-X$ ont même loi, par parité de la densité de $\mathcal{N}(0, 1)$!). Donc les X_n convergent en loi vers $-X$, pourtant, $X_n - X$ suit une loi normale $\mathcal{N}(0, 2)$ qui n'est pas la loi de $0 = X - X$.

Cependant, on a

Proposition 5.3.1 Si la suite $(X_n)_n$ converge en loi vers X et si $(Y_n)_n$ converge en loi vers 0 alors $X_n + Y_n$ converge en loi vers X .

Remarque : On peut expliquer intuitivement ce résultat : dire que $(X_n)_n$ tend en loi vers X , c'est dire que les répartitions statistiques de X_n et X ont tendance à être voisines. Dire que Y_n tend vers 0, c'est dire que statistiquement, Y_n est souvent voisin de 0. Il en résulte qu'ajouter Y_n à X_n modifie souvent peu X_n et donc également la répartition statistique des valeurs prises par X_n .

Démonstration : Soit t réel tel que la fonction de répartition F_X de X soit continue en t . Soit $\alpha > 0$, pour avoir $X_n + Y_n \leq t$, on a soit $X_n \leq t + \alpha$, soit $Y_n \leq -\alpha$, c'est à dire $\{X_n + Y_n \leq t\} \subset \{X_n \leq t + \alpha\} \cup \{Y_n \leq -\alpha\}$ d'où

$$\mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(X_n \leq t + \alpha) + \mathbb{P}(Y_n \leq -\alpha).$$

D'une façon analogue

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X_n \leq t - \alpha) - \mathbb{P}(Y_n \geq \alpha).$$

Fixons $\varepsilon > 0$, par continuité de F_X en t , il existe α tel que

$$F_X(t) = \mathbb{P}(X \leq t) \leq F_X(t + \alpha) = \mathbb{P}(X \leq t + \alpha) \leq \mathbb{P}(X \leq t) + \frac{\varepsilon}{3}.$$

On peut s'arranger pour choisir α tel que F_X soit continue en $t + \alpha$. En effet, comme F_X est croissante, (c'est une fonction réglée et donc) elle a au plus une infinité dénombrable de points de discontinuité, en particulier ses points de continuité sont partout denses. On peut donc choisir α' tel que $t < t + \alpha' < t + \alpha$ et tel que F_X soit continue en $t + \alpha'$. On a alors par croissance de F_X

$$\mathbb{P}(X \leq t) \leq \mathbb{P}(X \leq t + \alpha') \leq \mathbb{P}(X \leq t + \alpha) \leq \mathbb{P}(X \leq t) + \frac{\varepsilon}{3}.$$

De plus, par la convergence faible et la continuité de F_X en $t + \alpha'$:

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \leq t + \alpha') = \mathbb{P}(X \leq t + \alpha').$$

Il existe donc n_0 tel que pour tout $n \geq n_0$,

$$\mathbb{P}(X_n \leq t + \alpha') \leq \mathbb{P}(X \leq t + \alpha') + \frac{\varepsilon}{3}.$$

D'autre part, la convergence en loi de Y_n vers 0 entraîne que $\mathbb{P}(Y_n \leq -\alpha')$ tend vers $\mathbb{P}(0 \leq -\alpha') = 0$ (la fonction de répartition de la v.a. 0 est continue en $-\alpha'$). Il existe donc n_1 tel que pour $n \geq n_1$,

$$\mathbb{P}(Y_n \leq -\alpha') \leq \frac{\varepsilon}{3}.$$

Finalement, pour tout $n \geq N_0 = \max(n_0, n_1)$, on a

$$\begin{aligned} \mathbb{P}(X_n + Y_n \leq t) &\leq \mathbb{P}(X_n \leq t + \alpha') + \mathbb{P}(Y_n \leq -\alpha') \\ &\leq \mathbb{P}(X \leq t + \alpha') + \varepsilon/3 + \varepsilon/3 \\ &\leq \mathbb{P}(X \leq t) + \varepsilon. \end{aligned}$$

On montre de même (en raisonnant de l'autre côté) l'existence de N_1 tel que pour tout $n \geq N_1$,

$$\mathbb{P}(X_n + Y_n \leq t) \geq \mathbb{P}(X \leq t) - \varepsilon.$$

La convergence en loi en résulte : pour tout $n \geq \max(N_0, N_1)$, on a en tout point t de continuité de F_X :

$$\begin{aligned} \mathbb{P}(X \leq t) - \varepsilon &\leq \mathbb{P}(X_n + Y_n \leq t) \leq \mathbb{P}(X \leq t) + \varepsilon \\ F_X(t) - \varepsilon &\leq F_{X_n+Y_n}(t) \leq F_X(t) + \varepsilon. \end{aligned}$$

□

Théorème 5.3.1 Une suite de va X_n converge en loi vers X ssi pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée :

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

Démonstration : Nous nous contentons de prouver que l'une des deux implications (la plus intéressante) : si $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ quand $n \rightarrow +\infty$ pour toute fonction f continue bornée alors X_n converge en loi vers X .

Soit $x \in \mathbb{R}$ un point de continuité de F , la fonction de répartition de X . Choisissons $y > x$ (à préciser dans un instant) et soit f la fonction égale à 1 pour $u \leq x$ et égale à 0 pour $u \geq y$, et affine entre x et y . Cette fonction est continue et bornée par 1. Puis comme $\mathbf{1}_{]-\infty, x]} \leq f \leq \mathbf{1}_{]-\infty, y]}$, on a

$$\mathbb{P}(X_n \leq x) = \mathbb{E}[\mathbf{1}_{]-\infty, x]}(X_n)] \leq \mathbb{E}[f(X_n)] \leq \mathbb{E}[\mathbf{1}_{]-\infty, y]}(X_n)] \leq \mathbb{P}(X_n \leq y) \quad (5.1)$$

c'est à dire $F_n(x) \leq \mathbb{E}[f(X_n)] \leq F_n(y)$. De la même manière, on montre que $F(x) \leq \mathbb{E}[f(X)] \leq F(y)$.

Fixons $\varepsilon > 0$, par continuité de F en x , il existe $y_0 \geq x$ tel que $F(x) \leq F(y_0) \leq F(x) + \varepsilon/2$. On choisit la fonction f_0 associée comme précédemment à ce y_0 . Par hypothèse pour cette fonction f_0 continue bornée, il existe n_0 tel que

$$\forall n \geq n_0, \quad \mathbb{E}[f_0(X_n)] \leq \mathbb{E}[f_0(X)] + \varepsilon/2.$$

Il en résulte que pour tout $n \geq n_0$,

$$F_n(x) \leq \mathbb{E}[f_0(X_n)] \leq \mathbb{E}[f_0(X)] + \varepsilon/2 \leq F(y_0) + \varepsilon/2 \leq F(x) + \varepsilon. \quad (5.2)$$

En choisissant de façon analogue $y < x$, on construit des fonctions \tilde{f} pour lesquelles on obtient des inégalités inversées :

$$F_n(y) \leq \mathbb{E}[\tilde{f}(X_n)] \leq F_n(x), \quad F(y) \leq \mathbb{E}[\tilde{f}(X)] \leq F(x). \quad (5.3)$$

Par continuité de F , on peut alors choisir $y_1 \leq x$ tel que $F(x) - \varepsilon/2 \leq F(y_1) \leq F(x)$; puis pour la fonction \tilde{f}_1 associée à ce y_1 comme ci-dessus, l'hypothèse donne l'existence de n_1 tel que

$$\forall n \geq n_1, \quad \mathbb{E}[\tilde{f}_1(X_n)] \geq \mathbb{E}[\tilde{f}_1(X)] - \varepsilon/2$$

puis pour tout $n \geq n_1$,

$$F_n(x) \geq \mathbb{E}[\tilde{f}_1(X_n)] \geq \mathbb{E}[\tilde{f}_1(X)] - \varepsilon/2 \geq F(y_1) - \varepsilon/2 \geq F(x) - \varepsilon. \quad (5.4)$$

Finalement avec (5.2), (5.4), on a obtenu la convergence de $F_n(x)$ vers $F(x)$. \square

Idée : Pour la démonstration de l'implication réciproque, le cheminement est le suivant : on commence par montrer que l'on peut se ramener au cas de fonction f continue sur un intervalle compact $[-a, a]$. En effet, $\mathbb{P}(|X| > a)$ tend vers 0 quand a tend vers $+\infty$; puis il en est de même pour $\mathbb{P}(|X_n| > a)$ en utilisant la convergence en loi et un « bon » a pour majorer toutes ces probabilités par un même ε .

Une fois que l'on s'est ramené à un intervalle compact, on approche uniformément f par une fonction en escalier h à $\varepsilon > 0$ près. On a alors $|\mathbb{E}[f(X)] - \mathbb{E}[h(X)]| \leq \varepsilon$ (et de même avec X_n , pour tout n). Puis on montre que si les extrémités des escaliers sont des points de continuité de la fonction de répartition F_X de X alors $\mathbb{E}[h(X_n)]$ converge vers $\mathbb{E}[h(X)]$ ce qui donne la conclusion souhaitée pour h et donc celle pour f à ε près. Finalement on fait tendre ε vers 0 et on a la conclusion aussi pour f .

Il reste à remarquer que l'on peut toujours s'arranger pour que la condition sur les extrémités de l'escalier soit vérifiée.

Remarque : Un des intérêts de la convergence en loi est de permettre d'approcher les lois inconnues (ou connues mais difficilement calculables parfois) des X_n par celle (souvent plus simple) de la limite X , comme par exemple dans le théorème central limite (cf. chapitre 7).

D'où l'importance de critères garantissant cette convergence sans repasser par la définition. Le critère ci-dessus n'est en pratique pas très utilisable. Nous en aurons un bien plus pratique avec la notion de fonction caractéristique (théorème de Paul Lévy, cf. chapitre 7).

Proposition 5.3.2 *Si la suite $(X_n)_n$ converge presque sûrement vers X , elle converge en loi vers X .*

Démonstration : On voit la convergence en loi par le critère du théorème 5.3.1 : soit f une fonction continue bornée. Si $X_n \rightarrow X$, alors $f(X_n) \rightarrow f(X)$ presque sûrement et $f(X_n)$ est bornée par $\|f\|_\infty$, qui comme toutes les constantes est \mathbb{P} -intégrable. Mais alors par convergence dominée, $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$, ce qui justifie la convergence en loi par le théorème 5.3.1. \square

5.4 Convergence en probabilité

La convergence en probabilité de X_n vers X garantit que la probabilité que « X_n et X restent éloignées » tend vers 0.

Définition 5.4.1 Une suite $(X_n)_n$ de variables aléatoires converge en probabilité vers la variable aléatoire X si

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

On note la convergence en proba : $X_n \xrightarrow{\mathbb{P}} X$.

Convergences en proba et en loi

Proposition 5.4.1 La convergence en probabilité implique la convergence en loi.

Démonstration : Il s'agit de montrer que si $(X_n)_n$ converge en probabilité vers X , la suite $F_n(x) = \mathbb{P}(X_n \leq x)$ converge vers $F(x) = \mathbb{P}(X \leq x)$ en tout point x où F est continue.

Soit $\varepsilon > 0$, $X_n = X + (X_n - X)$ est inférieure à x si soit $X \leq x + \varepsilon$ soit $X_n - X \leq -\varepsilon$. On a donc $\{X_n \leq x\} \subset \{X \leq x + \varepsilon\} \cup \{X_n - X \leq -\varepsilon\}$ puis

$$\begin{aligned} \mathbb{P}(X_n \leq x) &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(X_n - X \leq -\varepsilon) \\ &\leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| \geq \varepsilon) \end{aligned}$$

Soit $\eta > 0$, par continuité de F_X en x , il existe ε_0 tel que $F_X(x) \leq F_X(x + \varepsilon_0) \leq F_X(x) + \eta/2$. Pour ce choix de ε_0 , il existe n_0 tel que pour tout $n \geq n_0$, $\mathbb{P}(|X - X_n| \geq \varepsilon_0) \leq \eta/2$. Finalement, pour tout $n \geq n_0$,

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x) \leq F_X(x + \varepsilon_0) + \mathbb{P}(|X_n - X| \geq \varepsilon_0) \leq F_X(x) + \eta/2 + \eta/2 \leq F_X(x) + \eta.$$

Un raisonnement analogue partant de l'inégalité

$$\mathbb{P}(X \leq x - \varepsilon) \leq \mathbb{P}(X_n \leq x) + \mathbb{P}(X - X_n \leq -\varepsilon)$$

réécrite sous la forme

$$\mathbb{P}(X_n \leq x) \geq \mathbb{P}(X \leq x - \varepsilon) - \mathbb{P}(X - X_n \leq -\varepsilon)$$

puis

$$F_{X_n}(x) \geq F_X(x - \varepsilon) - \mathbb{P}(|X - X_n| \geq \varepsilon)$$

permet de montrer la seconde inégalité et finalement la convergence de $F_{X_n}(x)$ vers $F_X(x)$, en x point de continuité de F_X . \square

Remarque : La réciproque est fautive en général : la convergence en loi n'entraîne pas la convergence en probabilité. Cependant, dans un cas bien particulier, on a une réciproque partielle :

Proposition 5.4.2 Si X_n converge en loi vers la constante c alors X_n converge en probabilité vers c .

Démonstration : En effet, soit pour $\varepsilon > 0$, $h = \mathbf{1}_{]-\infty, c-\varepsilon] \cup [c+\varepsilon, +\infty[}$ et g la fonction qui vaut 1 si $x \leq c - \varepsilon$ ou $x \geq c + \varepsilon$, 0 en c et affine entre $c - \varepsilon$ et c puis entre c et $c + \varepsilon$. La fonction g est continue, bornée et $h \leq g$. On a

$$\mathbb{P}(|X_n - c| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}}] = \int h(X_n) d\mathbb{P} \leq \int g(X_n) d\mathbb{P} \rightarrow \int g(c) d\mathbb{P} = g(c) = 0$$

en appliquant le théorème de convergence dominée. On a donc $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - c| \geq \varepsilon) = 0$. \square

Convergences en proba et en norme p

Proposition 5.4.3 *Si X_n et X sont dans $L^p(\Omega)$ alors la convergence en norme $\|\cdot\|_p$ $X_n \xrightarrow{L^p} X$ entraîne la convergence en probabilité $X_n \xrightarrow{\mathbb{P}} X$.*

Démonstration : Il suffit d'appliquer l'inégalité de Markov : soit $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^p \geq \varepsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} = \frac{\|X_n - X\|_p^p}{\varepsilon^p} \rightarrow 0, n \rightarrow +\infty.$$

Conséquence : La convergence en norme $\|\cdot\|_p$ entraîne donc aussi la convergence en loi.

Convergences en proba et ps

Proposition 5.4.4 *Si la suite $(X_n)_n$ converge presque sûrement vers X , elle converge en probabilité (et donc aussi en loi) vers X .*

Démonstration : Soit $\varepsilon > 0$ fixé. On pose $f_n = \mathbf{1}_{\{|X_n - X| > \varepsilon\}}$. Les fonctions f_n sont mesurables, positives et majorées par 1. Par ailleurs, d'après la convergence presque sûre, pour presque tout $\omega \in \Omega$, il existe n_0 tel que pour tout $n \geq n_0$,

$$|X_n(\omega) - X(\omega)| \leq \varepsilon$$

c'est à dire $f_n(\omega) = 0$ pour tout $n \geq n_0$. On a donc écrit que la suite f_n converge presque sûrement vers 0. D'après le théorème de convergence dominée

$$\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{E}[f_n] \longrightarrow \mathbb{E}[\lim_{n \rightarrow +\infty} f_n] = \mathbb{E}[0] = 0.$$

\square

Ce résultat admet une réciproque partielle qu'on obtient grâce au lemme de Borel-Cantelli :

Proposition 5.4.5 *Si la suite X_n converge en probabilité vers X , alors elle admet une sous suite $(X_{n_p})_p$ qui converge presque sûrement vers X .*

Démonstration : La convergence en probabilité avec $\varepsilon = 1/p$ donne :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq 1/p) = 0$$

donc pour n assez grand, $\mathbb{P}(|X_n - X| \geq 1/p)$ devient plus petit que $1/2^p$. Soit n_p tel que

$$\mathbb{P}(|X_{n_p} - X| \geq 1/p) \leq 1/2^p.$$

On peut supposer la suite $(n_p)_p$ croissante.

Comme $\sum_{p=1}^{+\infty} 1/2^p < +\infty$, d'après le (premier) lemme de Borel-Cantelli, avec les évènements $A_p = \{|X_{n_p} - X| \geq 1/p\}$, on a $\mathbb{P}(\overline{\lim}_p A_p) = 0$, c'est à dire que presque sûrement, seuls un nombre fini d'évènements A_p sont réalisés. Il existe p_0 tel que pour tout $p \geq p_0$, l'évènement A_p n'est pas réalisé, c'est à dire

$$|X_{n_p} - X| \leq 1/p.$$

En faisant tendre p vers $+\infty$, on obtient, $\lim_{p \rightarrow +\infty} X_{n_p} = X$, et ce avec probabilité 1. \square

On a un critère de convergence en proba qui s'énonce avec la convergence ps :

Proposition 5.4.6 *Une condition nécessaire et suffisante pour que $X_n \xrightarrow{\mathbb{P}} X$ est que pour toute sous suite $(X_{n_k})_k$ de $(X_n)_n$, il existe une sous suite encore plus fine $(X_{n_{k(p)}})_p$ telle que $X_{n_{k(p)}}$ converge ps vers X .*

Démonstration : • **Sens direct.** Si $X_n \xrightarrow{\mathbb{P}} X$ alors étant donnée une sous suite $\{n_k\}$, on choisit $\{n_{k(p)}\}$ telle que $k \geq k(p)$ implique

$$\mathbb{P}(|X_{n_k} - X| \geq 1/p) < 1/2^p.$$

D'après le premier lemme de Borel-Cantelli, on a ps $|X_{n_{k(p)}} - X| < 1/p$ pour tout p assez grand. Mais alors $\lim_{p \rightarrow +\infty} X_{n_{k(p)}} = X$ ps.

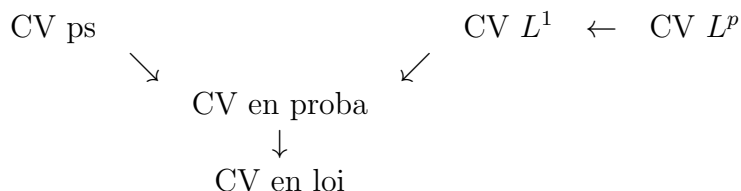
• **Sens indirect.** Si X_n ne converge pas en proba vers X , alors il existe $\varepsilon > 0$ et $\eta > 0$ pour lequel $\mathbb{P}(|X_{n_k} - X| > \varepsilon) > \eta$ le long d'une sous suite $\{n_k\}$. Aucune sous suite extraite de $(X_{n_k})_k$ ne peut converger en proba vers X et donc aucune ne peut converger vers X avec proba 1. \square

Conséquences.

- Unicité ps de la limite en proba : Si $X_n \xrightarrow{\mathbb{P}} X$ et $X_n \xrightarrow{\mathbb{P}} Y$ alors $X = Y$ ps.
- Si f est continue et $X_n \xrightarrow{\mathbb{P}} X$ alors $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

En effet, si $X_n \xrightarrow{\mathbb{P}} X$. Soit $(n_k)_k$ une sous suite, on peut extraire $(n_{k_p})_p \subset (n_k)_k$, telle que $X_{n_{k_p}}$ converge ps vers X . Comme f est continue, on a aussi $f(X_{n_{k_p}}) \rightarrow f(X)$ quand $p \rightarrow +\infty$. D'après le critère précédent, $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Remarque : Les liens entre les différentes convergences sont donnés par le diagramme suivant :



5.5 Lois des grands nombres (LGN)

5.5.1 Version faible de la 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 (dite faible) pour la convergence *en probabilité*.

Théorème 5.5.1 (LGN faible) *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes et de même loi avec un moment d'ordre 2. Alors*

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

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.

Démonstration : Ici, la v.a. limite est la constante $\mathbb{E}[X_1]$ (= $\mathbb{E}[X_i]$ pour tout i car les v.a. X_i ont même loi, donc même espérance). Il s'agit de vérifier

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

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

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

L'inégalité de Tchebychev appliquée à M_n donne alors pour tout $\varepsilon > 0$ et $n \in \mathbb{N}^*$:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} \left(\sum_{i=1}^n X_i\right) - \mathbb{E}[X_1]\right| \geq \varepsilon\right) &= \mathbb{P}(|M_n - \mathbb{E}[M_n]| \geq \varepsilon) \leq \frac{\text{Var}(M_n)}{\varepsilon^2} \\ &\leq \frac{\text{Var}(X_1)}{n\varepsilon^2}. \end{aligned} \quad (5.5)$$

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

Remarque 5.5.1 Plus que la convergence, nous avons obtenu la *vitesse* de convergence : d'après (5.5) elle est en $O(1/n)$. Si on connaît $\text{Var}(X_1)$, on peut donc pour une proportion donnée, trouver un rang n_0 tel que pour $n \geq n_0$ et pour cette proportion de $\omega \in \Omega$, on ait la moyenne arithmétique ε -proche de l'espérance. Ceci est utile pour déterminer des intervalles de confiance.

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

Corollaire 5.5.1 Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de Bernoulli de même paramètre p . Alors

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

Démonstration : $\mathbb{E}[X_i] = p$ car $X_i \sim b(p)$ et comme $\text{Var}(X_i) = p(1-p) < +\infty$, la LGN s'applique. \square

C'est ce résultat qui formalise le résultat intuitif sur le lancer des dés ou des pièces : avec $X_i = \mathbf{1}_{\{\text{obtenir un 4 au } i\text{-ème lancer}\}}$, on a $X_i \sim b(1/6)$ et $p = 1/6$ et $\frac{1}{n} \sum_{i=1}^n X_i$ désigne la fréquence d'apparition du 4 sur les n premiers lancers.

5.5.2 Application : estimation d'une proportion inconnue

On se propose d'estimer le paramètre p inconnu d'une loi de Bernoulli en observant un grand nombre de fois un phénomène aléatoire de loi de Bernoulli $b(p)$, c'est à dire en

observant les réalisations d'une suite de v.a. indépendantes $X_i(\omega)$, $1 \leq i \leq n$, de loi de Bernoulli $b(p)$.

Modélisons le problème de la façon suivante. On a une urne comportant des boules rouges en proportion inconnue p et des boules vertes (en proportion $1 - p$). On effectue n tirages d'une boule avec remise. Notons

$$X_i = \begin{cases} 0 & \text{si la boule tirée est rouge} \\ 1 & \text{si la boule tirée est verte} \end{cases} = \mathbf{1}\{\text{la boule est rouge au } i\text{-ème tirage}\}.$$

Désignons toujours par M_n la moyenne arithmétique des n premières v.a. X_i , ici cela correspond à la fréquence d'apparition des boules rouges lors des n premiers tirages. D'après la loi faible des grands nombres (ou plutôt son corollaire 5.5.1 pour les fréquences), M_n converge en probabilité vers p :

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p.$$

On v.a. donc estimer p par M_n pour n assez grand. En fait, on observe *une* valeur particulière $M_n(\omega)$ calculée à partir des n tirages réellement effectués correspondant à la réalisation ω du hasard. La question pratique qui se pose est de donner une fourchette pour l'approximation de p par la valeur observée $M_n(\omega)$ et contrôler le risque que cette approximation ne soit pas valable. L'inégalité de Tchebychev pour M_n s'écrit ici :

$$\mathbb{P}(|M_n - p| \geq t) \leq \frac{\text{Var}(X_1)}{nt^2} = \frac{p(1-p)}{nt^2} \leq \frac{1}{4nt^2}$$

car on majore facilement $p(1-p)$ par $1/4$ (la fonction $x \mapsto x(1-x)$ sur $[0, 1]$ a son maximum en $1/2$ et il vaut $1/4$). D'où

$$\mathbb{P}(M_n - t < p < M_n + t) = 1 - \mathbb{P}(|M_n - p| \geq t) \geq 1 - \frac{1}{4nt^2}. \quad (5.6)$$

En pratique, on observe $M_n(\omega)$ et on dit que $I(\omega) =]M_n(\omega) - t, M_n(\omega) + t[$ est un intervalle de confiance (ou fourchette). L'équation (5.6) permet de voir que la probabilité de se tromper (*i.e.* en fait $p \notin I(\omega)$) est majorée par $1/(4nt^2)$.

Si on se fixe un seuil d'erreur α , on trouve t_α tel que $1/(4nt_\alpha^2) = \alpha$ et l'intervalle $I_\alpha(\omega) =]M_n(\omega) - t_\alpha, M_n(\omega) + t_\alpha[$ est alors l'intervalle de confiance au niveau α .

Exemple (Sondage) : Avant un référendum, un institut de sondage interroge au hasard 1000 personnes dans la rue. On note p la proportion d'électeurs décidés à voter OUI dans la population totale. Dans l'échantillon sondé, cette proportion est égale à 0,54. Proposer un intervalle de confiance pour p au niveau 0,95.

Le sondage peut être assimilé à un tirage avec remise (la réponse d'un électeur de l'échantillon correspondant au tirage d'une boule d'une certaine couleur selon son choix de vote), on est alors ramené à la situation de l'exemple précédent. Ici, la fréquence

observée du choix du OUI sur les 1000 électeurs est $M_{1000}(\omega) = 0,54$ et l'intervalle de confiance est

$$I =]0,54 - t; 0,54 + t[\quad \text{avec un niveau } \alpha \geq 1 - 1/(4 \times 1000 \times t^2).$$

Ici, on veut un seuil de confiance $\alpha \geq 0,95$, il faut alors

$$1 - \frac{1}{4000 \times t^2} \geq 0,95 \iff \frac{1}{4000 \times t^2} \leq 0,05 \iff t \geq \frac{1}{10\sqrt{2}} \simeq 0,0707.$$

Avec $t = 0,071$, on trouve $I =]0,469; 0,611[$. On constate en particulier qu'une zone de l'intervalle de confiance correspond à $p < 1/2$, pour lequel le OUI ne serait pas gagnant alors que la proportion observée semblait lui garantir la victoire. On ne peut pas garantir la victoire du OUI avec une probabilité d'erreur inférieure à 5%.

Combien de personnes faut-il alors interroger pour donner une fourchette à $\pm 1\%$ avec un seuil de 95% ?

Reprenons de (5.6), avec une fourchette $t = 0,01$. On veut $\alpha \geq 0,95$ soit $1 - \alpha \leq 0,05$:

$$\frac{1}{4n \times 0,01^2} \leq 0,05.$$

On trouve $n = 50\,000$, ce qui donne au sondage un coût prohibitif. En gros, pour améliorer la précision d'un facteur 10, il faut interroger 100 fois plus de personnes et donc multiplier les coûts par 100.

5.5.3 Version forte de la LGN

La loi des grands nombres admet une version pour la convergence presque sûre : c'est la LGN forte.

Théorème 5.5.2 (LGN forte) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi avec un moment d'ordre 1 (i.e. $\mathbb{E}[|X_1|] < +\infty$, si bien que l'espérance est bien définie). Alors*

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

Pour commencer, on donne la preuve de la LGN avec l'hypothèse $\mathbb{E}[X_1^4] < +\infty$ (existence d'un moment d'ordre 4). Historiquement, c'est la 1ère preuve de la LGN forte.

Démonstration : Il suffit de prouver le théorème quand $\mathbb{E}[X_1] = 0$, le cas général s'obtenant par translation. Posons

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n = \sum_{i=1}^n X_i.$$

Soit $\varepsilon > 0$ et $D_\varepsilon = \overline{\lim}_{n \rightarrow +\infty} \{|M_n| \geq \varepsilon\}$.

On va utiliser le lemme de Borel-Cantelli pour montrer que $\mathbb{P}(D_\varepsilon) = 0$.

On conclura alors en montrant que $D = \bigcup_n D_{1/n}$ est de probabilité nulle. En effet on a $D = \{M_n \not\rightarrow 0\}$, donc le résultat est acquis si on montre que D est de probabilité nulle.

Afin d'utiliser le lemme de Borel-Cantelli, on montre la convergence de la série de terme général $\mathbb{P}(|M_n| \geq \varepsilon)$. Or

$$\mathbb{P}(|M_n| \geq \varepsilon) = \mathbb{P}(|S_n| \geq n\varepsilon) = \mathbb{P}(|S_n|^4 \geq n^4\varepsilon^4).$$

Par l'inégalité de Markov, on a alors

$$\mathbb{P}(|M_n| \geq \varepsilon) \leq \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4}.$$

Il reste à estimer $\mathbb{E}[S_n^4]$.

$$\begin{aligned} S_n^4 &= (X_1 + X_2 + \dots + X_n)^4 \\ &= \sum_{k_1, k_2, k_3, k_4 \in \{1, \dots, n\}^4} X_{k_1} X_{k_2} X_{k_3} X_{k_4} \\ &= \begin{cases} M(4) \sum_{i=1}^n X_i^4 & + M(1, 3) \sum_{1 \leq i < j \leq n} X_i^3 X_j \\ + M(2, 2) \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 & + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} X_i^2 X_j X_k \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} X_i X_j X_k X_l \end{cases} \end{aligned}$$

où $M(i_1, \dots, i_p)$ désigne le nombre de 4-uplets (u_1, \dots, u_4) de $\{1, \dots, n\}$ en prenant i_1 fois la valeur u_1, \dots, i_p fois la valeur u_p .

La linéarité et l'indépendance des X_i donnent alors $\mathbb{E}[S_n^4] =$

$$\begin{cases} M(4) \sum_{i=1}^n \mathbb{E}[X_i^4] & + M(1, 3) \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^3] \mathbb{E}[X_j] \\ + M(2, 2) \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] & + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j] \mathbb{E}[X_k] \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j] \mathbb{E}[X_k] \mathbb{E}[X_l] \end{cases}$$

Comme $\mathbb{E}[X_i] = 0$, les deuxième, quatrième et cinquième termes sont nuls. Comme on montre que $M(4) = 1$, $M(2, 2) = 6$, on obtient

$$E(S_n^4) = \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i^2] \mathbb{E}[X_j^2]$$

$$\begin{aligned}
&= n\mathbb{E}[X_1^4] + 6C_n^2(\mathbb{E}[X_1^2])^2 \\
&= n\mathbb{E}[X_1^4] + 3n(n-1)(\mathbb{E}[X_1^2])^2 \\
&\leq Mn + 3Mn(n-1) \\
&\leq 3Mn^2 < +\infty
\end{aligned}$$

où on a posé $M = \max(\mathbb{E}[X_1^2]^2, \mathbb{E}[X_1^4])$. On a alors

$$\mathbb{P}(|M_n| \geq \varepsilon) \leq \frac{\mathbb{E}[S_n^4]}{n^4\varepsilon^4} \leq \frac{3M}{n^2\varepsilon^4}.$$

Comme $3M/n^2\varepsilon^4$ est le terme général d'une série convergente, $\mathbb{P}(|M_n| \geq \varepsilon)$ aussi. Le lemme de Borel-Cantelli s'applique et donne $\mathbb{P}(D_\varepsilon) = 0$. Posons alors

$$D = \bigcup_{p=1}^{+\infty} D_{1/p}$$

On a $\mathbb{P}(D) = 0$ car D est réunion dénombrable d'ensembles $D_{1/p}$ de probabilités nulles. Prenons alors

$$\Omega_0 := D^c = \bigcap_{p \geq 1} D_{1/p}^c = \bigcap_{p \geq 1} \bigcup_{k \geq 1} \bigcap_{n \geq k} \{|M_n| \leq 1/p\}.$$

On a $\mathbb{P}(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$, par traduction dans le langage logique des symboles ensemblistes, pour tout $p \in \mathbb{N}^*$, il existe un entier k tel que pour tout $n \geq k$ $|M_n| \leq 1/p$.

On a donc M_n qui converge presque sûrement vers 0 ; ce qui achève la preuve de la LGN forte. \square

On donne maintenant une preuve de la LGN forte avec pour seule hypothèse l'existence du moment d'ordre 1. Cette preuve est issue du manuel de probabilité de l'école Polytechnique de J. Neveu. Son inconvénient est d'être moins naturelle que celle où le moment d'ordre 4 est supposé fini. Comme la convergence ps entraîne la convergence en probabilité, on obtient aussi la LGN faible sous l'existence du moment d'ordre 1, cf. théorème 5.5.1.

Démonstration : On suppose que $\mathbb{E}[|X_1|] < +\infty$. Notons

$$S_n = \sum_{k=0}^n X_k.$$

Nous allons montrer que pour tout $a \geq \mathbb{E}[X_1]$, la va $M_n = \sup_{n \geq 0} (S_n - na)$ est finie ps.

Étape 1. De l'inégalité $S_n \leq na + M$ (qui découle de la définition de M), on déduit ps

$$\limsup_{n \rightarrow +\infty} (S_n/n) \leq a$$

En faisant $a \searrow \mathbb{E}[X_1]$, on obtient avec proba 1

$$\limsup_{n \rightarrow +\infty} (S_n/n) \leq \mathbb{E}[X_1]. \quad (5.7)$$

Le résultat appliqué aux va $-X_n$ donne aussi avec proba 1 : $\limsup_{n \rightarrow +\infty} (-S_n/n) \leq \mathbb{E}[-X_1]$, càd avec proba 1

$$\liminf_{n \rightarrow +\infty} (S_n/n) \geq \mathbb{E}[X_1]. \quad (5.8)$$

Finalement, (5.7) et (5.8) sont vraies ensemble avec proba 1, càd ps

$$\lim_{n \rightarrow +\infty} \frac{S_n}{n} = \mathbb{E}[X_1].$$

Le but est donc de montrer que $\mathbb{P}(M < +\infty) = 1$.

Étape 2. On montre d'abord que $\mathbb{P}(M = +\infty)$ ne peut valoir que 0 ou 1. Pour cela, on introduit la suite de va positives intégrables

$$M_n = \max_{0 \leq k \leq n} (S_k - ka).$$

Cette suite croît vers M . L'évènement $\{M = +\infty\}$ est indépendant de X_1, \dots, X_p pour tout p car les bornes supérieures

$$M = \sup_{k \geq 0} (S_k - ka), \quad \sup_{k \geq p} (S_k - ka), \quad \text{et} \quad \sup_{k \geq p} (S_k - S_p - ka)$$

sont infinies en même temps et $\sup_{k \geq p} (S_k - S_p - ka)$ ne dépend que de X_{p+1}, X_{p+2}, \dots puisque $S_k - S_p = \sum_{i=p+1}^k X_i$ si $k > p$.

En particulier, $\{M = +\infty\}$ est indépendant de M_p , ce qui permet d'écrire pour tout $a < +\infty$ que

$$\mathbb{P}(M = +\infty, M_p \leq a) = \mathbb{P}(M = +\infty)\mathbb{P}(M_p \leq a).$$

Lorsque $p \rightarrow +\infty$, les évènements $\{M_p \leq a\}$ décroissent vers $\{M \leq a\}$ et les évènements $\{M = +\infty, M_p \leq a\}$ décroissent vers $\{M = +\infty, M \leq a\}$. On obtient alors à la limite $p \rightarrow +\infty$ par la propriété de monotonie séquentielle des probabilités :

$$0 = \mathbb{P}(M = +\infty, M \leq a) = \mathbb{P}(M = +\infty)\mathbb{P}(M \leq a).$$

En notant que $\{M \leq a\} \searrow \{M < +\infty\}$ quand $a \rightarrow +\infty$, il vient

$$0 = \mathbb{P}(M = +\infty)\mathbb{P}(M < +\infty).$$

Cela justifie que $\mathbb{P}(M = +\infty) = 0$ ou 1.

Étape 3. Il reste à montrer que $\mathbb{P}(M = +\infty) \neq 1$ si $a > \mathbb{E}[X_1]$. Pour cela, on introduit

$$\tilde{M}_n = \max_{0 \leq k \leq n} \left(\sum_{i=2}^{k+1} X_i - ka \right).$$

Ces M_n sont définies à partir de X_2, X_3, \dots de la même manière que les M_n à partir de X_1, X_2, \dots . Donc, il est clair que les M_n et \tilde{M}_n ont même loi pour tout $n \geq 1$ et aussi que $\tilde{M} = \lim_{k \rightarrow +\infty} \tilde{M}_k$ a même loi que M .

D'autre part, on voit facilement que

$$M_{n+1} = \max(0, X_1 - a + \tilde{M}_n),$$

ce qui se réécrit encore pour $n \geq 1$

$$M_{n+1} = \tilde{M}_n + X_1 - \min(a, \tilde{M}_n + X_1).$$

Noter que $\mathbb{E}[\tilde{M}_n] = \mathbb{E}[M_n] \leq \mathbb{E}[M_{n+1}]$ (car \tilde{M}_n et M_n ont même loi et car $M_n \leq M_{n+1}$). Donc, en prenant les espérances (finies par hypothèse sur le moment d'ordre 1), on a pour tout $n \geq 1$:

$$\mathbb{E}[X_1] - \mathbb{E}[\min(a, \tilde{M}_n + X_1)] = \mathbb{E}[M_{n+1}] - \mathbb{E}[M_n] \geq 0.$$

En passant à la limite $n \rightarrow +\infty$, on a donc

$$\mathbb{E}[\min(a, \tilde{M} + X_1)] \leq \mathbb{E}[X_1].$$

Si $\tilde{M} = +\infty$ avec proba 1, l'inégalité se réduit à $a \leq \mathbb{E}[X_1]$, ce qui est absurde.

On a donc $\mathbb{P}(\tilde{M} = +\infty) = 0$, càd $\mathbb{P}(M = +\infty) = 0$, ce qui conclut l'étape 3 et la preuve de la LGN. \square

Chapitre 6

Séries de variables aléatoires indépendantes et loi des grands nombres

Dans tout ce chapitre, on considère une suite $(X_n)_n$ de v.a.i (variables aléatoires indépendantes). On note $S_n = \sum_{k=1}^n X_k$ la somme et $M_n = S_n/n$ la moyenne arithmétique et on étudie la convergence de ces quantités quand $n \rightarrow +\infty$.

6.1 Sommes de variables aléatoires indépendantes

Proposition 6.1.1

- S_n converge ps ou S_n diverge ps.
- Soit $(a_n)_n$ une suite de réels, $a_n \rightarrow +\infty$ alors S_n/a_n diverge ps ou converge ps vers L une v.a. ps constante.

Démonstration :

- Pour tout p , on a $\{S_n \text{ converge}\} = \bigcap_{p \in \mathbb{N}} \{\sum_{k=p}^n X_k \text{ converge}\} \in \mathcal{F}^\infty$, la tribu asymptotique. Or comme \mathcal{F}^∞ est la tribu triviale, l'évènement $\{S_n \text{ converge}\}$ est de probabilité 0 ou 1.

- La convergence ou la divergence de S_n/a_n est due à la première partie. Puis, $\lim_{n \rightarrow +\infty} S_n/a_n = L$ est mesurable pour \mathcal{F}^∞ car

$$\left\{ \lim_n \frac{S_n}{a_n} \leq t \right\} = \bigcap_{p \in \mathbb{N}} \left\{ \lim_n \frac{\sum_{k=p}^n X_k}{a_n} \leq t \right\} \in \mathcal{F}^\infty.$$

On conclut avec le lemme qui suit. □

Lemme 6.1.1 Une v.a. L mesurable pour \mathcal{F}^∞ est ps constante : $\exists l \in \bar{\mathbb{R}}$ tel que $L = l$ ps.

Démonstration : Comme $\{L \leq t\} \in \mathcal{F}^\infty$, on a $F(t) = \mathbb{P}(L \leq t) \in \{0, 1\}$. Posons alors $l = \sup\{t \in \mathbb{R} \mid F(t) = 0\}$. On a $\mathbb{P}(L < l - \varepsilon) = 0$ et $\mathbb{P}(L > l + \varepsilon) = 0$ pour tout $\varepsilon > 0$. On en déduit $\mathbb{P}(L = l) = 1$. \square

Proposition 6.1.2 Soient $(X_n)_n$ des v.a. quelconques. Il existe $(a_n)_n$ telle que $X_n/a_n \rightarrow 0$ ps quand $n \rightarrow +\infty$.

Démonstration : Pour tout n , $\lim_{t \rightarrow +\infty} \mathbb{P}(|X_n| \geq t) \rightarrow 0$ donc il existe t_n tel que $\mathbb{P}(|X_n| \geq t_n) \leq 2^{-n}$. On a donc $\sum_n \mathbb{P}(|X_n| \geq t_n) \leq \sum_n 2^{-n} < +\infty$. D'après le lemme de Borel-Cantelli : ps, il existe $n_0(\omega)$ tel que pour $n \geq n_0$, on a $|X_n(\omega)| \leq t_n$. On prend alors $a_n = nt_n$. On a bien $|X_n|/a_n$ inférieur à $1/n$ pour n assez grand, ce qui conclut. \square

Proposition 6.1.3 Soit $(X_n)_n$ des v.a. et $\varepsilon_n > 0$. Si $\sum_{n=1}^{+\infty} \mathbb{P}(|X_{n+1} - X_n| \geq \varepsilon_n) < +\infty$ et $\sum_n \varepsilon_n < +\infty$. Alors X_n converge ps.

Démonstration : D'après le lemme de Borel-Cantelli : ps, il existe $n_0(\omega)$ tel que pour $n \geq n_0$, $|X_{n+1}(\omega) - X_n(\omega)| \leq \varepsilon_n$. On a donc pour tout $p, q \geq n_0(\omega)$,

$$|X_p - X_q| \leq \sum_{k=p}^q \varepsilon_k \leq \sum_{k=n}^{+\infty} \varepsilon_k \rightarrow 0, \quad n \rightarrow +\infty.$$

La suite $(X_n(\omega))_n$ est une suite de Cauchy de \mathbb{R} , elle est donc convergente. \square

6.2 Convergence des sommes S_n

6.2.1 Les convergences ps et L^1

Proposition 6.2.1 (critère de Cauchy) Soient $(X_n)_n$ une suite de va. On note

$$Y_n = \sup_{p, q \geq n} |X_p - X_q| \quad \text{et} \quad Z_n = \sup_{p \geq n} |X_p - X_n|.$$

On a alors les équivalences suivantes

$$(X_n)_n \text{ converge ps} \iff Y_n \xrightarrow{\mathbb{P}} 0 \iff Z_n \xrightarrow{\mathbb{P}} 0.$$

Démonstration : On a facilement $Z_n \leq Y_n \leq 2Z_n$, ce qui justifie la 2ème équivalence.

Puis si $X_n \rightarrow 0$ ps, alors X_n est ps une suite de Cauchy donc $Y_n \rightarrow 0$ ps et a fortiori en probabilité.

Supposons maintenant que $Z_n \xrightarrow{\mathbb{P}} 0$. On a

$$\{X_n \text{ converge}\} = \{Z_n \rightarrow 0\} = \bigcap_p \{\lim_n Z_n \leq 1/p\} = \bigcap_p \bigcup_n \{Z_n \leq 1/p\}$$

car $(Z_n)_n$ est décroissante. Or $\{Z_n \leq 1/p\}_n$ est une suite croissante d'évènements de proba qui tend vers 1 pour tout p car $Z_n \xrightarrow{\mathbb{P}} 0$. On a donc pour tout p , $\bigcup_n \{Z_n \leq 1/p\}$ de proba 1 et ça reste donc encore vrai pour l'intersection \bigcap_p . \square

Théorème 6.2.1 (Paul Lévy) Soit $(X_n)_n$ une suite de vai. On a

$$S_n \text{ converge ps} \iff S_n \text{ converge en proba.}$$

Démonstration : On suppose que S_n converge en proba. D'après le critère précédent, il suffit de voir que $Z_n = \sup_{p \geq n} |S_p - S_n|$ converge vers 0 en proba pour établir que S_n converge ps. Notons $\tau_n^\varepsilon = \inf(k \geq n : |S_k - S_n| > \varepsilon)$. On a

$$\mathbb{P}(Z_n > \varepsilon) = \mathbb{P}(\tau_n^\varepsilon < +\infty).$$

Noter que $\{\tau_n^\varepsilon = k\} \in \sigma(X_{n+1}, \dots, X_k)$ car ne concerne que les v.a. X_{n+1}, \dots, X_k . On a ensuite

$$\{|S_k - S_m| < \varepsilon/2, \tau_n^\varepsilon = k\} \subset \{|S_n - S_m| > \varepsilon/2, \tau_n^\varepsilon = k\}.$$

En effet, quand $\tau_n^\varepsilon = k$, on a $|S_n - S_m| + |S_m - S_k| \geq |S_n - S_k| > \varepsilon$ et donc

$$|S_m - S_n| > \varepsilon - |S_m - S_k| > \varepsilon/2$$

quand on a en plus $|S_m - S_k| < \varepsilon/2$.

On a donc pour $m > n$,

$$\begin{aligned} \mathbb{P}(|S_m - S_n| > \varepsilon/2) &\geq \sum_{k=m+1}^n \mathbb{P}(|S_m - S_n| > \varepsilon/2, \tau_n^\varepsilon = k) \\ &\geq \sum_{k=m+1}^n \mathbb{P}(|S_m - S_k| < \varepsilon/2, \tau_n^\varepsilon = k) \\ &\geq \sum_{k=m+1}^n \mathbb{P}(|S_m - S_k| < \varepsilon/2) \mathbb{P}(\tau_n^\varepsilon = k) \end{aligned}$$

où la dernière ligne est due à l'indépendance de $\{|S_m - S_k| < \varepsilon/2\} \in \sigma(X_{k+1}, \dots, X_m)$ et de $\{\tau_n^\varepsilon = k\} \in \sigma(X_{n+1}, \dots, X_k)$.

Comme par hypothèse, S_n converge en probabilité, avec $\alpha_n = \sup_{m \geq n} \mathbb{P}(|S_m - S_n| \geq \varepsilon/2)$, on a

$$\lim_{n \rightarrow +\infty} \alpha_n = 0.$$

On a aussi pour $k \geq n$, $\mathbb{P}(|S_m - S_k| < \varepsilon/2) \geq 1 - \alpha_n$, on en déduit :

$$\alpha_n \geq \sum_{k=n+1}^m (1 - \alpha_n) \mathbb{P}(\tau_n^\varepsilon = k)$$

$$\frac{\alpha_n}{1 - \alpha_n} \geq \sum_{k=n+1}^m \mathbb{P}(\tau_n^\varepsilon = k) = \mathbb{P}(\tau_n^\varepsilon \leq m).$$

Comme c'est vrai pour tout $m \geq n$, on déduit $\mathbb{P}(\tau_n^\varepsilon < +\infty) \leq \frac{\alpha_n}{1 - \alpha_n}$. Comme $\lim_{n \rightarrow +\infty} \alpha_n = 0$, on a bien $\lim_{n \rightarrow +\infty} \mathbb{P}(\tau_n^\varepsilon < +\infty) = 0$, ce qui (comme expliqué précédemment) permet de conclure. \square

6.2.2 Convergence ps et L^2

Lemme 6.2.1 (Kolmogorov) Soit $(X_n)_n$ une suite de v.a. centrées, de carrés intégrables alors

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > \alpha\right) \leq \frac{\mathbb{E}[S_n^2]}{\alpha^2}.$$

Cette inégalité est meilleure que celle de Tchebychev appliquée à S_n qui donne

$$\mathbb{P}(|S_n| > \alpha) \leq \frac{\mathbb{E}[S_n^2]}{\alpha^2}.$$

Démonstration : Posons $\tau = \inf(k \geq 1 | |S_k| \geq \alpha)$. τ est une variable aléatoire.

On a $\mathbb{P}(\tau \leq n) = \sum_{k=1}^n \mathbb{P}(\tau = k)$ et $\mathbb{E}[S_n^2] \geq \sum_{k=1}^n \mathbb{E}[S_n^2; \tau = k]$.

Mais $\{\tau = k\} \in \sigma(X_1, \dots, X_k)$ et $S_n - S_k = \sum_{i=k+1}^n X_i$ ne dépend que de X_1, \dots, X_n donc est mesurable pour $\sigma(X_{k+1}, \dots, X_n)$.

On a donc $(S_n - S_k) \perp \{\tau = k\}$. Mais alors

$$\begin{aligned} \mathbb{E}[S_n^2; \tau = k] &= \mathbb{E}[(S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2; \tau = k] \\ &\geq \mathbb{E}[S_k^2; \tau = k] + 2\mathbb{E}[S_n - S_k]\mathbb{E}[S_k; \tau = k] + 0 \\ &\geq \mathbb{E}[S_k^2; \tau = k] \\ \mathbb{E}[S_n^2] &\geq \sum_{k=1}^n \mathbb{E}[S_k^2; \tau = k] \geq \sum_{k=1}^n \mathbb{E}[\alpha^2; \tau = k] \geq \alpha^2 \sum_{k=1}^n \mathbb{P}(\tau = k) \\ &\geq \alpha^2 \sum_{k=1}^n \mathbb{P}(\tau = k) \geq \alpha^2 \mathbb{P}(\tau \leq n) = \alpha^2 \mathbb{P}(\max_{1 \leq k \leq n} |S_k| \geq \alpha). \end{aligned}$$

\square

Théorème 6.2.2 Soit $(X_n)_n$ une suite de v.a. centrées L^2 . Si $\sum_n \mathbb{E}[X_n^2] < +\infty$ alors S_n converge ps.

Démonstration : Comme les v.a. sont centrées, on a

$$E\left[\left(\sum_{k=1}^n X_k\right)^2\right] = \sum_{k=1}^n \mathbb{E}[X_k^2] + \sum_{k \neq l} \mathbb{E}[X_k X_l] = \sum_{k=1}^n \mathbb{E}[X_k^2]$$

car $\mathbb{E}[X_k X_l] = \mathbb{E}[X_k] \mathbb{E}[X_l] = 0$. On a donc d'après le lemme de Kolmogorov

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_{m+k} - S_m| \geq \alpha\right) \leq \frac{\mathbb{E}[(S_{m+n} - S_m)^2]}{\alpha^2} = \frac{\mathbb{E}[(\sum_{k=m+1}^n X_k)^2]}{\alpha^2} = \sum_{k=m+1}^n \frac{\mathbb{E}[X_k^2]}{\alpha^2}.$$

On a donc (avec $n \rightarrow +\infty$)

$$\mathbb{P}\left(\sup_{k \geq m} |S_k - S_m| \geq \alpha\right) \leq \sum_{k=m+1}^{+\infty} \frac{\mathbb{E}[X_k^2]}{\alpha^2} \rightarrow 0, \quad m \rightarrow +\infty.$$

On a montré que $\sup_{k \geq m} |S_k - S_m| \xrightarrow{\mathbb{P}} 0$, ce qui conclut par le critère de Cauchy vu. \square

Application. Soit $\epsilon_n = \pm 1$ avec des proba $1/2$ (ie ϵ_n est de loi de Bernoulli $b(1/2)$). On considère la somme $\sum_n \frac{\epsilon_n}{n^\alpha}$. Alors si $\alpha > 1/2$, elle converge ps.

Le résultat suivant donne des conditions nécessaires et des conditions suffisantes pour la convergence d'une série $\sum_n X_n$ en fonction des distributions individuelles des X_n . On note $X_n^c = X_n \mathbf{1}_{|X_n| \leq c}$ la troncature de X_n au seuil $c > 0$.

Théorème 6.2.3 (trois séries) Soit $(X_n)_n$ une suite de vai. On considère les trois séries

$$\sum_n \mathbb{P}(|X_n| > c), \quad \sum_n \mathbb{E}[X_n^c], \quad \sum_n \text{Var}(X_n^c). \quad (6.1)$$

Si les 3 séries de (6.1) convergent pour tout $c > 0$ alors $\sum_n X_n$ converge ps.

Si $\sum_n X_n$ converge ps alors il existe $c > 0$ tel que les 3 séries de (6.1) convergent.

Le sens nécessaire utilise le TCL (théorème 7.3.1).

Démonstration :

• **Sens suffisant.** On suppose que les 3 séries de (6.1) convergent.

Posons alors $m_n^c = \mathbb{E}[X_n^c]$. D'après le théorème 6.2.2, comme $\sum_n \text{Var}(X_n^c) < +\infty$, $\sum_n (X_n^c - m_n^c)$ converge ps. Comme $\sum_n m_n^c$ converge aussi, $\sum_n X_n^c$ converge ps. Mais d'après le lemme de Borel-Cantelli $\mathbb{P}(\limsup_n \{X_n \neq X_n^c\}) = 0$, en effet

$$\sum_n \mathbb{P}(X_n \neq X_n^c) = \sum_n \mathbb{P}(|X_n| > c) < +\infty.$$

On a donc ps $\exists k, \forall n > k$ tel que $X_n = X_n^c$. Autrement dit ps $X_n = X_n^c$ pour n assez grand et donc $\sum_n X_n$ converge aussi ps.

• **Sens nécessaire.** Ce sens utilise le TCL qui sera vu seulement plus tard.

On suppose maintenant que $\sum_n X_n$ converge ps. On fixe $c > 0$. Comme $X_n \rightarrow 0$ avec proba 1, on a $\sum_n \mathbb{P}(|X_n| > c) < +\infty$ (sinon, on a une contradiction par le deuxième lemme de Borel-Cantelli).

On a aussi $X_n = X_n^c$ pour n assez grand et donc $\sum_n X_n^c$ converge aussi ps. Notons

$$S_n^c = \sum_{k=1}^n X_k^c, \quad M_n^c = \mathbb{E}[S_n^c], \quad s_n^c = \sqrt{\text{Var}(S_n^c)}.$$

Par l'absurde, si $s_n^c \rightarrow +\infty$ alors comme $X_n^c - m_n^c$ sont uniformément bornées, par le TCL

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(x < \frac{S_n^c - M_n^c}{s_n^c} \leq y \right) = \frac{1}{\sqrt{2\pi}} \int_x^y e^{-t^2/2} dt \quad (6.2)$$

Comme $\sum_n X_n^c$ converge ps, $s_n^c \rightarrow +\infty$ entraîne $S_n^c/s_n^c \rightarrow 0$ ps et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|S_n^c/s_n^c| \geq \varepsilon) = 0. \quad (6.3)$$

On aboutit alors à une contradiction. En effet

$$\mathbb{P} \left(x \leq \frac{S_n^c - M_n^c}{s_n^c} \leq y, \left| \frac{S_n^c}{s_n^c} \right| < \varepsilon \right) \geq \mathbb{P} \left(x \leq \frac{S_n^c - M_n^c}{s_n^c} \leq y \right) - \mathbb{P}(|S_n^c/s_n^c| \geq \varepsilon) > 0$$

pour n assez grand, compte tenu de (6.2) et de (6.3) quand $x < y$. Mais alors, pour tout $x < y$, on a

$$x - \varepsilon < -M_n^c/s_n^c < y + \varepsilon.$$

Comme cet encadrement est déterministe, il ne peut être vrai simultanément pour $(x - \varepsilon, y + \varepsilon) = (-1, 0)$ et $(x - \varepsilon, y + \varepsilon) = (0, 1)$.

Finalement, s_n^c ne tend pas vers $+\infty$ et on a bien la convergence de $\sum_n \text{Var}(X_n^c)$.

Maintenant, en appliquant le théorème 6.2.2, on a $\sum_n (X_n^c - m_n^c)$ converge ps. Comme $\sum_n X_n^c$ converge ps, on a aussi $\sum_n \mathbb{E}[X_n^c]$ converge ps.

6.3 Loi des grands nombres (LGN)

6.3.1 LGN L^2

Lemme 6.3.1 (Kronecker) Soit $\sum_n x_n$ une série convergente dans un espace vectoriel normé E ($x_n \in E$). Soient $(a_n)_{n \geq 0}$ une suite de réels positive qui converge en croissant vers $+\infty$. Alors

$$\frac{1}{a_n} \sum_{k=1}^n a_k x_k \rightarrow 0, \quad n \rightarrow +\infty.$$

Démonstration : Il s'agit de faire une transformation d'Abel en écrivant $x_k = S_k - S_{k-1}$:

$$\begin{aligned} \sum_{k=1}^n a_k x_k &= \sum_{k=1}^n a_k (S_k - S_{k-1}) = \sum_{k=1}^{n-1} (a_k - a_{k-1}) S_k + a_n S_n \\ \frac{1}{a_n} \sum_{k=1}^n a_k x_k &= S_n - \sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} S_k. \end{aligned} \quad (6.4)$$

Or d'après le théorème de Césaro pour les suites : comme S_k converge vers S et $\sum_{k=1}^n \frac{a_{k+1} - a_k}{a_n} = 1$, on a $\sum_{k=1}^{n-1} \frac{a_{k+1} - a_k}{a_n} S_k \rightarrow S$ quand $n \rightarrow +\infty$ ce qui conclut le lemme quand on passe à la limite dans (6.4). \square

Théorème 6.3.1 Soient $(X_n)_n$ une suite de v.a. centrées L^2 et $a_n > 0$ des réels qui convergent en croissant vers $+\infty$. Alors si $\sum_{n=1}^{+\infty} \frac{\mathbb{E}[X_n^2]}{a_n^2} < +\infty$, on a $\frac{S_n}{a_n} \xrightarrow{ps, L^2} 0$.

Corollaire 6.3.1 Si les v.a. X_n sont non centrées, alors $\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n^2)}{a_n^2} < +\infty$ implique $\frac{S_n - \mathbb{E}[S_n]}{a_n} \xrightarrow{ps, L^2} 0$.

Démonstration : On montre d'abord la convergence L^2 : On a

$$\frac{\mathbb{E}[S_n^2]}{a_n^2} = \frac{1}{a_n^2} \sum_{k=1}^n \frac{\mathbb{E}[X_k^2] a_k^2}{a_k^2}.$$

Comme $\sum_{k=1}^n \frac{\mathbb{E}[X_k^2]}{a_k^2}$ converge, le lemme de Kronecker s'applique avec $a_n^2 \rightarrow +\infty$ et donne $\mathbb{E}[S_n^2]/a_n^2 \rightarrow 0$, d'où la convergence L^2 de S_n/a_n vers 0.

Ensuite, pour la convergence ps : $\sum_k \frac{X_k}{a_k}$ converge ps d'après le théorème précédent. Le lemme de Kronecker s'applique encore et donne $\sum_{k \leq n} X_k/a_n \rightarrow 0$ ps. \square

Corollaire 6.3.2 Etant donnée $(X_n)_n$ une suite de v.a. centrées L^2 , de variances bornées $\mathbb{E}[X_n^2] \leq c$ (par exemple avec les X_n toutes de même loi), on a

- 1) $\sum_n \frac{1}{a_n^2} < +\infty \implies \frac{S_n}{a_n} \rightarrow 0$ L^2 et ps.
- 2) Dès que $\alpha > 1/2$, on a $S_n/n^\alpha \rightarrow 0$ L^2 et ps.
- 3) $S_n/n \rightarrow 0$ L^2 et ps.
- 4) Pour une suite $(X_n)_n$ de v.a. L^2 , on a $S_n/n \rightarrow \mathbb{E}[X_1]$ L^2 et ps.

6.3.2 LGN L^1

On pose dans cette section $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}$, les v.a. tronquées et $T_n = \sum_{k=1}^n Y_k$.

Lemme 6.3.2 (Troncature) Soit $(X_n)_n$ des v.a. L^1 de même loi. On a

- 1) $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[X_1]$ et $\mathbb{E}[\frac{T_n}{n}] \rightarrow \mathbb{E}[X_1]$.
- 2) $\sum \frac{\text{Var} Y_n}{n^2} < +\infty$ et $\sum \frac{\mathbb{E}[Y_n^2]}{n^2} < +\infty$.
- 3) $\mathbb{P}(\lim_n \{Y_n = X_n\}) = 1$.

Démonstration :

1) $\mathbb{E}[Y_n] = \mathbb{E}[X_n; |X_n| \leq n] = \mathbb{E}[X_1; |X_1| \leq n] \rightarrow \mathbb{E}[X_1]$ par le théorème de convergence dominée. Puis le théorème de Césaro implique $\mathbb{E}[\frac{T_n}{n}] \rightarrow \mathbb{E}[X_1]$.

2) Comme les v.a. ont toutes la même loi, on a $\mathbb{E}[Y_n^2] = \mathbb{E}[X_1^2; |X_1| \leq n] = \sum_{k=1}^n \mathbb{E}[X_1^2; k-1 < |X_1| \leq k]$, puis

$$\begin{aligned} \sum_n \frac{\mathbb{E}[Y_n^2]}{n^2} &= \sum_n \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}[X_1^2; k-1 < |X_1| \leq k] \\ &= \sum_{k \geq 1} \mathbb{E}[X_1^2; k-1 < |X_1| \leq k] \sum_{n \geq k} \frac{1}{n^2} \\ &\leq \sum_{k \geq 1} \mathbb{E}[X_1^2; k-1 < |X_1| \leq k] \times \frac{2}{k} \\ &\leq \sum_{k \geq 1} 2\mathbb{E}[|X_1|; k-1 < |X_1| \leq k] = 2\mathbb{E}[|X_1|] < +\infty. \end{aligned}$$

Puis comme $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[X_1]$, $\sum_n \frac{\mathbb{E}[Y_n]^2}{n^2}$ est de même nature que $\sum_n \frac{\mathbb{E}[X_1]^2}{n^2}$ donc convergente.

3) $\overline{\lim}_n \{Y_n \neq X_n\} = \overline{\lim}_n \{|X_n| \geq n\} = \overline{\lim}_n \{|X_1| \geq n\}$. Le lemme de Borel-Cantelli conclura si on montre que $\sum_n \mathbb{P}(|X_1| \geq n) < +\infty$. Or $\sum_n \mathbb{P}(|X_1| \geq n) \simeq \mathbb{E}[|X_1|] < +\infty$. \square

Théorème 6.3.2 (LGN L^1) Soit $(X_n)_n$ une suite de v.a. iid L^1 . Alors

$$\frac{S_n}{n} \xrightarrow{ps, L^1} \mathbb{E}[X_1].$$

Lemme 6.3.3 (Scheffé) Soient $X_n \rightarrow X$ ps. Alors

$$\mathbb{E}[|X_n|] \rightarrow \mathbb{E}[|X|] \quad \text{ssi} \quad \mathbb{E}[|X_n - X|] \rightarrow 0 \quad \text{ie } X_n \rightarrow X \text{ dans } L^1.$$

Démonstration : cf poly de Calcul Intégral.

Démonstration : [LGN- L^1] Les v.a. Y_n sont bornées donc sont dans L^2 . D'après le 2) du lemme de troncature, on peut appliquer la LGN L^2 , on a alors $\frac{T_n - \mathbb{E}[T_n]}{n} \rightarrow 0$ ps et L^2 . Puis d'après le 1) du lemme de troncature, on a $\mathbb{E}[\frac{T_n}{n}] \rightarrow \mathbb{E}[X_1]$ donc $\frac{T_n}{n} \rightarrow \mathbb{E}[X_1]$ dans L^2 et ps.

Par ailleurs d'après le 3), ps il existe $n_0(\omega)$ tel que pour $n \geq n_0$, on a $X_n(\omega) = Y_n(\omega)$. On a alors

$$S_n = T_n + \sum_{k=1}^{n_0} (X_k - Y_k), \quad \frac{S_n}{n} = \underbrace{\frac{T_n}{n}}_{\rightarrow \mathbb{E}[X_1]} + \underbrace{\frac{S_{n_0} - T_{n_0}}{n}}_{\rightarrow 0, n \rightarrow +\infty}.$$

On a donc déjà la convergence ps de S_n/n vers $\mathbb{E}[X_1]$.

Écrivons maintenant $X_n = X_n^+ - X_n^-$ et $S_n = \sum_{k=1}^n X_k^+ - \sum_{k=1}^n X_k^- = S_n^1 - S_n^2$. Les v.a. X_n^+ sont indépendantes et ont même loi. On a $S_n^1 \geq 0$ et $\frac{S_n^1}{n} \rightarrow \mathbb{E}[X_1^+]$ ps puis par convergence dominée $\mathbb{E}[\frac{S_n^1}{n}] \rightarrow \mathbb{E}[X_1^+]$.

Le lemme de Scheffé s'applique et donne $\frac{S_n^1}{n} \rightarrow \mathbb{E}[X_1^+]$ dans L^1 .

On fait de même pour $\frac{S_n^2}{n}$, ce qui permet de conclure la LGN :

$$\frac{S_n}{n} = \frac{S_n^1}{n} - \frac{S_n^2}{n} \rightarrow \mathbb{E}[X_1^+] - \mathbb{E}[X_1^-] = \mathbb{E}[X_1].$$

□

Corollaire 6.3.3 Soit $(X_n)_n$ une suite de v.a.i.d. Alors

- 1) Si $\mathbb{E}[|X_1|] = +\infty$ alors $\frac{S_n}{n}$ diverge ps.
- 2) Si $\mathbb{E}[X_1^+] = +\infty$ et $\mathbb{E}[X_1^-] < +\infty$ alors $\frac{S_n}{n} \rightarrow +\infty$ ps.

Démonstration : Soit pour $a > 0$, $\tilde{X}_n = X_n \wedge a$. Comme on a $\mathbb{E}[|\tilde{X}_n|] < +\infty$, on a $\frac{\tilde{S}_n}{n} \xrightarrow{ps} \mathbb{E}[\tilde{X}_1]$. Or $S_n \geq \tilde{S}_n$ et

$$\mathbb{E}[\tilde{X}_1] \geq \underbrace{\mathbb{E}[X_1^-]}_{< +\infty} + \underbrace{\mathbb{E}[X_1^+; 0 \leq X_1 \leq a]}_{\rightarrow +\infty, a \rightarrow +\infty}.$$

On a donc $\frac{S_n}{n} \rightarrow +\infty$ ps. □

Théorème 6.3.3 Soit $(X_n)_n$ une suite de v.a. de même loi intégrable et indépendantes 2 à 2. Alors $\frac{S_n}{n} \xrightarrow{ps, L^1} \mathbb{E}[X_1]$.

Démonstration : (idée) Passer par $\frac{S_{2^n}}{2^n}$ puis par $\frac{S_{[a^n]}}{[a^n]}$ pour $a > 1$.

6.4 Applications

6.4.1 Estimateurs

Définition 6.4.1 Étant donnée X une va, on appelle échantillon indépendant de la loi de X la suite $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ où la suite $(X_i)_i$ est iid de même loi que X . On définit alors

- la moyenne empirique $M_n = \frac{1}{n} \sum_{k=1}^n X_k$.
- la variance empirique $V_n = \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)^2$.

Proposition 6.4.1 M_n et V_n sont des estimateurs consistants de $\mathbb{E}[X]$ et $\text{Var}(X)$:

$$\begin{aligned} M_n &\rightarrow \mathbb{E}[X] \text{ ps} & V_n &\rightarrow \text{Var}(X) \text{ ps} \\ \mathbb{E}[M_n] &= \mathbb{E}[X] & \mathbb{E}[V_n] &= \text{Var}(X). \end{aligned}$$

Démonstration : Pour la moyenne empirique M_n : c'est clair.

Pour la variance empirique, on a

$$\begin{aligned} V_n &= \frac{1}{n-1} \left(\sum_{k=1}^n X_k^2 - 2 \sum_{k=1}^n M_n X_k + n M_n^2 \right) = \frac{1}{n-1} \left(\sum_{k=1}^n X_k^2 - 2n M_n^2 + n M_n^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{k=1}^n X_k^2 - M_n^2 \right) \simeq \frac{1}{n-1} (n \mathbb{E}[X^2] - n \mathbb{E}[X]^2) \\ &\rightarrow \text{Var}(X). \end{aligned}$$

$$\text{Puis } \mathbb{E}[V_n] = \frac{1}{n-1} (n \mathbb{E}[X^2] - n \mathbb{E}[M_n^2]).$$

Or $\mathbb{E}[M_n^2] = \frac{1}{n^2} \mathbb{E}[(\sum_{k=1}^n X_k)^2] = \frac{1}{n^2} (n \mathbb{E}[X^2] + n(n-1) \mathbb{E}[X]^2)$ car les v.a. sont iid.

On a donc $\mathbb{E}[V_n] = \frac{1}{n-1} (n \mathbb{E}[X^2] - (\mathbb{E}[X^2] - (n-1) \mathbb{E}[X]^2)) = \text{Var}(X)$. \square

6.4.2 Méthode de Monte Carlo

Soit $f : B \rightarrow \mathbb{R}^2$ et $(X_n)_n$ une suite de v.a. uniformes sur B . On a

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{ps} \frac{1}{\lambda(B)} \int_B f d\lambda, \quad n \rightarrow +\infty.$$

En effet, $\mathbb{E}[f(X_1)] = \int f(x) \mathbb{P}_{X_1}(dx) = \int f(x) \frac{\mathbf{1}_B(x) dx}{\lambda(B)}$, puis on applique la LGN aux v.a. $f(X_i)$.

La méthode de Monte-Carlo permet donc de faire un calcul approché d'intégral en utilisant la LGN.

Application : l'aiguille de Buffon

On jette des aiguilles de même taille l sur un parquet composé de lattes d'épaisseur L . On peut modéliser ce parquet par un plan où sont dessinées des droites parallèles distantes de L . On compte le nombre d'intersection aiguilles/droites. Quand on lance N aiguilles, le nombre d'intersection est équivalent à $\frac{2l}{\pi L} N$ quand N est grand.

Notons X_k le nombre de droites intersectées par l'aiguille n° k . Comme toutes les aiguilles sont identiques et lancées de la même façon, les v.a. X_k sont iid. On cherche donc $\sum_{k=1}^N X_k$. D'après la LGN, si elle s'applique, on a $\sum_{k=1}^N X_k \simeq N \mathbb{E}[X_1]$.

On suppose le problème isotrope (indépendant de la direction) et invariant par translation (le parquet est infini).

On repère une aiguille par la position R de son milieu et par sa direction θ . Compte tenu des hypothèses, R est une v.a. uniforme sur $[0, L/2]$ et θ est une v.a. uniforme sur $[-\pi/2, \pi/2]$.

On suppose $l \leq L$, si bien que $X_k = 0$ ou 1 (et donc X_k est bornée et la LGN s'applique sans problème).

Il y a intersection entre une aiguille et une droite si $\theta \in [\theta_1(R), \theta_1(R)]$ où $\cos \theta_1(R) = \frac{R}{(l/2)}$ pour $0 \leq R \leq l/2$.

On a $\{X = 1\} = \{\text{intersection}\} = \{\theta \in [-\theta_1(R), \theta_1(R)]\}$ et donc

$$\mathbb{E}[X] = \mathbb{P}(\text{intersection}) = \int_0^{l/2} \frac{2\theta_1(R)}{\pi} \frac{dR}{(L/2)}$$

Par ailleurs, on peut aussi caractériser l'intersection par des conditions sur R , à θ fixé : il faut $0 \leq R \leq (l/2) \cos \theta$ lorsque $\theta \in [-\pi/2, \pi/2]$.

On a donc aussi

$$\mathbb{E}[X] = \mathbb{P}(\text{intersection}) = \int_{-\pi/2}^{\pi/2} \frac{(l/2) \cos \theta}{(L/2)} \frac{d\theta}{\pi} = \frac{l}{\pi L} \int_{-\pi/2}^{\pi/2} \sin \theta d\theta = \frac{2l}{\pi L}.$$

Si $l > L$, on généralise en considérant une courbe rectifiable de longueur l : si C est une réunion de segments $\cup_i S_i$, on a

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[\#\{\text{intersection de } C\}] = \sum_i \mathbb{E}[\#\{\text{intersection de } S_i\}] = \sum_i \mathbb{E}[X_i] \\ &= \sum_i \frac{2l(S_i)}{\pi L} = \frac{2l(C)}{\pi L}. \end{aligned}$$

On fait de même pour une courbe rectifiable C , limite de réunion de segments, par convergence dominée on montre alors

$$\mathbb{E}[X(C)] = \frac{2l(C)}{\pi L}.$$

Chapitre 7

Fonction caractéristique

7.1 Définition et propriétés

Définition 7.1.1 Soit X une variable aléatoire définie sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle fonction caractéristique de X la fonction de \mathbb{R} dans \mathbb{C}

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX} d\mathbb{P}. \quad (7.1)$$

Remarque 7.1.1 • ϕ_X est toujours bien définie car $|e^{itX}| = 1$ est \mathbb{P} -intégrable.

- Si X est de densité f , la fonction caractéristique s'écrit

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx = \mathcal{F}(f)\left(\frac{t}{2\pi}\right).$$

À quelques notations près, il s'agit de la transformée de Fourier $\mathcal{F}(f)$ de la densité f de la variable X .

- Si X est une variable discrète, la fonction caractéristique peut se voir encore comme une transformée de Fourier mais il faut alors introduire une transformée de Fourier par rapport à une mesure discrète.

- De façon générale

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itx} d\mathbb{P}_X(x).$$

La fonction caractéristique ϕ_X est donc la transformée de Fourier de la mesure \mathbb{P}_X , loi de la v.a. X .

Proposition 7.1.1 La fonction caractéristique caractérise la loi : si $\phi_X = \phi_Y$ alors les variables X et Y ont même loi et réciproquement.

Démonstration : La justification vient du résultat d'inversion des transformées de Fourier.

Comme $|\phi_X(t)| \leq \mathbb{E}[|e^{itX}|] = 1$, Φ_X est bornée donc intégrable par rapport à \mathbb{P} . En interprétant $\phi_X(t)$ comme la transformée de Fourier de la loi \mathbb{P}_X , le théorème d'inversion des transformées de Fourier s'applique et donne

$$\mathbb{P}_X = \mathcal{F}(\phi_X).$$

Si $\phi_X = \phi_Y$, on en déduit donc l'égalité des lois $\mathbb{P}_X = \mathbb{P}_Y$. □

Exemples :

- Pour des variables à densité :

Variables	Densité f	Intervalle	Fonction caractéristique ϕ_X
Loi normale standard	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$-\infty < x < +\infty$	$e^{-t^2/2}$
Loi normale	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-m)^2/(2\sigma^2)}$	$-\infty < x < +\infty$	$e^{imt-\sigma^2t^2/2}$
Loi uniforme	1	$0 \leq x \leq 1$	$\frac{e^{it}-1}{it}$
Loi exponentielle	e^{-x}	$0 \leq x < +\infty$	$\frac{1}{1-it}$
Loi de Cauchy	$\frac{1}{\pi(1+x^2)}$	$-\infty < x < +\infty$	$e^{- t }$

- Pour des variables discrètes X de domaine $X(\Omega) = \{x_1, \dots, x_n, \dots\}$ de masse ponctuelle $p_k = \mathbb{P}(X = x_k)$ en x_k , on a

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \sum_{k=1}^{+\infty} e^{itx_k} p_k$$

Variables	Domaine $X(\omega)$	Probabilités ponctuelles p_k	Fonction caractéristique ϕ_X
Loi de Bernoulli	$\{0, 1\}$	$1 - p, p$	$1 - p + pe^{it}$
Loi binomiale	$\{0, \dots, n\}$	$C_n^k p^k (1 - p)^{n-k}$	$(1 - p + pe^{it})^n$
Loi de Poisson	\mathbb{N}	$e^{-\alpha} \frac{\alpha^k}{k!}$	$e^{\alpha(e^{it}-1)}$

La régularité de la fonction caractéristique ϕ_X est liée à l'intégrabilité de la v.a. X :

Proposition 7.1.2 • $\phi_X(t)$ est définie pour tout $t \in \mathbb{R}$.

- Si X a un moment d'ordre 1, alors ϕ_X est dérivable (et même C^1) avec

$$\phi'_X(t) = i\mathbb{E}[Xe^{itX}].$$

En particulier : $i\mathbb{E}[X] = \phi'_X(0)$.

- Plus généralement, si X a un moment d'ordre p alors ϕ_X est p fois dérivable (et même de classe C^p), avec

$$\phi_X^{(p)}(t) = i^p \mathbb{E}[X^p e^{itX}].$$

En particulier $i^p \mathbb{E}[X^p] = \phi_X^{(p)}(0)$.

Démonstration : Pour les résultats de dérivation, il suffit d'appliquer le théorème de convergence dominée pour dériver sous le signe E . La domination est à chaque fois garantie par l'existence des moments nécessaires :

$$|X^p e^{itX}| \leq |X|^p \quad \text{est } \mathbb{P}\text{-intégrable}$$

par existence du moment d'ordre p . □

Théorème 7.1.1 *Si deux variables aléatoires X et Y sont indépendantes alors*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Plus généralement, pour n variables aléatoires X_1, \dots, X_n indépendantes on a

$$\phi_{X_1+\dots+X_n}(t) = \phi_{X_1}(t) \dots \phi_{X_n}(t).$$

Démonstration : On a

$$\phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX} e^{itY}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}]$$

d'après la proposition 2.3.3. La généralisation au cas de n variables est immédiate. □

On définit aussi la fonction caractéristique d'un vecteur de la façon suivante :

Définition 7.1.2 *Soit (X, Y) un vecteur aléatoire, on définit sa fonction caractéristique comme la fonction de 2 variables, donc de \mathbb{R}^2 dans \mathbb{C} :*

$$\phi_{(X,Y)}(t, s) = \mathbb{E}[e^{i\langle (t,s), (X,Y) \rangle}] = \mathbb{E}[e^{i(tX+sY)}]$$

où $\langle (t, s), (x, y) \rangle = tx + sy$ est le produit scalaire euclidien de \mathbb{R}^2 .

Il s'agit de la transformée de Fourier de la mesure de \mathbb{R}^2 $\mathbb{P}_{(X,Y)}$, loi du vecteur (X, Y) . On a alors un critère d'indépendance de deux variables aléatoires :

Proposition 7.1.3 *Deux variables aléatoires X et Y sont indépendantes ssi*

$$\phi_{(X,Y)}(t, s) = \phi_X(t)\phi_Y(s).$$

Démonstration : On raisonne par équivalence :

$$\begin{aligned} X \perp Y &\Leftrightarrow d\mathbb{P}_{X,Y}(dx, dy) = d\mathbb{P}_X(dx)d\mathbb{P}_Y(dy) \\ &\Leftrightarrow \int_{\mathbb{R}^2} e^{i(tx+sy)} d\mathbb{P}_{X,Y}(dx, dy) = \int_{\mathbb{R}^2} e^{i(tx+sy)} d\mathbb{P}_X(dx)d\mathbb{P}_Y(dy) \\ &\Leftrightarrow \int_{\mathbb{R}^2} e^{i(tx+sy)} d\mathbb{P}_{X,Y}(dx, dy) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(dx) \int_{\mathbb{R}} e^{isy} d\mathbb{P}_Y(dy) \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \int_{\Omega} e^{i(tX+sY)} d\mathbb{P} = \int_{\Omega} e^{itX} d\mathbb{P} \int_{\Omega} e^{isY} d\mathbb{P} \\ &\Leftrightarrow \phi_{(X,Y)}(t, s) = \phi_X(t)\phi_Y(s) \end{aligned}$$

où on admet l'équivalence de la deuxième ligne (c'est la même que celle sur la caractérisation de la loi de X par ϕ_X) et où on a utilisé ensuite les théorèmes de Fubini, puis de transfert. \square

On généralise immédiatement au cas de n variables la définition de la fonction caractéristique du vecteur (X_1, \dots, X_n)

$$\phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \mathbb{E}[e^{i(t_1 X_1 + \dots + t_n X_n)}]$$

et le critère d'indépendance de n variables aléatoires.

On a maintenant le critère suivant pour la convergence en loi :

Théorème 7.1.2 (Paul Lévy) *La convergence en loi des variables aléatoires est équivalente à la convergence simple de leur fonction caractéristique :*

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{ssi} \quad \forall t \in \mathbb{R}, \phi_{X_n}(t) \rightarrow \phi_X(t).$$

Démonstration : • Dans le sens direct : on sait que la convergence en loi est équivalente à $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ pour toute fonction f continue bornée. Avec $f(x) = e^{itx}$, on a $\phi_{X_n}(t) \rightarrow \phi_X(t)$.

• Le sens indirect est plus difficile et admis. \square

Application. LGN faible L^1 : Soit $(X_i)_i$ une suite de v.a.i. de même loi avec un moment d'ordre 1 fini. Alors

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

Preuve. Notons $S_n = \sum_{i=1}^n X_i$, on a

$$\begin{aligned} \phi_{S_n/n}(t) &= \mathbb{E}[e^{itS_n/n}] = \mathbb{E}\left[\prod_{j=1}^n e^{itX_j/n}\right] = \prod_{j=1}^n \mathbb{E}[e^{itX_j/n}] = \mathbb{E}[e^{itX_1/n}]^n \\ &= \phi_{X_1}(t/n)^n = \left(\phi_{X_1}(0) + \phi'_{X_1}(0)\frac{t}{n} + \epsilon(1/n)\right)^n = \left(1 + i\mathbb{E}[X_1]\frac{t}{n} + \epsilon(1/n)\right)^n \\ &= \exp\left(n \ln\left(1 + i\mathbb{E}[X_1]\frac{t}{n} + \epsilon(1/n)\right)\right) = \exp\left(i\mathbb{E}[X_1]t + \epsilon(1)\right) \end{aligned}$$

où ϵ est une fonction de limite nulle en 0. On a donc

$$\lim_{n \rightarrow +\infty} \phi_{S_n/n}(t) = \exp(it\mathbb{E}[X_1]) = \phi_{\mathbb{E}[X_1]}(t).$$

D'après le théorème de Paul Lévy, on a donc $\frac{S_n}{n} \xrightarrow{\mathcal{L}} \mathbb{E}[X_1]$.

Comme $\mathbb{E}[X_1]$ est une constante, d'après la Prop. 5.4.2, on a aussi la convergence en proba

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1].$$

7.2 Variables et vecteurs gaussien

Variabes gaussiennes

Définition 7.2.1 Une v.a. X suit la loi normale standard $\mathcal{N}(0,1)$ si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

De façon générale, une v.a. X suit la loi normale $\mathcal{N}(m, \sigma^2)$ si elle admet pour densité

$$t \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Si $\sigma^2 = 0$, la loi est dégénérée, la v.a. X est constante égale à m . Sa loi est un dirac en m : $\mathbb{P}_X = \delta_m$.

Exercice. Normalisation de la loi normale $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.

Notons $I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx$ et montrons que $I^2 = 2\pi$. On a

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-x^2/2} dx \times \int_{-\infty}^{+\infty} e^{-y^2/2} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{-y^2/2} dx dy = \int \int_{\mathbb{R} \times \mathbb{R}} e^{-(x^2+y^2)/2} dx dy \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{+\infty} r e^{-r^2/2} dr = 2\pi \left[-e^{-r^2/2} \right]_0^{+\infty} = 2\pi \end{aligned}$$

où on a utilisé le théorème de Fubini à la 2ème ligne puis on a fait un changement de variables en polaires à la 3ème ligne.

Rappelons que :

Proposition 7.2.1 Une v.a. $X \simeq \mathcal{N}(m, \sigma^2)$ peut se voir comme la translatée et la dilatée d'une v.a. X_0 de loi normale standard $\mathcal{N}(0,1)$ par

$$X = m + \sigma X_0.$$

Autrement dit si $X \simeq \mathcal{N}(m, \sigma^2)$, on définit la variable centrée réduite $\tilde{X} = \frac{X - m}{\sigma}$. Elle suit la loi $\mathcal{N}(0, 1)$.

Rappelons encore que :

Proposition 7.2.2 Une v.a. X de loi $\mathcal{N}(m, \sigma^2)$ a pour

- *Espérance* : $\mathbb{E}[X] = m$.
- *Variance* : $\text{Var}(X) = \sigma^2$.
- *Fonction caractéristique* : $\phi_X(t) = \exp(imt - \sigma^2 t^2/2)$.

Démonstration : On se contente de montrer l'expression de la fonction caractéristique. Les calculs de l'espérance et de la variance ont déjà été faits dans les chapitres précédents. Comme X a même loi que $\sigma X_0 + m$, elle a aussi même fonction caractéristique :

$$\begin{aligned}\phi_X(t) &= \phi_{\sigma X_0 + m}(t) = \mathbb{E}[e^{it(\sigma X_0 + m)}] \\ &= \mathbb{E}[e^{it\sigma X_0} e^{itm}] = e^{itm} \phi_{X_0}(t\sigma).\end{aligned}$$

Il suffit donc de montrer que $\phi_{X_0}(t) = e^{-t^2/2}$. Or

$$\begin{aligned}\phi_{X_0}(t) &= \int_{-\infty}^{+\infty} e^{itx} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2 - 2itx}{2}\right) \frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{(x - it)^2 - (it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{(x - it)^2}{2}\right) e^{-t^2/2} \frac{dx}{\sqrt{2\pi}} \\ &= e^{-t^2/2} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x - it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \\ &= e^{-t^2/2}\end{aligned}$$

car avec le changement de variable (complexe) $y = x - it$, on a :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x - it)^2}{2}\right) \frac{dx}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{2}\right) \frac{dy}{\sqrt{2\pi}} = 1.$$

□

Proposition 7.2.3 Soient $X_1 \simeq \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \simeq \mathcal{N}(m_2, \sigma_2^2)$ indépendantes. Alors $X_1 + X_2 \simeq \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Démonstration : En effet, on sait que $X_1 + X_2$ suit une loi normale $\mathcal{N}(m, \sigma^2)$, puis que ses paramètres sont

$$m = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = m_1 + m_2$$

$$\sigma^2 = \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = \sigma_1^2 + \sigma_2^2.$$

□

Vecteur gaussien

Définition 7.2.2 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est gaussien ssi toutes les combinaisons linéaires de ces coordonnées $\langle a, X \rangle = a_1 X_1 + \dots + a_n X_n$ suit une loi gaussienne dans \mathbb{R} (pour tout $a = (a_1, \dots, a_n) \in \mathbb{R}^n$).

Définition 7.2.3 La matrice de covariance d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ est la matrice carrée symétrique, positive

$$K = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n}.$$

L'espérance de $X = (X_1, \dots, X_n)$ est le vecteur des espérances de ses marginales

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n]).$$

Si $\mathbb{E}[X] = 0$, le vecteur X est dit centré.

Fonction caractéristique gaussienne en dimension n

Si $X = (X_1, \dots, X_n)$ est un vecteur gaussien alors $\langle a, X \rangle = \sum_{i=1}^n a_i X_i$ suit une loi normale de paramètres

$$\begin{aligned} \mathbb{E}[\langle a, X \rangle] &= \mathbb{E}[a_1 X_1 + \dots + a_n X_n] = a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n] = \langle a, \mathbb{E}[X] \rangle \\ \text{Var}(\langle a, X \rangle) &= \text{Var}(a_1 X_1 + \dots + a_n X_n) = \sum_{i,j=1}^n a_i a_j \text{Cov}(X_i, X_j) = a^t \text{Cov}(X) a \end{aligned}$$

en utilisant des notations matricielles.

La v.a. gaussienne $\langle a, X \rangle$ suit donc la loi $\mathcal{N}(\langle a, \mathbb{E}[X] \rangle, a^t \text{Cov}(X) a)$, sa fonction caractéristique est donnée par

$$\phi_{\langle a, X \rangle}(x) = \exp \left(ix \langle a, \mathbb{E}[X] \rangle - \frac{1}{2} (a^t \text{Cov}(X) a) x^2 \right). \quad (7.2)$$

D'après les définitions des fonctions caractéristiques d'une variable aléatoire et d'un vecteur aléatoire

$$\phi_X(x) = \mathbb{E}[e^{i\langle x, X \rangle}] = \phi_{\langle x, X \rangle}(1).$$

On déduit de (7.2) :

Proposition 7.2.4 La fonction caractéristique d'un vecteur gaussien $X = (X_1, \dots, X_n)$ est donnée par

$$\begin{aligned} \phi_X(x) &= \exp \left(i \langle x, \mathbb{E}[X] \rangle - \frac{1}{2} (x^t \text{Cov}(X) x) \right) \\ &= \exp \left(i \langle x, \mathbb{E}[X] \rangle - \frac{1}{2} \langle x, \text{Cov}(X) x \rangle \right). \end{aligned} \quad (7.3)$$

Remarques :

- La loi d'un vecteur gaussien est connue dès qu'on a le vecteur moyenne $\mathbb{E}[X]$ et la matrice de covariance $\text{Cov}(X)$.
- On parle du vecteur gaussien standard en dimension n lorsque $\mathbb{E}[X] = 0$ et $\text{Cov}(X) = I_n$. Sa fonction caractéristique est alors

$$\phi_X(x) = \exp(-\langle x, x \rangle / 2) = \exp(-\|x\|^2 / 2).$$

- Pour un vecteur gaussien centré, $\mathbb{E}[X] = 0$ et on montre que

$$\langle x, \text{Cov}(X)x \rangle = \mathbb{E}[\langle x, X \rangle^2],$$

la fonction caractéristique devient donc

$$\phi_X(x) = \exp\left(-\frac{1}{2}\langle x, \text{Cov}(X)x \rangle\right) = \exp\left(-\frac{1}{2}\mathbb{E}[\langle x, X \rangle^2]\right).$$

- En prenant $x = (x_1, 0, \dots, 0)$, on a

$$\phi_X(x) = \phi_{X_1}(x_1) = \exp(i\mathbb{E}[X_1]x_1 - \text{Var}(X_1)x_1^2/2).$$

On retrouve que $X_1 \simeq \mathcal{N}(\mathbb{E}[X_1], \text{Var}(X_1))$. Plus généralement, pour tout $1 \leq i \leq n$, on a $X_i \simeq \mathcal{N}(\mathbb{E}[X_i], \text{Var}(X_i))$.

Comme pour les v.a. gaussiennes, on peut se ramener à un vecteur gaussien standard en centrant et en réduisant un vecteur gaussien quelconque non dégénéré. On a en effet

Proposition 7.2.5 *Soit $X \simeq \mathcal{N}(m, K)$ un vecteur gaussien non dégénéré de moyenne $m \in \mathbb{R}^n$ et de matrice de covariance K . Alors*

$$\sqrt{K}^{-1}(X - m) \simeq \mathcal{N}(0, I_n). \quad (7.4)$$

Démonstration : Comme le vecteur X est non dégénéré, sa matrice de covariance K est définie (càd inversible), symétrique. Il existe donc une matrice $A = \sqrt{K}$ inversible telle que $K = A^t A$. Il est donc légitime d'utiliser \sqrt{K}^{-1} dans (7.4).

On montre maintenant que $\tilde{X} = \sqrt{K}^{-1}(X - m)$ est gaussien, standard en calculant sa fonction caractéristique :

$$\begin{aligned} \phi_{\tilde{X}}(x) &= \mathbb{E}[\exp(i\langle x, \tilde{X} \rangle)] \\ &= \mathbb{E}[\exp(i\langle x, \sqrt{K}^{-1}(X - m) \rangle)] \\ &= \mathbb{E}[\exp(i\langle (\sqrt{K}^{-1})^t x, X - m \rangle)] \\ &= \mathbb{E}[\exp(i\langle (\sqrt{K}^{-1})^t x, X - m \rangle) e^{-i\langle (\sqrt{K}^{-1})^t x, m \rangle}] \\ &= \mathbb{E}[\exp(i\langle (\sqrt{K}^{-1})^t x, X \rangle)] \times e^{-i\langle (\sqrt{K}^{-1})^t x, m \rangle} \\ &= \phi_X((\sqrt{K}^{-1})^t x) \times e^{-i\langle (\sqrt{K}^{-1})^t x, m \rangle} \end{aligned}$$

$$\begin{aligned}
&= \exp \left(i \langle m, (\sqrt{K}^{-1})^t x \rangle - \frac{1}{2} \langle (\sqrt{K}^{-1})^t x, K (\sqrt{K}^{-1})^t x \rangle \right) \times e^{-i \langle (\sqrt{K}^{-1})^t x, m \rangle} \\
&= \exp \left(-\frac{1}{2} \langle (\sqrt{K}^{-1})^t x, K (\sqrt{K}^{-1})^t x \rangle \right) \\
&= \exp \left(-\frac{1}{2} \langle x, \sqrt{K}^{-1} K (\sqrt{K}^{-1})^t x \rangle \right) \\
&= \exp \left(-\frac{1}{2} \langle x, x \rangle \right) = \exp(-\|x\|^2/2).
\end{aligned}$$

□

Remarque : Une v.a. $X \simeq \mathcal{N}(m, K)$ avec K inversible peut se voir comme une translatée et dilatée du vecteur gaussien standard $X_0 \simeq \mathcal{N}(0, I_n)$:

$$X \simeq \sqrt{K} X_0 + m.$$

Indépendance de variables gaussiennes

Proposition 7.2.6 Soient (X, Y) un couple gaussien. Alors X et Y sont indépendantes ssi $\text{Cov}(X, Y) = 0$.

Démonstration : Le sens direct est vrai quelque soit la loi de X et de Y . Pour la réciproque, on sait que X et Y sont indépendantes si $\phi_{(X,Y)}(t_1, t_2) = \phi_X(t_1)\phi_Y(t_2)$. Or (X, Y) est un vecteur gaussien de matrice de covariance diagonale

$$\begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$$

car $\text{Cov}(X, Y) = \text{Cov}(Y, X) = 0$. On déduit de (7.3) que

$$\begin{aligned}
\phi_{(X,Y)}(t_1, t_2) &= \exp(im_1 t_1 + im_2 t_2 - \frac{1}{2}(t_1^2 \sigma_X^2 + t_2^2 \sigma_Y^2)) \\
&= \exp(im_1 t_1 - \frac{1}{2} t_1^2 \sigma_X^2) \times \exp(-im_2 t_2 - \frac{1}{2} t_2^2 \sigma_Y^2) \\
&= \phi_X(t_1) \phi_Y(t_2),
\end{aligned}$$

ce qui justifie l'indépendance. □

De la même façon, pour des vecteurs :

Proposition 7.2.7 Soit $(X_1, \dots, X_n, Y_1, \dots, Y_p)$ un vecteur gaussien de dimension $n + p$. Les deux vecteurs aléatoires $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_p)$ sont indépendants ssi les covariances $\text{Cov}(X_i, Y_j)$, $1 \leq i \leq n, 1 \leq j \leq p$ sont toutes nulles.

Densité gaussienne en dimension n

Soit $X \simeq \mathcal{N}(0, I_n)$ un vecteur gaussien standard en dimension n . Comme $\text{Cov}(X) = I_n$, les marginales X_1, \dots, X_n sont toutes indépendantes. La loi du vecteur $X = (X_1, \dots, X_n)$ est donc la loi produit de ses marginales

$$\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

En terme de densité, la densité de X est donnée par le produit tensoriel

$$\begin{aligned} f_X(x_1, \dots, x_n) &= f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n) \\ &= \left(\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \right) \times \dots \times \left(\frac{1}{\sqrt{2\pi}} e^{-x_n^2/2} \right) \\ &= \frac{1}{\sqrt{2\pi}^n} \exp(-(x_1^2 + \dots + x_n^2)/2). \end{aligned}$$

On peut énoncer :

Proposition 7.2.8 *La densité d'un vecteur gaussien standard en dimension n est*

$$f_X(x) = \frac{1}{\sqrt{2\pi}^n} \exp(-(x_1^2 + \dots + x_n^2)/2).$$

Pour passer au cas général d'un vecteur gaussien $X \simeq \mathcal{N}(m, K)$ non dégénéré (càd avec $K = \text{Cov}(X)$ inversible), on va utiliser le vecteur centré réduit donné en (7.4) : on a $X \simeq \sqrt{K}X_0 + m$ avec $X_0 \simeq \mathcal{N}(0, I_n)$.

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(\sqrt{K}X_0 + m \in A) \\ &= \mathbb{P}(X_0 \in \sqrt{K}^{-1}(A - m)) \\ &= \int_{\sqrt{K}^{-1}(A - m)} f(x) dx \\ &= \int_A f(\sqrt{K}^{-1}(y - m)) \frac{dy}{\det \sqrt{K}} \\ &= \int_A \frac{\exp(-\|\sqrt{K}^{-1}(y - m)\|^2/2)}{((2\pi)^n \det K)^{1/2}} dy \\ &= \int_A \frac{\exp(-\langle (x - m), K^{-1}(x - m) \rangle/2)}{((2\pi)^n \det K)^{1/2}} dx \end{aligned}$$

où on a fait à la 4ème ligne le changement de variable $y = \sqrt{K}x + m$. On a prouvé

Proposition 7.2.9 *La densité d'un vecteur gaussien $X \simeq \mathcal{N}(m, K)$ non dégénéré est*

$$f_X(x) = \frac{\exp(-\langle (x - m), K^{-1}(x - m) \rangle/2)}{((2\pi)^n \det K)^{1/2}}.$$

Vecteurs non gaussiens avec des variables marginales gaussiennes

On a déjà vu que si un vecteur $\bar{X} = (X_1, \dots, X_n)$ est gaussien alors ses marginales X_i le sont aussi, de même les combinaisons linéaires de ses marginales le sont.

La réciproque est fautive : Si des variables aléatoires sont gaussiennes alors le vecteur formé par ces variables n'est pas nécessairement gaussien.

En effet, prenons X une v.a. de loi $\mathcal{N}(0, 1)$ et Y de loi donnée, pour $a > 0$ fixé, par

$$Y = \begin{cases} X & \text{si } |X| \leq a, \\ -X & \text{si } |X| > a. \end{cases}$$

Alors Y est de loi $\mathcal{N}(0, 1)$ en effet comme la loi de X est symétrique : $\mathcal{L}(X) = \mathcal{L}(-X)$:

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}[e^{itY}] = \mathbb{E}[e^{itX} \mathbf{1}_{|X| \leq a}] + \mathbb{E}[e^{-itX} \mathbf{1}_{|X| > a}] \\ &= \mathbb{E}[e^{itX} \mathbf{1}_{|X| \leq a}] + \mathbb{E}[e^{itX} \mathbf{1}_{|-X| > a}] = \mathbb{E}[e^{itX} \mathbf{1}_{|X| \leq a}] + \mathbb{E}[e^{itX} \mathbf{1}_{|X| > a}] \\ &= \mathbb{E}[e^{itX} (\mathbf{1}_{|X| \leq a} + \mathbf{1}_{|X| > a})] \\ &= \mathbb{E}[e^{itX}] = e^{-t^2/2}. \end{aligned}$$

Puis, la variable $X + Y$ est donnée par

$$\begin{aligned} X + Y &= \begin{cases} X + X = 2X & \text{si } |X| \leq a \\ X - X = 0 & \text{si } |X| > a \end{cases} \\ &= 2X \mathbf{1}_{|X| \leq a}. \end{aligned}$$

La combinaison linéaire $X + Y$ ne suit donc pas une loi gaussienne. En effet, $X + Y$ a un atome en 0 :

$$\mathbb{P}(X + Y = 0) = \mathbb{P}(|X| > a) > 0.$$

Le couple aléatoire (X, Y) n'est donc pas gaussien, sinon on aurait $X + Y$ de loi gaussienne.

De plus, cet exemple montre aussi que dans la proposition 7.2.6, l'hypothèse (X, Y) gaussien est nécessaire et il ne suffit pas de supposer que X et Y sont des variables gaussiennes. En effet,

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY] = \mathbb{E}[X^2 \mathbf{1}_{|X| \leq a}] - \mathbb{E}[X^2 \mathbf{1}_{|X| > a}] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X^2 \mathbf{1}_{|X| > a}] - \mathbb{E}[X^2 \mathbf{1}_{|X| > a}] \\ &= 1 - 2\mathbb{E}[X^2 \mathbf{1}_{|X| > a}]. \end{aligned}$$

La fonction $u(a) = \mathbb{E}[X^2 \mathbf{1}_{|X| < a}]$ tend vers 0 en $+\infty$ par convergence dominée, est continue et vaut $\mathbb{E}[X^2] = 1$ en 0. Il existe donc a tel que $u(a) = 1/2$ c'est-à-dire $\text{Cov}(X, Y) = 0$.

Pourtant, X et Y sont non indépendantes sinon la loi du couple (X, Y) serait

$$\mathbb{P}_{(X, Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y = e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} \times e^{-y^2/2} \frac{dy}{\sqrt{2\pi}} = e^{-\frac{x^2+y^2}{2}} \frac{dxdy}{2\pi}.$$

qui est gaussienne, ce qui est faux. On a donc des variables gaussiennes X et Y non corrélées mais non indépendantes.

7.3 Théorème central limite (TCL)

Dans la suite i.i.d. signifiera indépendant(e)s et identiquement distribué(e)s, c'est à dire de même loi.

On déduit du théorème de Paul Lévy un résultat fondamental en probabilité et en statistiques : le théorème central limite (TCL) dû à Paul Lévy.

Théorème 7.3.1 (TCL) *Soit $(X_n)_n$ une suite de variables aléatoires réelles indépendantes et de même loi, d'espérance m et de variance finie σ^2 . Soit $S_n = X_1 + \dots + X_n$ la somme partielle. Alors quand $n \rightarrow +\infty$*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarque 7.3.1 • D'un certain point de vue ce résultat complète la loi des grands nombres : la LGN dit que la moyenne arithmétique

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n}$$

converge quand $n \rightarrow +\infty$ vers l'espérance $m = \mathbb{E}[X_1]$ (la moyenne probabiliste) en probabilité ou presque sûrement selon la version de la LGN. On sait donc que S_n est équivalent à nm . Le TCL indique (en un sens faible) que la vitesse de cette convergence est en $1/\sqrt{n}$ (sens faible car la convergence du TCL est une convergence assez faible : la convergence en loi).

• En plus dans le TCL, apparaît à la limite la loi $\mathcal{N}(0, 1)$ alors que les v.a. X_i sont de lois arbitraires : ce résultat justifie le rôle universel de la loi normale. Elle modélise les petites variations de n'importe quelle loi (avec un moment d'ordre 2) par rapport à sa moyenne.

Démonstration : Posons $Y_i = X_i - m$, si bien que les v.a. Y_i sont indépendantes de même loi avec $\mathbb{E}[Y_i] = 0$, $\text{Var}(Y_i) = \text{Var}(X_i)$. Notons $S'_n = Y_1 + \dots + Y_n$ et $Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{S'_n}{\sigma\sqrt{n}}$. On a

$$\begin{aligned} \phi_{Z_n}(t) &= \mathbb{E}[\exp\{it \frac{S'_n}{\sigma\sqrt{n}}\}] = \mathbb{E}[\exp\{i \frac{t}{\sigma\sqrt{n}} S'_n\}] = \phi_{S'_n}(\frac{t}{\sigma\sqrt{n}}) \\ &= \phi_{Y_1}(\frac{t}{\sigma\sqrt{n}}) \dots \phi_{Y_n}(\frac{t}{\sigma\sqrt{n}}) \\ &= \left(\phi_{Y_1}(\frac{t}{\sigma\sqrt{n}}) \right)^n \end{aligned}$$

en utilisant $\phi_{Y_1+\dots+Y_n} = \phi_{Y_1} \dots \phi_{Y_n} = \phi_{Y_1}^n$ par indépendance et identique distribution des variables Y_i .

Comme Y_1 a un moment d'ordre 2, ϕ_{Y_1} est dérivable 2 fois avec $\phi_{Y_1}(0) = 1$, $\phi'_{Y_1}(0) = i\mathbb{E}[Y_1] = 0$ et $\phi''_{Y_1}(0) = i^2\mathbb{E}[Y_1^2] = -\sigma^2$. La formule de Taylor à l'ordre 2 en 0 donne alors

$$\begin{aligned}\phi_{Y_1}(x) &= \phi_{Y_1}(0) + x\phi'_{Y_1}(0) + \frac{x^2}{2}\phi''_{Y_1}(0) + x^2\epsilon(x) \\ &= 1 - \frac{\sigma^2 x^2}{2} + x^2\epsilon(x)\end{aligned}$$

où la fonction ϵ vérifie $\lim_{x \rightarrow 0} \epsilon(x) = 0$. On a donc

$$\begin{aligned}\phi_{Z_n}(t) &= \left(\phi_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{\sigma^2 t^2}{2\sigma^2 \sqrt{n}^2} + \frac{t}{\sigma^2 \sqrt{n}^2} \epsilon(t/(\sigma\sqrt{n}))\right)^n = \left(1 - \frac{t^2}{2n} + \frac{1}{n} \epsilon(1/\sqrt{n})\right)^n \\ &= \exp\left(n \ln\left(1 - \frac{t^2}{2n} + \frac{1}{n} \epsilon(1/\sqrt{n})\right)\right) = \exp\left(n\left(-\frac{t^2}{2n} + \frac{1}{n} \epsilon(1/\sqrt{n})\right)\right) \\ &= \exp\left(-\frac{t^2}{2} + \epsilon(1/\sqrt{n})\right).\end{aligned}$$

D'où pour chaque $t \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \phi_{Z_n}(t) = \exp\left\{-\frac{t^2}{2}\right\} = \phi_{\mathcal{N}(0,1)}(t).$$

Le théorème de Paul Lévy affirme alors que Z_n converge en loi vers $\mathcal{N}(0, 1)$, c'est bien la conclusion du TCL. \square

Remarque 7.3.2 En général, lorsque n est grand, on approxime la loi d'une somme de v.a. i.i.d. de $L^2(\Omega)$ par une loi normale grâce au TCL de la façon suivante. Soit $S_n = X_1 + \dots + X_n$ la somme de v.a. i.i.d. X_i avec $\sigma_X^2 < +\infty$, on a d'après le TCL

$$\frac{X_1 + \dots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Quand n est grand on approxime alors la loi de $\frac{X_1 + \dots + X_n - n\mathbb{E}[X_1]}{\sigma\sqrt{n}}$ par celle de $Y = \mathcal{N}(0, 1)$. Si bien que la loi de la somme $S_n = X_1 + \dots + X_n$ est approximée par celle de

$$n\mathbb{E}[X_1] + \sigma\sqrt{n}Y \simeq \mathcal{N}(n\mathbb{E}[X_1], \sigma^2 n).$$

Règle statistique : La somme S_n d'une suite de v.a.i.i.d. L^2 de moyenne m et de variance σ^2 s'approxime par

$$S_n \approx \mathcal{N}(nm, \sigma^2 n).$$

En particulier comme une loi binomiale $\mathcal{B}(n, p)$ peut se voir comme la loi d'une somme X_n de n va X_i de loi de Bernoulli $b(p)$ indépendante, on a l'approximation d'une loi binomiale par une loi normale

$$\mathcal{B}(n, p) \approx \mathcal{N}(np, np(1-p)).$$

Exemple : Un joueur lance une pièce équilibrée : lorsqu'il obtient pile, il gagne 100 Euros, lorsqu'il obtient face, il perd 100 Euros. Estimer le nombre maximal de lancers à effectuer pour que ce joueur ait plus de 95 chances sur 100 de perdre au plus 2000 Euros.

Notons n le nombre de lancers effectués, la v.a. S_n égale au nombre de piles obtenus sur les n premiers lancers suit une loi $\mathcal{B}(n, 1/2)$ et le gain (algébrique) vaut :

$$G_n = 100 \times S_n - 100 \times (n - S_n) = 200S_n - 100n.$$

On cherche alors n tel que $\mathbb{P}(G_n \geq -2000) \geq 0,95$. Or $\{G_n \geq -2000\} = \{S_n - n/2 \geq -10\}$.

Comme S_n de loi binomiale, peut être vue comme une somme $S_n = \epsilon_1 + \dots + \epsilon_n$ de v.a. de loi $b(1/2)$, d'après le TCL

$$S_n^* = \frac{S_n - n/2}{\sqrt{n/4}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Si bien qu'on approxime la loi de S_n^* par $\mathcal{N}(0, 1)$ et donc celle de S_n par la loi normale $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$.

Chercher n tel que $\mathbb{P}(G_n \geq -2000) = \mathbb{P}(S_n - n/2 \geq -10) \geq 0,95$ revient à estimer n tel que

$$\mathbb{P}(\mathcal{N}(0, 1) \geq -20/\sqrt{n}) \geq 0,95 \quad \text{ou} \quad \mathbb{P}(\mathcal{N}(0, 1) \leq 20/\sqrt{n}) \geq 0,95$$

par symétrie de la loi $\mathcal{N}(0, 1)$. La table de la loi $\mathcal{N}(0, 1)$ donne alors

$$\frac{20}{\sqrt{n}} = 1,65 \quad \text{c'est à dire} \quad n = \left(\frac{20}{1,65}\right)^2 = 146.$$

Remarque 7.3.3 (Autres applications statistiques) Le TCL est fondamental en statistique pour établir des intervalles de confiance ou pour faire des tests d'hypothèses à partir d'échantillons aléatoires (test du χ^2 , test de Student ou autres).

Lorsque les v.a.i. ne sont plus identiquement distribuées, le TCL classique n'est pas applicable. On a cependant un résultat du type TCL lorsque les variances sont uniformément contrôlées (cf. (7.5)).

Théorème 7.3.2 (TCL de Lindeberg) Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.i. centrées et de variances σ_n^2 finies pour tout $n \in \mathbb{N}^*$. On note $S_n^2 = \sum_{k=1}^n \sigma_k^2$. Si pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow +\infty} \frac{1}{S_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}] = 0 \quad (7.5)$$

alors on a la convergence en loi

$$\frac{1}{S_n} \sum_{k=1}^n X_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarque 7.3.4 Le TCL classique découle de celui de Lindeberg.

En effet, d'abord quitte à travailler avec $Y_i = X_i - \mathbb{E}[X_i]$, il n'y a pas de restriction à supposer les X_i centrées. Puis dans le TCL classique, les v.a. X_n sont iid, donc $\sigma_n^2 = \sigma^2$ et $S_n^2 = n\sigma^2$. La condition (7.5) devient $A_n := \frac{1}{n\sigma^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon\sigma\sqrt{n}}] \rightarrow 0, \quad n \rightarrow +\infty$.

Or

$$\begin{aligned} A_n &= \frac{1}{n\sigma^2} \sum_{k=1}^n \int_{[-\varepsilon\sigma\sqrt{n}, \varepsilon\sigma\sqrt{n}]^c} x^2 \mathbb{P}_{X_k}(dx) \\ &= \frac{1}{n\sigma^2} n \int_{[-\varepsilon\sigma\sqrt{n}, \varepsilon\sigma\sqrt{n}]^c} x^2 \mathbb{P}_{X_1}(dx) \\ &= \frac{1}{\sigma^2} \int_{[-\varepsilon\sigma\sqrt{n}, \varepsilon\sigma\sqrt{n}]^c} x^2 \mathbb{P}_{X_1}(dx) \rightarrow 0, \quad n \rightarrow +\infty \end{aligned}$$

car comme le domaine d'intégration $[-\varepsilon\sigma\sqrt{n}, \varepsilon\sigma\sqrt{n}]^c \rightarrow \emptyset$ et X_1 admet un moment d'ordre 2 fini, le théorème de convergence dominée s'applique.

La condition (7.5) de Lindeberg est donc remplie et le TCL de Lindeberg donne la conclusion du TCL classique.

On commence la preuve du TCL de Lindeberg par 3 lemmes préliminaires

Lemme 7.3.1 Pour tout $k \in \mathbb{N}$ et $u \in \mathbb{R}$, on a

$$\left| e^{iu} - \sum_{j=0}^{k-1} \frac{(iu)^j}{j!} \right| \leq \frac{u^k}{k!}. \quad (7.6)$$

Démonstration : On raisonne par récurrence sur k . En effet pour $k = 1$, on a

$$|e^{iu} - 1|^2 = (\cos u - 1)^2 + \sin^2 u = 2 - 2\cos u = 2 \int_0^u \sin v dv \leq 2 \int_0^u v dv = [v^2]_0^u = u^2.$$

La majoration (7.6) est vérifiée pour $k = 1$. Supposons la vérifiée au rang k et montrons la au rang $k + 1$. On a

$$\left| e^{iu} - \sum_{j=0}^k \frac{(iu)^j}{j!} \right| = \left| \int_0^u (e^{iv} - \sum_{j=0}^{k-1} \frac{(iv)^j}{j!}) dv \right| \leq \int_0^u |e^{iv} - \sum_{j=0}^{k-1} \frac{(iv)^j}{j!}| dv \leq \int_0^u \frac{v^k}{k!} dv = \frac{u^{k+1}}{(k+1)!}$$

ce qui prouve (7.6) au rang $k + 1$ et donc (7.6) par récurrence. \square

Lemme 7.3.2 Sous la condition (7.5), on a

$$\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq n} (\sigma_k / S_n) = 0. \quad (7.7)$$

Démonstration : En effet pour $\varepsilon > 0$ fixé

$$\begin{aligned} \max_{1 \leq k \leq n} (\sigma_k^2 / S_n^2) &\leq \max_{1 \leq k \leq n} \frac{\mathbb{E}[X_k^2]}{S_n^2} \leq \max_{1 \leq k \leq n} \frac{\mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}]}{S_n^2} + \max_{1 \leq k \leq n} \frac{\mathbb{E}[X_k^2 \mathbf{1}_{|X_k| \leq \varepsilon S_n}]}{S_n^2} \\ &\leq \underbrace{\sum_{k=1}^n \frac{\mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}]}{S_n^2}}_{\rightarrow 0, \text{ d'après (7.5)}} + \underbrace{\max_{1 \leq k \leq n} \frac{\mathbb{E}[X_k^2 \mathbf{1}_{|X_k| \leq \varepsilon S_n}]}{S_n^2}}_{\leq \varepsilon^2}. \end{aligned}$$

D'où pour tout $\varepsilon > 0$, on a $\lim_{n \rightarrow +\infty} \max_{1 \leq k \leq n} (\sigma_k / S_n) < \varepsilon$, le résultat (7.7) suit quand $\varepsilon \searrow 0$. \square

Lemme 7.3.3 Pour $|a_k| \leq 1$ et $|a_k + b_k| \leq 1$, on a

$$\left| \prod_{k=1}^n a_k - \prod_{k=1}^n (a_k + b_k) \right| \leq \sum_{k=1}^n |b_k|. \quad (7.8)$$

Démonstration : Récurrence facile.

On démontre maintenant le TCL de Lindeberg.

Démonstration : Soit $\varepsilon > 0$, notons ϕ_k la fonction caractéristique de X_k . On a

$$\phi_k\left(\frac{t}{S_n}\right) = \mathbb{E}[\exp(i \frac{t}{S_n} X_k)] = \mathbb{E}[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| > \varepsilon S_n}] + \mathbb{E}[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| \leq \varepsilon S_n}].$$

Puis (7.6) donne pour $k = 2$ et 3 respectivement

$$\begin{aligned} \left| \exp(i \frac{t}{S_n} X_k) - \left(1 + i \frac{t}{S_n} X_k\right) \right| &\leq \frac{|t|^2}{2S_n^2} X_k^2, \\ \left| \exp(i \frac{t}{S_n} X_k) - \left(1 + i \frac{t}{S_n} X_k - \frac{t^2}{2S_n^2} X_k^2\right) \right| &\leq \frac{|t|^3}{6S_n^3} X_k^3. \end{aligned}$$

Comme $\mathbb{E}[X_k] = 0$ et $\mathbb{E}[X_k^2] = \sigma_k^2$, on a (1ère égalité) :

$$\begin{aligned} &\left| \phi_k\left(\frac{t}{S_n}\right) - \left(1 - \frac{t^2 \sigma_k^2}{2S_n^2}\right) \right| \\ &= \left| \mathbb{E}[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| > \varepsilon S_n}] + \mathbb{E}[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| \leq \varepsilon S_n}] \right. \\ &\quad \left. - E \left[\left(1 + i \frac{t}{S_n} X_k - \frac{t^2}{2S_n^2} X_k^2\right) (\mathbf{1}_{|X_k| > \varepsilon S_n} + \mathbf{1}_{|X_k| \leq \varepsilon S_n}) \right] \right| \\ &= \left| E \left[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| \leq \varepsilon S_n} - \left(1 + i \frac{t}{S_n} X_k - \frac{t^2}{2S_n^2} X_k^2\right) \mathbf{1}_{|X_k| \leq \varepsilon S_n} \right] \right| \\ &\quad + \left| E \left[\exp(i \frac{t}{S_n} X_k) \mathbf{1}_{|X_k| > \varepsilon S_n} - \left(1 + i \frac{t}{S_n} X_k - \frac{t^2}{2S_n^2} X_k^2\right) \mathbf{1}_{|X_k| > \varepsilon S_n} \right] \right| \end{aligned}$$

$$\begin{aligned}
&\leq E \left[\left| \exp\left(i\frac{t}{S_n}X_k\right) - \left(1 + i\frac{t}{S_n}X_k - \frac{t^2}{2S_n^2}X_k^2\right) \right| \mathbf{1}_{|X_k| \leq \varepsilon S_n} \right] \\
&\quad + E \left[\left| \exp\left(i\frac{t}{S_n}X_k\right) - \left(1 + i\frac{t}{S_n}X_k\right) \right| \mathbf{1}_{|X_k| > \varepsilon S_n} \right] + \frac{t^2}{2S_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}] \\
&\leq \frac{|t|^3}{3!} E \left[\frac{X_k^3}{S_n^3} \mathbf{1}_{|X_k| \leq \varepsilon S_n} \right] + \frac{t^2}{2!} E \left[\frac{X_k^2}{S_n^2} \mathbf{1}_{|X_k| > \varepsilon S_n} \right] + \frac{t^2}{2S_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}] \\
&\leq \frac{|t|^3}{3!} \frac{\varepsilon}{S_n^2} \sigma_k^2 + \frac{t^2}{S_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}]. \tag{7.9}
\end{aligned}$$

Notons $a_k = \phi_k(t/S_n)$ et $b_k = \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right) - \phi_k(t/S_n)$.

On a $|a_k| \leq 1$ et d'après (7.7) : $|a_k + b_k| = \left|1 - \frac{t^2\sigma_k^2}{2S_n^2}\right| \leq 1$ pour n assez grand et pour tout k et tout $t \in \mathbb{R}$. On a par ailleurs

$$\left| \prod_{k=1}^n a_k - \prod_{k=1}^n (a_k + b_k) \right| \leq \sum_{k=1}^n |b_k| \tag{7.10}$$

$$\leq \sum_{k=1}^n \frac{|t|^3}{3!} \frac{\varepsilon}{S_n^2} \sigma_k^2 + \frac{t^2}{S_n^2} \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}] \tag{7.11}$$

$$\begin{aligned}
&\leq \varepsilon \frac{|t|^3}{6} \frac{\sum_{k=1}^n \sigma_k^2}{S_n^2} + \frac{t^2}{S_n^2} \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbf{1}_{|X_k| > \varepsilon S_n}] \\
&\leq \varepsilon (|t|^3/6 + t^2) \tag{7.12}
\end{aligned}$$

pour n assez grand, d'après l'hypothèse (7.5). Notez que (7.9) a été utilisée dans (7.11).

Notons φ_n la fonction caractéristique de $\frac{X_1 + \dots + X_n}{S_n}$. Comme les v.a. X_k sont indépendantes, on a $\varphi_n(t) = \prod_{k=1}^n \phi_k(t/S_n)$ et donc pour n assez grand d'après (7.12) :

$$\left| \varphi_n(t) - \prod_{k=1}^n \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right) \right| \leq \varepsilon \left(\frac{|t|^3}{6} + t^2 \right).$$

Finalement on a montré que pour tout $t \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \lim_{n \rightarrow +\infty} \prod_{k=1}^n \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right)$$

Or comme $S_n \rightarrow +\infty$, on a

$$\begin{aligned}
\ln \left(\prod_{k=1}^n \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right) \right) &= \sum_{k=1}^n \ln \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right) = \sum_{k=1}^n -\frac{t^2\sigma_k^2}{2S_n^2} - \frac{t^2\sigma_k^2}{2S_n^2} \varepsilon(1/S_n) \\
&= -\frac{t^2}{2} \frac{\sum_{k=1}^n \sigma_k^2}{S_n^2} - \frac{t^2}{2} \frac{\sum_{k=1}^n \sigma_k^2}{S_n^2} \varepsilon(1/S_n) = -\frac{t^2}{2} (1 + \varepsilon(1/S_n)) \\
\lim_{n \rightarrow +\infty} \ln \left(\prod_{k=1}^n \left(1 - \frac{t^2\sigma_k^2}{2S_n^2}\right) \right) &= -\frac{t^2}{2}.
\end{aligned}$$

car $\sum_{k=1}^n \sigma_k^2 = S_n^2$. On a donc

$$\lim_{n \rightarrow +\infty} \varphi_n(t) = \exp(-t^2/2) = \phi_{\mathcal{N}(0,1)}(t).$$

D'après le théorème de Paul Lévy, la v.a. $\frac{X_1 + \dots + X_n}{S_n}$ converge en loi vers $\mathcal{N}(0, 1)$. \square